2023

Exploring Modelling Assumptions and their Impact on Extreme Discharges for the Meuse Catchment

MASTER'S THESIS

ELIZABETH TAYLOR

Exploring Modelling Assumptions and their Impact on Extreme Discharges for the Meuse Catchment

FINAL REPORT

By

Elizabeth TAYLOR

In partial fulfillment of the requirements for the degree of

Master of Science

Civil Engineering

At the Delft University of Technology,

To be defended publicly October 24, 2023 at 1:30 PM

Thesis Committee:	Dr. Elisa Ragno	TU Delft
	Dr. Markus Hrachowitz	TU Delft
	Dr. Laurène Bouaziz	Deltares
	Dr. Anaïs Couasnon	Deltares

ABSTRACT

Floods are the most frequent natural disaster and due to climate change the frequency and intensity of these events are increasing. Therefore, it is becoming increasingly important to obtain accurate estimations of extreme discharges. Statistical modelling is widely used to estimate extreme discharges by fitting observed extreme discharges to an extreme value distribution. However, limited historical data makes it difficult to confidently model the tail behavior of extremes. Additionally, several modelling assumptions impact extreme discharge estimates including selection of the nonstationary method, extreme value distribution, parameter estimation method, and the impact of seasonality. In an effort to reduce uncertainties, a new method has been developed to derive design discharges for the Meuse in the Netherlands. This method, GRADE (Generator of Rainfall and Discharge Extremes) consists of three components: a stochastic weather generator, a hydrological model, and an extreme value analysis (EVA). However, the stochastic weather generator is not capable of producing daily rainfall that exceeds the range of historical data. Therefore, a physically based climate model, RACMO, is now being studied. RACMO is capable of generating 1,040 years of synthetic meteorological data that can be routed in a hydrological model to obtain 1,040 years of synthetic discharges. The physically based climate model makes it possible to capture the underlying physical processes of extreme events and the hydrological model can provide discharge information at locations where there are no observations. This thesis evaluates the impact various modelling assumptions have on estimated discharges using synthetic data generated by the RACMO through application of a case study in the Meuse.

The identified modelling assumptions are individually evaluated to assess their impact on extreme discharge estimates and their uncertainty. Based on the results presented in this thesis, longer record lengths provide considerable value in an EVA. At the six stations of interest, the width of the confidence interval of 100-year discharges obtained using Gumbel (GEV) for a record length of 1,040 years decreased between 75 and 85 percent (70 and 89 percent) from estimates obtained from a record length of 65 years. However, this increase in certainty could still lead to significant over and underestimations of extreme discharge if the other modelling assumptions are not carefully considered.

It is discovered the tail behavior of extreme discharges varies throughout the Meuse, so it is important to compare the fit of distributions at each location where estimates are required. At the six stations of interest, GEV and Gumbel (GP and Gumbel) estimates of the 100-year discharge obtained from the 1,040 years of synthetic data vary between 3 and 14 (4 and 9) percent. Additionally, the MLE and MoM parameter estimation methods used to fit the distribution to the 1,040 years of synthetic data resulted in up to a 9 percent difference in estimates of the 100-year discharge. Neglecting seasonality resulted in up to a 2 percent difference in 100-year discharges obtained from the 1,040 years of synthetic data. However, it is discovered there are limited number of summer extremes in the modelled discharges. Neglecting seasonality in historical data led to under and overestimations up to 24 and 22 percent, respectively.

The results presented in this thesis demonstrate the added value of estimating extreme discharges using ensemble members from physically based climate models. Additionally, it is emphasized that each modelling assumption impacts estimated discharges and should, therefore, be carefully considered.

ACKNOWLEDGMENTS

I decided to pursue my master's because, I knew if didn't take this step, it would be a decision I'd later regret. The past two years have thoroughly challenged me. I've studied in a foreign country, worked with people with diverse perspectives, spent weeks studying from sticky notes attached to my kitchen, bathroom, and bedroom walls, and even had the opportunity to travel to Romania on fieldwork. Through these experiences, I've been blessed to have met many international friends who make the Netherlands feel more like home.

During the past eight months I've been fortunate to receive invaluable guidance and support from my thesis committee. I am extremely grateful to my primary supervisor, Dr. Elisa Ragno, who consistently made time for discussions and generously shared several papers relevant to my research. I am thankful for Dr. Markus Hrachowitz who shared constructive feedback that inspired me to critically assess the assumption of stationarity in the generation of synthetic discharges. Additionally, I'd like to thank Dr. Anais Couasnon for her feedback during our weekly meetings and Dr. Laurène Bouaziz for her guidance in performing Thiessen polygons in python.

I'd like to thank my family in the US who made a trip last fall to support me and my sister run (and beat one of my professors!) in the Amsterdam Marathon. To my mom, who has made countless sacrifices to ensure my academic success, thank you for encouraging me to continuously challenge myself. Somehow, I think you knew I had the mind of an engineer before I did. And to my sisters, Emily and Lauren, thank you for brightening my days through frequent video calls. I feel blessed to have such supportive sisters who make the distance feel insignificant.

To my husband, Cesare, thank you for your unwavering support and encouragement throughout the past two demanding years as I pursue my master's. Your belief in me has been a constant source of motivation. I am extremely fortunate to have someone as sweet, patient, and loving, by my side.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Name
AIC	Akaike Information Criterion
AM	Annual Maxima
aST	Approximated Stationary Model
CDF	Cumulative Distribution Function
CI	Confidence Interval
DEM	Digital Elevation Model
ECDF	Empirical Cumulative Distribution Function
EVA	Extreme Value Analysis
EVT	Extreme Value Model
GEV	Generalized Extreme Value distribution
GOF	Goodness of fit
GRADE	Generator of Rainfall and Discharge Extremes
GP	Generalized Pareto Distribution
i.i.d.	Independent and identically distributed
KNMI	Koninlijk Nederlands Meteorologisch Instituut (Royal Netherlands Meteorological Institute)
KS	Kolmogorov Smirnov
MK	Mann Kendall
MLE	Maximum Likelihood Estimation
MoM	Method of Moments
MRLP	Mean Residual Life Plot
NEVA	Nonstationary Extreme Value Analysis
NS	Nonstationary Linear Model
PDF	Probability Distribution Function
POT	Peak over Threshold
PWM	Probability Weighted Moments
QGIS	Quantum Geographic Information Software
Q-Q plot	Quantile-Quantile plot
Q_{100}	Discharge corresponding to the 100-year return period
RACMO	Regional Atmospheric Climate MOdel
RP	Return Period
SPW	Service Public de Wallonie
uST	Updated Stationary Model

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Floods are the most frequent natural disaster that can devastate communities causing property damage and loss of life. Over 2 billion people were affected by floods globally between 1998 and 2017 (World Health Organization, 2023). Floods are increasing in frequency and intensity as a result of human intervention, natural climatic variability, and climate change; therefore, accurate estimation of the magnitude and frequency of floods is becoming increasingly important to manage and design flood defenses (Salas & Obeysekera, 2014).

Statistical modelling is widely used to estimate extreme discharges by fitting extreme discharges to an extreme value distribution. Annual maxima (AM) or peaks over a threshold (POT) are methods commonly used to obtain extremes. Extreme value models are used to estimate return periods (RP) which describe the probability of an extreme event to occur in any given year. In the Netherlands, discharge corresponding to the 1250-year RP is used for the design and maintenance of flood defenses in riverine areas. In coastal areas the 10,000-year RP is used. Due to limited historical data, statistical extrapolation of the distribution of AM discharges is required to obtain the 1250-year design RP which introduces large uncertainties in estimated discharges (Parmet & Mulders, 1999).

In an effort to reduce uncertainties, a new method has been developed to derive design discharges for the Meuse in the Netherlands. This method, GRADE (Generator of Rainfall and Discharge Extremes) consists of three components: a weather generator, a hydrological model, and an extreme value analysis (EVA). The stochastic weather generator uses a nonparametric resampling technique to simulate 50,000 years of daily rainfall and temperature data (Hegnauer M., 2014). However, the stochastic weather generator had two main limitations: it was limited to producing observed rainfall amounts which would make it difficult to incorporate climate change and it was not capable of producing daily rainfall that exceeded the range of historical data (Hegnauer M., 2014).

Due to the limitations of the stochastic weather generator, a physically based climate model capable of generating long synthetic meteorological time series is now being studied. The climate model relies on equations derived from knowledge of climate processes to generate time series long enough to eliminate the need for statistical extrapolation to evaluate the 1250-year design RP. While physically based models are able to capture the underlying processes that statistical extrapolation cannot, there is still more to be learned about this new methodology. This thesis evaluates the impact various modelling assumptions have on estimated discharges using synthetic data generated by the physically based climate model, RACMO, through application of a case study in the Meuse catchment.

1.1. BACKGROUND

A large portion of the Netherlands lies below sea level; therefore, the design and maintenance of flood protection is of critical importance. Large areas surrounding the Meuse River are prone to floods indicating a need to reduce the uncertainty in extreme discharge estimates used in design of flood defenses (Tu et al., 2005a).

In July 2021, slow propagation of a low-pressure system led to extreme precipitation, well beyond a 100-year event, that caused significant flooding in the Rhine and Meuse catchments in central Europe (Mohr et al., 2023). However, what made this event particularly unique was that it occurred during the summer, while most flood events in the Meuse occur during the winter. Extremes from different seasons can be driven by different flood mechanisms indicating a mixed distribution that should not be fit to a single extreme value distribution. Additionally, discharge estimates are often obtained from limited historical data, therefore, there may not be observed summer events. This makes it difficult to obtain reliable estimates of rare events, like what occurred in July 2021, from short historical records.

In the Netherlands, GRADE (Generator of Rainfall and Discharge Extremes) is used to obtain long synthetic discharges that can be used to obtain more robust estimates. GRADE consists of three components: a weather generator, hydrological model, and a flood frequency analysis. A statistical weather generator to obtain meteorological time series as long as 50,000 years that can be used in a hydrological model to obtain 50,000 years of synthetic discharges (Bouaziz et al., 2022). As shown in Figure 1, GRADE eliminates the need for statistical extrapolation of extreme value distributions since this process is capable of generating time series long enough to evaluate higher RP. Rather than performing an EVA on short historical records of river discharge, the analysis is performed on the long series of synthetic discharge produced using GRADE.

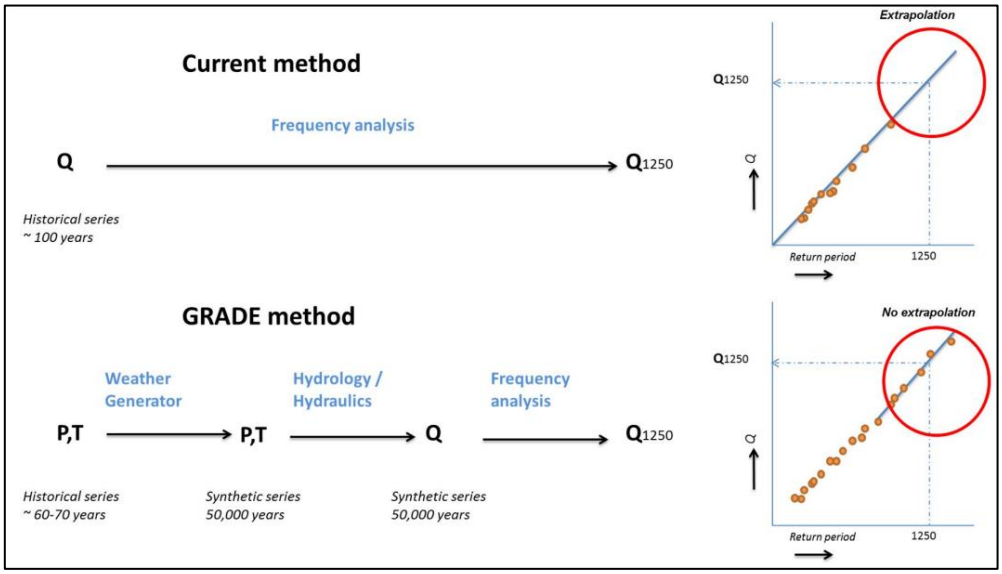


Figure 1: Comparison of Statistical Methods to the GRADE Methodology (*Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse Basins-Final Report of GRADE 2.0*, n.d.)

However, this method cannot be used to obtain estimates for smaller tributaries which have faster reactions and require sub-daily time steps to capture extreme discharges. Additionally, the statistical weather generator uses statistical resampling of observed time series and, therefore, is not capable of producing daily rainfall exceeding the range of historical data. No summer event in the 50,000 years generated by GRADE reached the magnitude of the July 2021 event. Therefore, although the increased record length reduces the statistical uncertainty in discharge estimates, the statistical weather generator used in GRADE is not able to capture the physics of extreme events.

The new methodology used to obtain long series of discharges is shown in Figure 2. RACMO, a physically based climate model is used to generate 16 ensemble members of 65 years, each with slightly different initial conditions resulting in 1,040 years of synthetic meteorological data. RACMO assumes the natural variability of the current climate and the data is bias corrected so that the resulting modelled data is considered to be stationary. Additionally, it is able to capture climate change and preserve the statistical properties of observations (Bouaziz et al., 2022). The long meteorological time series is used in a hydrological model to obtain longer discharge time series. No anthropogenic forcing is added to the hydrological model; therefore, climate change could be evaluated by comparing 30 years of modelled meteorological data with future scenarios of an equally long time series. The physically based climate model makes it possible to capture the underlying physical processes of extreme events and the hydrological model can provide discharge information at locations where there are no observations available.

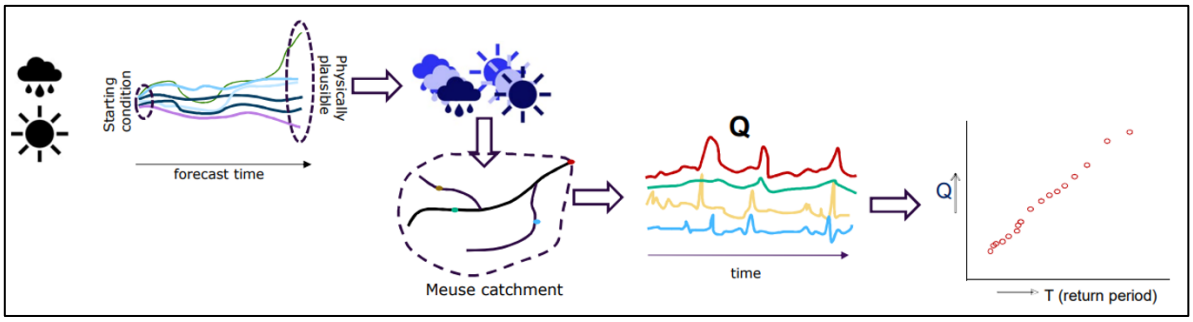


Figure 2: Method used to Generate Long Series of Discharges for EVA (Slomp, 2021)

1.1.1. MODELLING ASSUMPTIONS IMPACTING DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

Several modelling assumptions impacting discharge estimates and their uncertainty are identified including nonstationarity in observations, selection of an extreme value distribution, method used to estimate distribution parameters, effects of seasonality, and impact of record length. The modelling assumptions are presented Figure 3 and briefly described below.

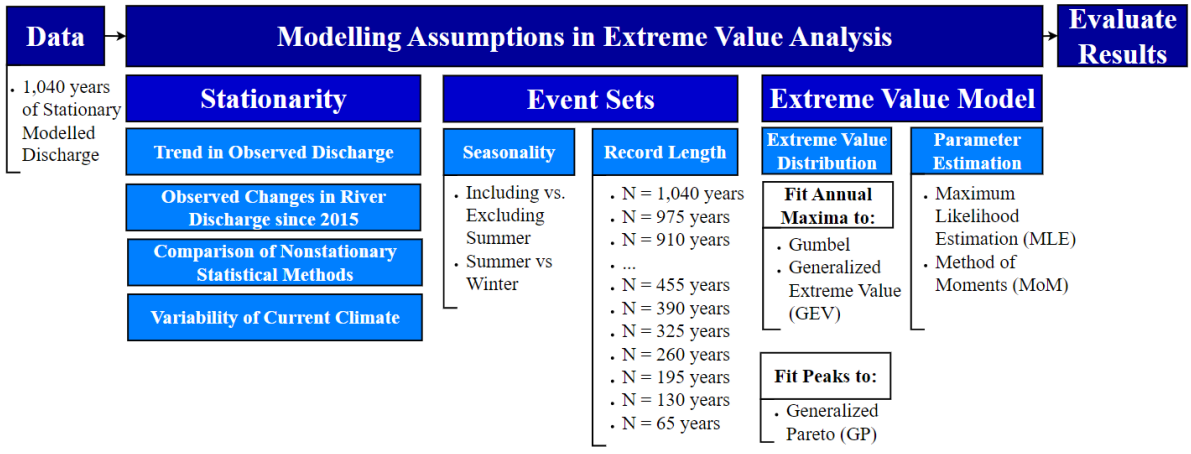


Figure 3: Modelling Assumptions Impacting Discharge Estimates and their Uncertainty

Stationarity
The synthetic data generated by RACMO is bias corrected so that the time series input into the hydrological model is stationary. Since no land use change is incorporated in the hydrological model, the generated modelled discharges are stationary. However, climate change and human intervention have led to increasing rainfall and river discharges. Therefore, before performing an EVA on observed discharge it is necessary to identify whether nonstationarity is present. However, statistical tests should not be solely relied on to infer the persistence of nonstationarity. The predictive capability of various nonstationary methods is compared to demonstrate the potential implications of extrapolating a nonstationary linear trend.

One of the limitations of GRADE is that the stochastic weather generator is limited to resampling from observed time series and is, therefore, not capable of producing daily rainfall exceeding the range of historical data. However, since the implementation of GRADE in 2015, events like what occurred in July 2021 add additional information about the statistics of extreme discharges. Therefore, observed changes in river discharges are studied to evaluate how much discharge estimates have changed since 2015.

Lastly, uncertainty in discharge estimated from performing a nonstationary EVA on observations is compared to uncertainty in estimates from various simulations of the current climate.

Extreme Value Model
Selection of an extreme value model is expected to impact discharge estimates. Therefore, various extreme value distributions will be compared to explore their impact on estimated discharges. In addition, the parameter estimation methods used to fit extreme value distributions can impact return level estimates and will also be included in this study.

Event Sets
Event sets describe the time series that is used to obtain discharge estimates such as the seasonality and record length. Event sets are expected to impact estimations of extreme discharge because they affect the amount of data that is included in the analysis. The availability of data and frequency of events is often limited; therefore, event sets are important to consider when studying the impact of extreme discharge estimates.

Limited historical data presents a significant impact on the sensitivity of estimates. The long synthetic discharge series generated by RACMO can be used to evaluate estimated discharges and their uncertainty at various record lengths.

Flood events of the Meuse generally result from back-to-back rainfall events that occur in the winter (Hegnauer et al., 2014). Tu et al. (2005a) found that 82 of 91 AM at Borgharen, where the Meuse enters the Netherlands, during the period 1912 through 2002 occurred in the winter. However, as shown by the recent flooding in July 2021, the Meuse is also susceptible to summer floods. To study the effect of seasonality, discharge estimates can be compared

when including and neglecting summer rainfall events. In addition, an EVA can be performed by partitioning summer and winter events before combining the results.

1.2. OBJECTIVES AND RESEARCH QUESTIONS

The goal of this master's thesis is to evaluate the impact various modelling assumptions have on estimated discharges using synthetic data generated by RACMO. Estimates and their uncertainty will be compared for various extreme value models, parameter estimation methods, record lengths, and seasonal assumptions. Before analyzing the stationary modelled data, available observations are studied to identify any potential trends in the time series and evaluate the performance of various nonstationary methods.

The following research questions have been developed to fulfill the objective of this thesis:

- *How much impact can a very extreme event, such as what occurred in July 2021 in central Europe, have on the statistics of extreme discharge?*
- *To what extent do nonstationary statistical methods impact discharge estimates when a trend in observed discharge is identified?*
- *How does the confidence of discharges estimated from various simulations of the current climate compare to the confidence of estimates resulting from a nonstationary extreme value analysis?*
- *To what extent do the modelling assumptions in an extreme value analysis affect the estimation of extreme discharge return levels?*
    - *How much influence do event sets have on discharge estimates and their uncertainty?*
    - *How much influence do extreme value models have on discharge estimates and their uncertainty?*

1.3. THESIS OUTLINE

Chapter 2 provides a brief theoretical background on the methodology to answer the research questions through application of a case study introduced in Chapter 3. The methods and results that address the research questions in this thesis are split into three main chapters:

- Chapter 4: Nonstationarity of Observed Discharge
- Chapter 5: Extreme Value Models Influence on Discharge Estimates and Their Uncertainty
- Chapter 6: Event Sets Influence on Discharge Estimates and Their Uncertainty

Chapter 4 discusses nonstationarity of observed discharge, compares the performance of various nonstationary statistical methods, and compares the uncertainty of discharge estimates resulting from various simulations of the current climate to uncertainty resulting from a NEVA. The impact selection of extreme value distributions and parameter estimation methods have on discharge estimates and their uncertainty is evaluated in Chapter 5. Chapter 6 studies the impact event sets, record length and seasonality, have on discharge estimates and their uncertainty. A comprehensive discussion of the results presented in this thesis is included in Chapter 7 and conclusion and recommendations are presented in Chapter 8.

2. LITERATURE REVIEW

This chapter describes the theoretical background of the methods and tools used throughout this thesis. Section 2.1 describes the frequency and probability of exceedance of extreme events. The theoretical background behind a stationary EVA is discussed in Section 2.2. Parameter estimation methods are described in Section 2.3. Statistical tests to evaluate or compare the fit of distributions are discussed in Section 2.4. Statistical methods that are used to analyze observed trends are described in Section 2.5. Various methods of performing NEVA are introduced in Section 2.6. Lastly, a method of combining partitioned EVA, which is used when summer and winter extremes are separated, is discussed in Section 2.7.

2.1. FREQUENCY AND PROBABILITY OF EXCEEDANCE

When analyzing extreme events, frequency and probability of exceedance are often used interchangeably. However, there is a difference. Exceedance probability quantifies the chance of an event to occur at least once during a certain period, such as a year, and is equal to the inverse of the RP of annual extremes, T_a. Frequency quantifies the number of times an event occurs on average during a certain period and is equal to the inverse of the RP of an event, T.

For example, an event with an exceedance probability of 1/100 per year (T_a = 100 year) indicates that there is a probability of 1 in 100 that that the event will occur in any given year. The exceedance frequency of this event is also 1/100 per year (T = 100 year) which indicates the event will occur on average once every 100 years. However, probability and frequency differ for smaller RP; the lowest annual maxima could also occur during years with higher annual maxima. For example, an event with a frequency of 2/year occurs on average twice per year and has a RP, T, of 0.5 year. However, probability must be between 0 and 1 and the RP of annual maxima, T_a, must be greater than 1. Therefore, for very extreme events with large RPs probability and frequency are nearly equal because it is unlikely that a very extreme event will occur more than once per year. A common assumption is that extremes can be modelled as a Poisson process so the probability can be estimated from the frequency using the following equation.

$$P_{exceed} = 1 - e^{-N} \tag{2-1}$$

Where P_{exceed} is the probability of exceedance and N is the frequency of exceedance.

2.2. STATIONARY EXTREME VALUE ANALYSIS

Extreme Value Theory (EVT) is applied to model the stochastic behavior of extremes through an EVA so that extreme discharges can be estimated. While long synthetic time series generated from RACMO provide more information about the statistical behavior of extremes there is still some uncertainty for events with higher RPs. Events with higher RPs occur less frequently, so there are larger differences in the empirical distribution function for these events. Therefore, although more extremes are available it is still necessary to perform an EVA.

Application of EVT requires two main assumptions: events are independent and identically distributed (i.i.d.) and there are a large enough number of events each year so that the distribution of the extremes is asymptotic (Marra et al., 2019).

EVA refers to the process of modelling events that deviate from the median of a probability distribution, e.g., fitting a distribution to either the AM or values above a specified threshold. The two methods, Block Maxima and Peak over Threshold, are described in Section 2.2.1 and Section 2.2.2.

2.2.1. BLOCK MAXIMA

The block maxima approach involves dividing the time series into blocks, usually years, and taking the maximum value within each block. An advantage to taking the AM is that the extremes that are extracted will be independent from one another. According to the Fisher-Tippet-Gnedenko theorem, the maximum values of a sample of independent random values should converge to the Gumbel, Fréchet, or reverse Weibull distribution. These three distributions were combined to become the Generalized Extreme Value (GEV) distribution shown below (Nerantzaki & Papalexiou, 2022).

$$F(x) = \exp\left(-\left(1+\gamma\left(\frac{x-\alpha}{\beta}\right)\right)^{-1/\gamma}\right) \qquad 1+\frac{\gamma(x-\alpha)}{\beta} \geq 0 \; and \; \gamma \neq 0$$

$\gamma < 0$	Reverse Weibull (Type III) "light-tailed"
$\gamma = 0$	Gumbel (Type I)
$\gamma > 0$	Fréchet (Type II), "heavy-tailed"

(2-2)

Where: α, β, γ are the location, scale, and shape parameters, respectively.

The location parameter describes the position of the distribution along the x-axis and influences the up and down movement of the discharge frequency curve. The scale parameter describes the spread of the data and influences the slope of the discharge frequency curve. The shape parameter distinguishes the tail behavior displayed by the three types of the GEV distribution. The tail behavior of a distribution describes how quickly the probabilities decrease with increasing extremes. Heavy tail behavior refers to slowly decreasing probabilities and indicates that very extreme events occur frequently. While light tailed behavior refers to probabilities that decrease quickly indicating that very extreme events are not as common.

When the shape parameter is less than zero, equal to zero, or greater than zero the GEV is equivalent to the Weibull, Gumbel, or Fréchet distribution, respectively. The parent distribution determines the shape parameter of the resulting GEV distribution. As shown in Figure 4, the shape parameter is influenced by the tail of the distribution. Distributions whose tails decrease exponentially, such as the normal, exponential, gamma, lognormal, or Weibull distribution, lead to a shape parameter of zero. Distributions whose tail exhibit a polynomial decay, such as the Pareto or Student-t distribution, result in a positive shape parameter.

Distributions with finite tails, such as the uniform or beta distribution, lead to a negative shape parameter.

Sample sizes of AM are usually small, especially due to limited length of observations, resulting in return values with large uncertainties. Therefore, an alternative method, peak over threshold, is described below.

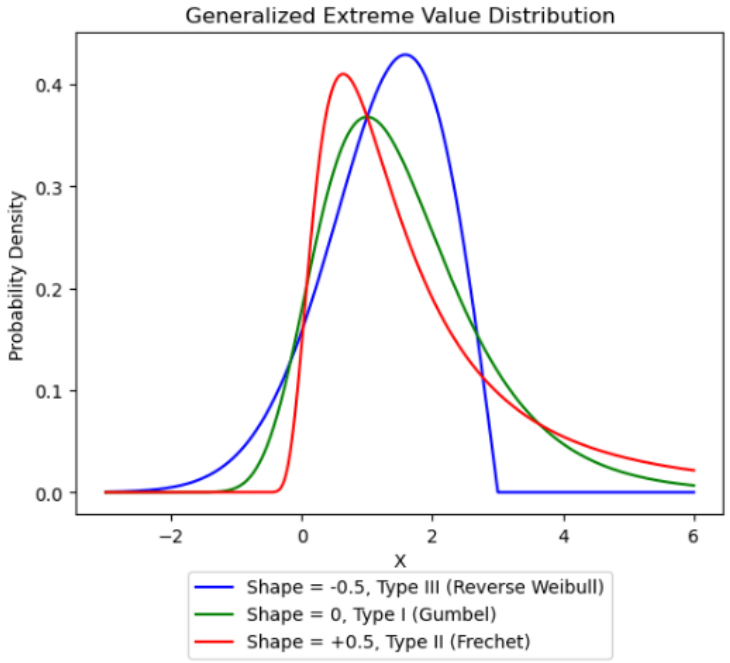


Figure 4: GEV Distribution Types

2.2.2. PEAK OVER THRESHOLD

The peak over threshold (POT) approach involves taking values of observations that exceed a certain threshold. According to the Balkema and de Haan and Pickands theorem, as the threshold increases, the peaks over the threshold converge to the Generalized Pareto (GP) distribution shown below (Nerantzaki & Papalexiou, 2022).

$$F(x) = \begin{cases} 1 - \left(1 + \gamma \frac{x-\alpha}{\beta}\right)^{-1/\gamma}, & for\ \gamma \neq 0 \\ 1 - \exp\left(-\frac{x-\propto}{\beta}\right), & for\ \gamma = 0 \end{cases} \quad (2\text{-}3)$$

$$-\infty < \gamma < \infty, x \geq \alpha\ when\ \gamma \geq 0, and\ \alpha \leq x \leq \alpha - \frac{\beta}{\gamma} when\ \gamma < 0$$

Where β and γ are the scale and shape parameters and α corresponds to the threshold, x_p.

Similar to GEV, the GP distribution has a type I, type II, or type III tail if the shape parameter is equal to zero, greater than zero, or less than zero, respectively. For a shape parameter equal to zero, GP is the exponential distribution. For a positive shape parameter, it is the Pareto distribution and for a negative shape parameter it is a special case of the beta distribution.

In order to apply the Balkema and de Haan and Pickands theorem, the data should be i.i.d as required by EVT, and the threshold should be optimized for convergence to the GP distribution. Hydrological time series are typically not i.i.d due to strong temporal

autocorrelation; hydrological processes are dependent on past conditions. Application of POT could lead to dependent data, therefore before applying EVT, declustering is required to ensure the peaks obtained from in the time series are independent.

Threshold Selection
Selection of a threshold should be carefully considered in a POT analysis because the choice of a threshold can have a strong impact on EVA results. High thresholds will result in less peaks and a large uncertainty while thresholds that are too low will result in a sample that is poorly modelled by the GP distribution.

According to threshold stability property of the GP distribution, if GP is valid for excesses over a threshold u_0, then it is valid for excesses over thresholds $u > u_0$. The expected value of threshold excesses given that they are greater than the threshold can be calculated using the following equation (Coles, 2001):

$$E[X - u|X > u] = \frac{\sigma_{u_0} + \xi u}{1 - \xi} \tag{2-4}$$

Where σ_{u_0} is the scale parameter for excesses over the threshold u_0 and ξ is the shape parameter.

From this equation it is clear that for thresholds $u > u_0$, the mean of excesses of the threshold u, $E[X - u|X > u]$, is a linear function of u. A Mean Residual Life Plot (MRLP) can be created by plotting the mean threshold excess against u. To determine a sufficiently high threshold for modelling GP, the MRLP can be used to select the lowest threshold where the graph is linear with increasing thresholds. The linear portion describes a valid range for thresholds because there will be higher variance and uncertainty for thresholds that are too high and have few exceedances and for thresholds that are too low such that the sample is poorly modelled. A slope of zero where the MRLP plot is linear indicates the extremes closely follow an exponential tail (El Adlouni et al., 2008).

2.3. PARAMETER ESTIMATION METHODS

One difficulty of applying statistical distributions is accurately estimating the unknown parameters. There are several methods to fit a parametric distribution to data. Method of Moments (MoM) and Maximum Likelihood Estimation (MLE) are among the most widely known and used methods for estimating parameters of Gumbel (Aydin & Şenoğlu, 2015a). Extreme discharges at Borgharen, where the Meuse enters the Netherlands, are closely modelled by Gumbel. Additionally, MoM and MLE are the two available methods used in the Scipy Stats Python package used to fit distribution parameters (Virtanen et al., 2020). Therefore, to evaluate the influence of parameter estimation methods, AM modelled discharge is fit to Gumbel using MoM and MLE.

2.3.1. MAXIMUM LIKELIHOOD ESTIMATION

One common method of estimating distribution parameters is maximum likelihood estimation (MLE). The probability density function (PDF) of the parametric distribution, *f,* to which the dataset, T = $\{t_i, \ldots, t_n\}$ will be fit has a given parameter set, θ. To fit a dataset to a parametric

distribution the estimated parameters, $\hat{\theta}$, must be found so that the theoretical distribution closely matches the distribution of the observations.

MLE is a method of finding the estimated parameters $\hat{\theta}$ by maximizing the likelihood of observing the data. The likelihood is the joint probability distribution of a specific probability distribution and its parameters. The likelihood function can be used to describe how likely a dataset is given the estimated parameters. Assuming observations are independent so that the total probability is equal to the product of the marginal probabilities, the likelihood function can be written as (Cousineau et al., 2004):

$$L(\hat{\theta}, T) = \prod_{i=1}^{n} f(t_i|\hat{\theta}) \tag{2-5}$$

2.3.2. METHOD OF MOMENTS

Method of moments (MoM) is another parameter estimation technique that could be advantageous because of its simplicity. MoM estimates parameters by equating sample and theoretical moments to obtain the unknown parameters (Hazelton, n.d.).

The statistical model is defined by a parameter vector, $\theta = (\theta_1,\ldots,\theta_p)^T$, the kth moment about zero of a random variable X is, $\mu_k = E[X_k]$, and the moment is a function of θ, $\mu_k = \mu_k(\theta)$.

For a random sample $X_1,\ldots,X_n$ the method of moments estimator $\hat{\theta}$ can be determined from the following equations*:*

$$\mu(\theta) = \hat{\mu}_k \qquad k = 1,2, \ldots, q \tag{2-6}$$

Where q is the smallest integer for which the system has a unique solution. The kth sample moment can be calculated as $\hat{\mu}_k = n^{-1}\sum_{i=1}^{n} X_i^k$

2.4. FORMAL AND INFORMAL TESTS

Formal and informal tests are commonly used to indicate how well a distribution fits a dataset. Formal tests are hypothesis-based tests that apply confidence levels that define the probability of rejecting a null hypothesis when it is true. For example, a confidence level of 0.05 indicates there is a 5 percent chance of incorrectly rejecting a null hypothesis. The Kolmogorov Smirnov test is a formal, hypothesis-based test that is commonly used to indicate how well a distribution fits data. Informal tests refer to tests that are not based on a formal hypothesis. Akaike's information criteria (AIC) is an informal test that is commonly used to compare the performance of statistical models.

2.4.1. FORMAL TESTS

The goodness-of-fit (GOF) of a distribution to a dataset can be evaluated through various methods including the Kolmogorov-Smirnov (KS) test and a quantile-quantile (Q-Q) plot.

The KS test is a formal test used to assess the GOF by calculating the maximum difference between the empirical cumulative distribution and the parametric cumulative distribution.

The empirical cumulative distribution function (ECDF) needs to be determined to evaluate how well an extreme value distribution fits observed data. The ECDF, $\hat{F}(x)$, is a step

function that describes the sample of observations that are less than or equal to a value x defined by the following equation (Taboga, 2021).

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n+1} 1_{\{x_i \leq x\}} \tag{2-7}$$

Where n is the sample size and $1_{\{x_i \leq x\}}$ is equal to 1 if $x_i \leq x$ and zero if $x_i > x$.

A graphical representation of the KS test is shown in Figure 5. If the deviation between the two distributions is small, then it is reasonable to assume the data fits the parametric distribution. The KS test statistic is calculated using the following equation.

$$D_n = sup_x|\hat{F}(x) - F(x)| \tag{2-8}$$

Where $\hat{F}(x)$ is the empicial cumulative distribution and $F(x)$ is the parametric distribution.

The null hypothesis of the KS test is that the random variable follows the distribution. If the p-value of the KS test is less than the desired confidence level, such as 0.05, the null hypothesis is rejected indicating the distribution is not a good fit.

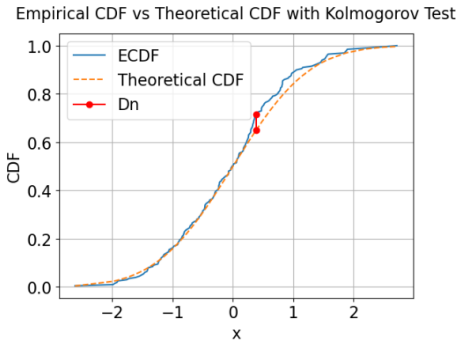


Figure 5: Kolmogorov-Smirnov Test

Another method for assessing GOF is a Q-Q plot, which is a graphical method. As shown in Figure 6, the x-axis contains the quantiles of the observations, and the y-axis contains the quantiles predicted by the fitted distribution. If the points lie on a 45-degree line y=x, the two distributions have a perfect fit. In the specific case shown in Figure 6, higher quantiles stray from the 45-degree line. For higher values, the theoretical quantiles estimated from the extreme value distribution are lower than the empirical quantiles indicating that the tail of the distribution underestimates high quantiles.

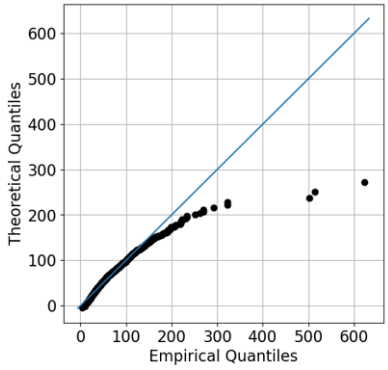


Figure 6: Q-Q Plot

2.4.2. INFORMAL TESTS

Akaike's information criteria (AIC) is an informal test to compare performance between statistical models and is commonly used to select the best distribution. The AIC is defined below (Akaike H., 1974):

$$AIC = 2k - 2\ln(\hat{L}) + \frac{2k^2 + 2k}{n - k - 1} \quad (2\text{-}9)$$

Where k is the number of estimated parameters in the model, $\hat{L}$ is the maximized value of the likelihood function, and n is the number of selected extremes.

As shown in the equation, the AIC penalizes distributions that require more parameters that need to be estimated. The model with the lowest AIC is determined to be the best fit model.

Burnham et al. (2011) state that models with AIC values that have a difference of 2 or less are equally good, models with a difference of 5 indicates the model with the lower AIC is better, and models with a difference of 10 has strong evidence that the model with the lowest AIC is a good fit.

2.5. TEMPORAL TREND ANALYSIS

2.5.1. MANN KENDALL TEST

Prior to performing an EVA, it is necessary to determine whether trends exist in the dataset. While parametric techniques assume the data follows a normal distribution, nonparametric techniques make no assumption about the underlying distribution. Nonparametric techniques are recommended to detect trends in time series because it is typically unknown whether the data in a time series is normally distributed or not. In addition, nonparametric techniques are less affected by missing data points compared to parametric techniques (Kamal et al., 2018). The Mann-Kendall (MK) test is a widely used nonparametric, hypothesis-based test to determine if there is a monotonic trend in a time series. The MK test can be used to evaluate if there is a trend in extreme discharges. MK's test statistic, S, describes the number of times the variable of interest in a time series, m_t, increases by calculating the number of times m_{t_2} is greater than m_{t_1}, minus the number of times m_{t_1} is greater than m_{t_2}. A positive value indicates the time series increases more than it decreases (Kendall M.G., 1955; Mann, 1945).

$$S = \sum_{t_1=1}^{T-1} \sum_{t_2=t_1+1}^{T} sgn(m_{t_2} - m_{t_1}) \quad (2\text{-}10)$$

Where m_t represents the variable of interest (extreme discharge), $sgn\ (\ )$ indicates the sign of $(m_{t_2} - m_{t_1})$, t= 1,…,T.

For datasets with more than 10 observations the MK's test statistic is normalized by calculating a Z score. The value of Z indicates an increasing trend if the value is positive and a decreasing trend if the value is negative. The Z value can be determined using the following formula (Kamal et al., 2018).

$$Z = \begin{cases} \frac{S-1}{\sqrt{var(s)}} & if\ S > 0 \\ 0 & if\ S = 0 \\ \frac{S+1}{\sqrt{var(s)}} & if\ S < 0 \end{cases} \quad (2\text{-}11)$$

The null hypothesis of the MK test is that there is no trend in the data and the alternative hypothesis is that there is a trend. The MK test assumes that under the null hypothesis the data are independently distributed in time. The null hypothesis is rejected if the specified significance level, α, is greater than the p-value.

The significance level is the probability of incorrectly rejecting a true null hypothesis. For example, a significance level of 0.10 indicates there is a 10 percent chance of incorrectly rejecting a true null hypothesis. For the MK test this means there is a 10 percent chance of concluding there is a trend in the data when there is not. In other words, there is a 90 percent confidence level that there is a trend if one is identified.

A two-tailed test is used such that the relationship is tested in both directions, both increasing and decreasing trends are considered. For a two-tailed test the p-value corresponding to a Z score can be determined using the following formula: p-value=2*(1-Area to left of Z-score) where the area to the left of a Z-score can be found using the standardized Z-tables, or standard normal tables.

In the example provided in Figure 7, the standard normal tables are used to determine the area to the left of the Z-score table as 0.9032. Since the total area under the curve is 1, the area to the right is 1-0.9032=0.0968 or 9.68% of the area under the curve. For a two-tailed test the area of rejection is equal to 9.68% of the area under the upper tail and 9.68% of the area under the lower tail as shown in Figure 7. This results in a p-value of 2*(0.9032) =0.1936. The null hypothesis is rejected if the alpha value exceeds this p-value.

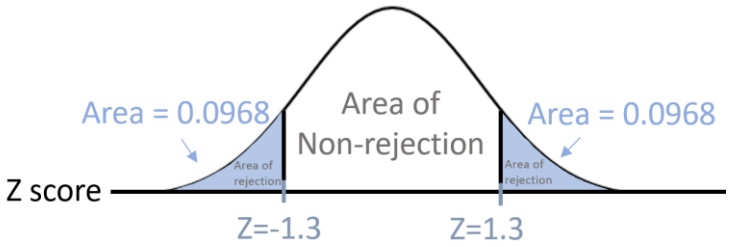


Figure 7: Example of two-tailed test, not drawn to scale.

When performing the MK test, it is important to consider the record length because the test results in higher accuracy for longer time series.

2.5.2. SEN'S SLOPE ESTIMATOR

While the MK test determines whether or not a trend exists in the AM of a time series, Sen's slope estimator can be used to determine the magnitude of that trend (Sen, 1968). Sen's slope estimator is a more robust method in determining the magnitude of a trend compared to linear regression because it is less affected by outliers and errors in data. The slope is determined using the following equation (Sen, 1968).

$$Q_i = \frac{x_j - x_i}{j - i}, i = 1, 2, \dots, N \quad (2\text{-}12)$$

Where x_j and x_i are values at time j and i (j>i). For n values of x_i in a time series, there will be N = n(n-1)/2 slope estimates.

2.6. NONSTATIONARY EXTREME VALUE ANALYSIS

A time series is stationary if it does not exhibit trends, shifts, or cyclicity and the statistical parameters are constant in time. While stationary time series contains statistical parameters that are constant in time, the statistical parameters of a nonstationary time series vary in time (Mudersbach & Jensen, 2010). The change in extremes can be modelled using other covariates such as land use and temperature or be modelled linearly, quadratically, or exponentially. However, increasing the number of parameters that need to be estimated increases model complexity and the uncertainty of the estimates. The three nonstationary methods that are studied in this thesis include the Nonstationary Linear Model, Approximated Stationary Model, and the Updated Stationary Model.

2.6.1. NONSTATIONARY (NS) LINEAR MODEL

In this thesis, Gumbel is used to compare the predictive capabilities of the various nonstationary statistical methods because the behavior of extremes at Borgharen is closely modelled by Gumbel. The nonstationary form of Gumbel with time varying parameters is shown below.

$$F(x) = e^{-e^{-(x-\alpha(t))/\beta(t)}} \quad (2\text{-}13)$$

Where, $\alpha(t)$ is time varying Gumbel location parameter at time t, $\beta(t)$ is the time varying Gumbel scale parameter.

In this thesis a Nonstationary Linear (NS) Model is used assuming the change in distribution parameters can be modelled as a linear function in time as shown in the equations below.

$$\alpha(t) = \alpha_0 + \alpha_1 t \quad (2\text{-}14)$$

$$\beta(t) = \beta_0 + \beta_1 t \quad (2\text{-}15)$$

Where, $\alpha(t)$ is the Gumbel location parameter at time t, $\beta(t)$ is the Gumbel scale parameter at time t, and α_0, α_1, β_0, and β_1 are the slope and intercept of the Gumbel parameters estimated from observations.

For this research, a sliding window length of 30 years is chosen, which covers approximately one climate period, so that the data within the window may be considered stationary (Mudersbach & Jensen, 2010). Therefore, the NS linear model used in this assumes that data within each 30-year window is stationary. An overview of the linear NS model adopted from Mudersbach and Jensen is provided in Figure 8 (Mudersbach & Jensen, 2010).

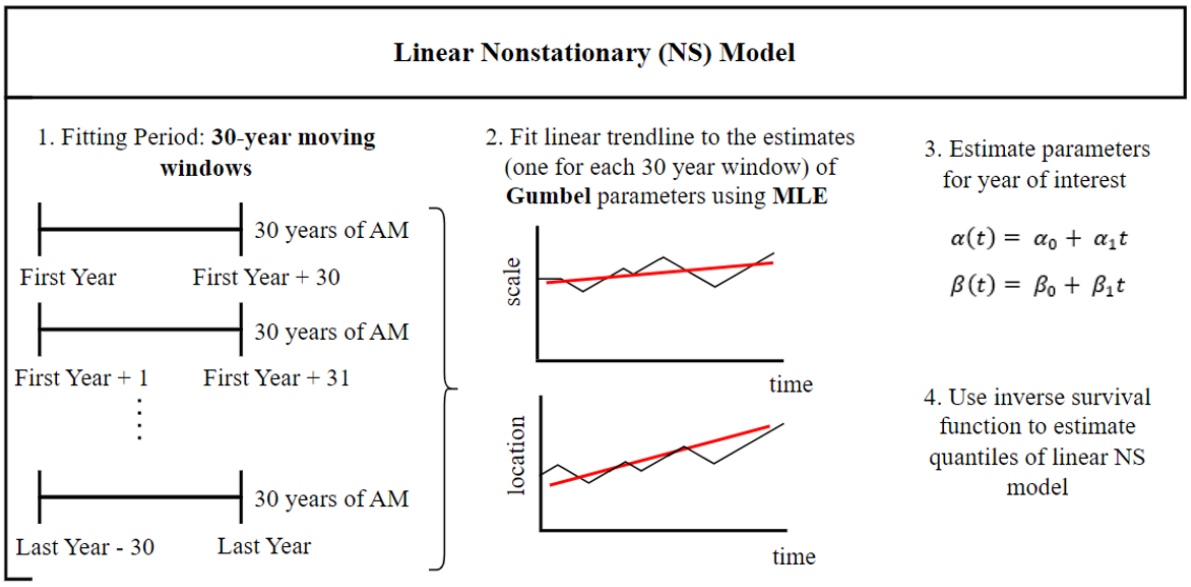


Figure 8: Overview of the Linear NS Model for the Gumbel Distribution (First Year and Last Year refer to the first and last year in the record or the first and last year in the fitting period of the distribution parameters) – Methodology adopted from (Mudersbach & Jensen, 2010)

2.6.2. APPROXIMATED STATIONARY (aST) MODEL

If it is uncertain whether observed trends will persist, approximated stationary models may be preferred over NS models. In the approximated stationary approach, the median Gumbel parameters within the fitting period are used to approximate the return levels for the evaluation period under the assumption of stationarity. Therefore, this approach assumes that the median return value of the linear trend within the fitting period can be used to obtain return values under stationary conditions (Luke et al., 2017). An illustration of the aST model is shown in blue in Figure 9.

2.6.3. UPDATED STATIONARY (uST) MODEL

Luke et. al. (2017) found that updated stationary models may be preferred over aST and NS models if discharges within the fitting period are impacted by physical changes in the watershed or if persistence of a trend is uncertain. The updated stationary model uses Gumbel parameters estimated by the NS model at the end of the fitting period to obtain return values under stationary conditions. An example of the uST model is shown in yellow in Figure 9.

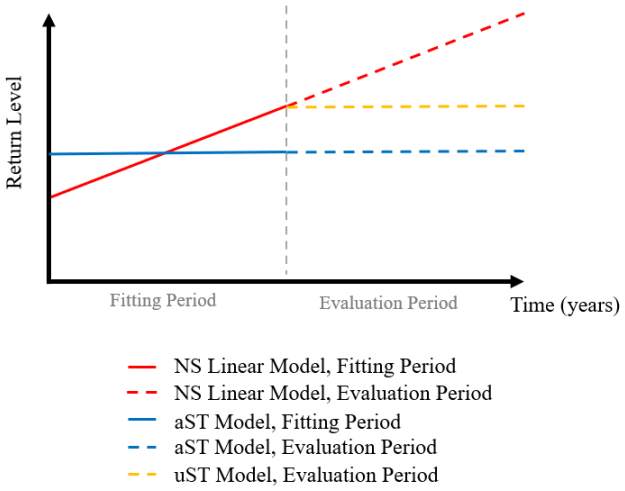


Figure 9: Nonstationary Modelling Techniques

2.7. COMBINING PARTITIONED EVA

Palutikof et al. (1999) describe seasonality as the seasonal variation in underlying meteorological drivers of extremes. Partitioning data can be useful when incorporating seasonality in an EVA and can result in more accurate estimates because separating the data results in simpler datasets with less meteorological drivers. For example, extremes from different seasons can be driven by different meteorological mechanisms and belong to different populations. Therefore, splitting the analysis between two seasons allows the two populations to be first treated individually. In this research project, the impact of seasonality is studied by splitting each year into summer and winter months and obtaining summer and winter maxima; two maxima are selected for each year. The combined distribution of summer and winter events can be determined using the following equation (Dullaart et al., 2021).

$$RP_{combined} = \frac{1}{\frac{1}{RP_{summer}} + \frac{1}{RP_{winter}}} \quad (2\text{-}16)$$

Where RP is the combined return period, RP_{summer} is the return period of summer events, and RP_{winter} is the return period of winter events.

When using this methodology to combine partitioned analyses it is important to distinguish between the frequency of exceedance and the probability of exceedance discussed in Section 2.1. For less extreme events, the RP and RP of annual extremes will differ. A RP less than 1 will result in a frequency greater than 1, indicating the event may occur on average more than once per year. However, the RP of annual maxima does not go below 1 and results in a probability of exceedance between 0 and 1.

3. CASE STUDY

3.1. THE MEUSE RIVER

The Meuse basin covers approximately 34,550 km^2 in France, Belgium, Germany, Luxembourg, and the Netherlands. It has a mean annual discharge of approximately 10 km^3/year (Descy et al., 2022). The river extends 874 kilometers from the source in France (Berger, 1992).

The catchment is split into two regions as shown in Figure 10. Ardennes covers a steeper, rocky region of the Meuse within Belgium between Chooz, France and Borgharen, Netherlands. Lorraine covers the broad and gently sloping river valley upstream of Chooz, France. It consists of more porous soils which allows for more infiltration. Therefore, Lorraine has a much slower response to rainfall compared to Ardennes. As a result, rainfall events typically lead to two distinct peaks in discharge: the first from the runoff produced in the Ardennes and the second from the slow response of Lorraine.

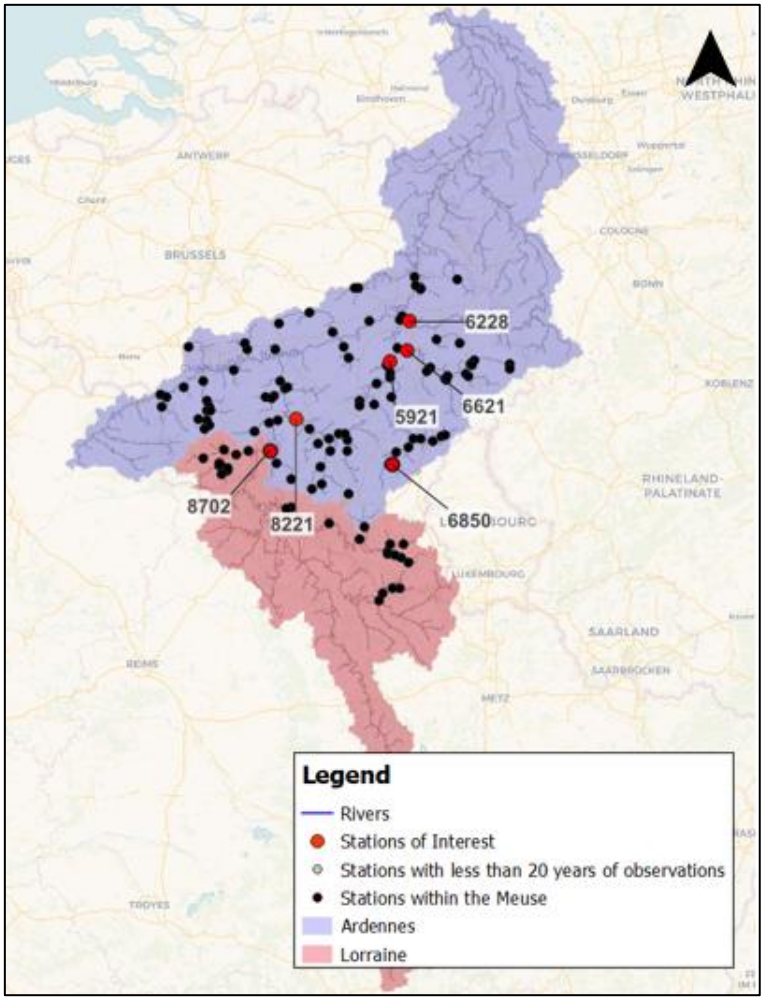


Figure 10: Meuse Catchment

The main tributaries of the Meuse are Chiers, Viroin, Semois, Lesse, Sambre, Ourthe, Roer, Niers, and Dieze. The longitudinal profile of the main tributaries is shown in Figure 11. The

Ourthe, Vesdre, and Ambleve have very steep slopes and the Belgian Meuse is also relatively steep. As shown in Figure 12, the Ardennes, shown in the middle of the figure, is at a much higher elevation than the rest of the catchment resulting in higher annual precipitation. High precipitation and the impermeability of the soil results in a high flood wave contributed by the Ardennes region of the Meuse (Berger, 1992).

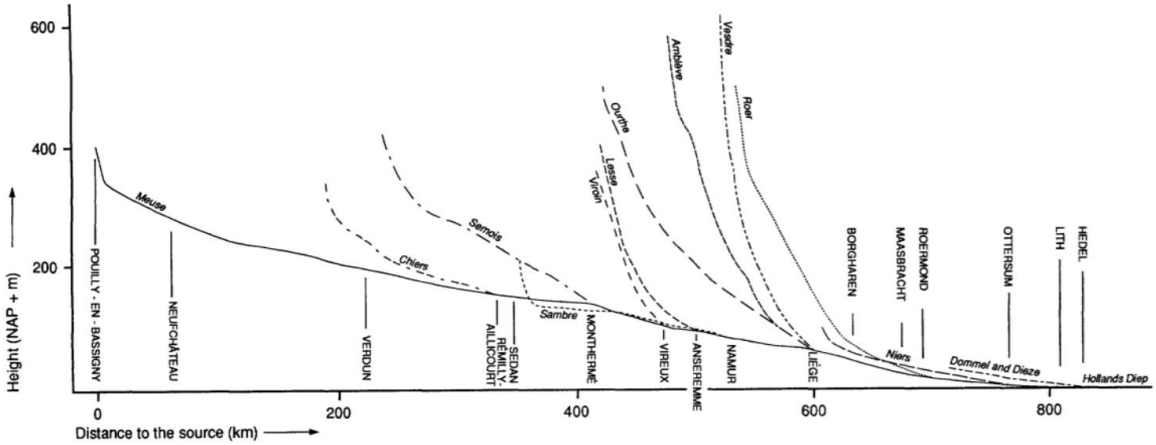


Figure 11: Longitudinal Profile of the Meuse (Berger, 1992)

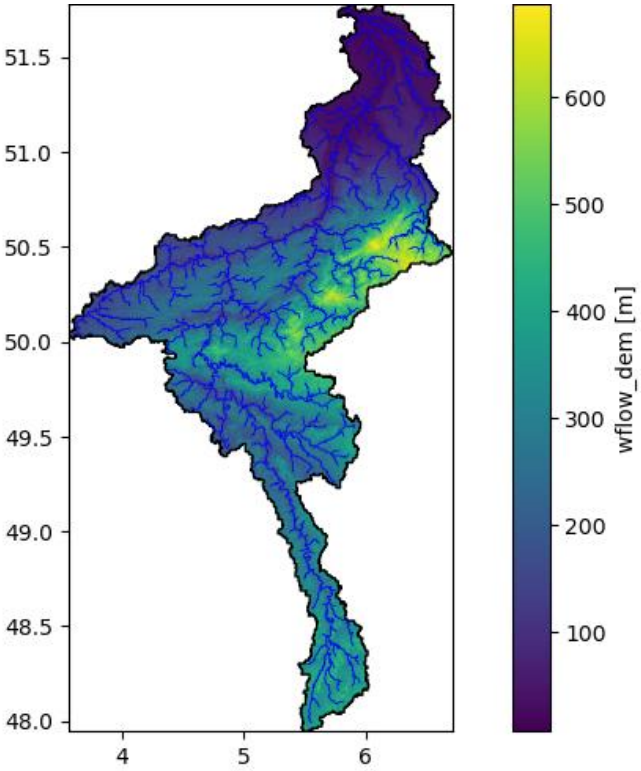


Figure 12: Meuse Elevation Map

3.1.1. LAND USE

The land use of the Meuse basin is shown in Table 1. Out of the 34,550 km^2 of the Meuse basin, the major uses are arable, forest, and pasture which account for 38.7, 28.7, and 18.1 percent of the total area.

Table 1: Land Use of the Meuse River Basin (Descy et al., 2022)

Land Use Category	Percentage of Catchment
Urban	11.4
Arable	38.7
Pasture	18.1
Forest	28.7
Natural Grassland	1.7
Sparse Vegetation	0.1
Wetland	0.4
Freshwater Bodies	0.9
Protected Area	0.2

3.2. DATA

RACMO provides synthetic precipitation and temperature data that are input into the hydrologic model. The hydrologic model provides the discharge time series that is used to obtain extreme value discharges used in the EVA.

In this study, up to 55 years of hourly discharge observations collected from 192 stations throughout the Meuse are available. In addition, 65 years for each of the 16 members, which represent 16 plausible initial climate conditions, result in 1,040 years of daily synthetic discharge. Average daily discharge is obtained so that observed and modelled results can be compared. Reported discharge obtained from the two-parameter lognormal distribution for the 50-, 100-, and 1000-year RP, provided by Service Public de Wallonie (SPW), is also available for comparison.

Hourly rainfall data is available at 68 locations throughout the Meuse with the earliest record on January 1, 1983, and the latest record on December 31, 2021.

3.2.1. DATA CURATION

Before determining the locations of interest for this study, careful data curation was performed to remove discharge stations with short records or those heavily impacted by humans such as canals or stations just upstream or downstream of hydraulic structures. Stations with a minimum record length of 20 years are chosen to maximize the number of stations included while excluding stations with short records. As shown in Figure 13, 140 out of the 192 stations in the Meuse have record lengths longer than 20 years. A summary of the justifications for discharge stations removed from this study is provided in Table 16. Further details are provided in Appendix B.

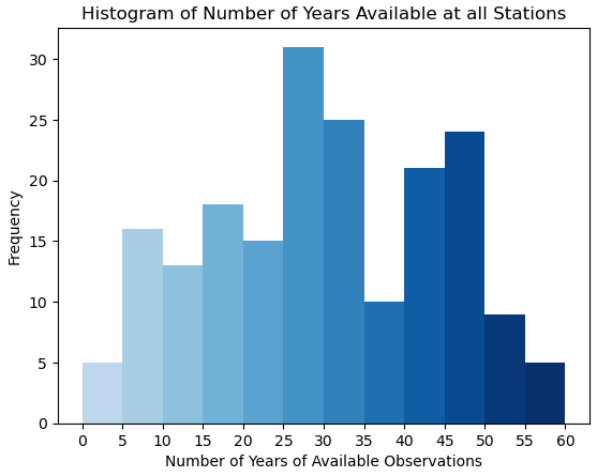


Figure 13 : Histogram of Number of Years Available at all Stations

Table 2: Summary of Stations (with more than 20 years) Removed from Study

Station ID	Reason for Removing
5771	Irregular flow pattern, Located along canal
7132	Flow influenced by nearby hydraulic structure
7831	Flow influenced by nearby hydraulic structure
8017	Flow influenced by nearby hydraulic structure
8022	Flow influenced by nearby hydraulic structure
9214	Flow influenced by nearby hydraulic structure
6220	Irregular flow pattern
6340	Irregular flow pattern
6440	Irregular flow pattern
5820	Change in base flow

Out of the remaining stations, three did not include information regarding the upstream subcatchment area. The Digital Elevation Model (DEM) for the Meuse was exported as a raster from the wflow model and the subcatchment area was calculated in Quantum Geographic Information Software (QGIS). The DEM was converted from a geographic coordinate system to a projected coordinate system, UTM zone 32N, with units of meters. The resulting area for each of these stations is included in Table 3. Although these stations are not the stations of interest that will be introduced in Section 3.3, obtaining the area is necessary to calculate the normalized Sen's slope which will be calculated at stations in the Meuse with more than 20 years of observations and a trend in AM observed discharge.

Table 3: Subcatchment Areas

Station ID	Subcatchment Area
5451	20,530 km^2
5904	2,760 km^2
5436	20,720 km^2

3.3. LOCATIONS OF INTEREST

Stations 5921, 8221, 6228, and 6621 are outlets of major tributaries in the Ardennes and are of interest to the Belgian authorities because if it rains in these catchments water will quickly reach the Netherlands; some of these stations were strongly impacted by the July 2021 flood.

In addition to these four stations, stations 6850 and 8702 with the highest increasing and decreasing trends in observed AM discharge, respectively, will also be analyzed. The stations of interest are shown in Figure 10.

Table 4: Stations of Interest

ID	River	Name	Subcatchment Area	Observed Record Length
5921	Ourthe Moyenne	Tabreux	1610 km^2	54
8221	Lesse	Gendron	1290 km^2	55
6228	Vesdre	Chaudfontaine	680 km^2	56
6621	Ambleve	Martinrive	1070 km^2	49
6850	Ruisseau de Laval	Sprimont	70 km^2	20
8702	Haute Meuse	Chooz	10,120 km^2	38

4. NONSTATIONARITY OF OBSERVED DISCHARGE

Currently, the meteorological time series generated from RACMO is bias corrected so the time series input into the hydrological model is stationary. No land use changes are modelled in the hydrological model; therefore, the generated modelled discharges are stationary. However, rainfall intensities and extreme river discharges are expected to increase in Northern Europe as a result of climate change (Diermanse et al., 2010). In addition, Milly et al. (2008) argue that as a result of changing climate and land use, the use of time-invariant probability distributions should not be the default assumption when estimating flood risk.

However, application of NS methods should not be solely based on results of statistical trend tests applied to relatively short historical records. Serinaldi et al. (2018) state that the results of these statistical tests should be interpreted carefully and should not be solely relied on to infer the persistence of nonstationarity. Instead, they emphasize the importance of studying the underlying physical mechanisms before applying NS methods. Even if a trend is identified there is still uncertainty on whether it will persist and how it will persist. For example, if an identified trend is indeed present it is still uncertain whether it will continue linearly, exponentially, or continue at all!

In this chapter the following research questions are addressed:

- *How much impact can a very extreme event, such as what occurred in July 2021 in central Europe, have on the statistics of extreme discharge?*
- *To what extent do nonstationary statistical methods impact discharge estimates when a trend in observed discharge is identified?*
- *How does the confidence of discharges estimated from various simulations of the current climate compare to the confidence of estimates from a nonstationary extreme value analysis?*

Before performing an EVA on observed discharge, it is necessary to identify whether nonstationarity is present. Therefore, trends in AM observed discharge and corresponding rainfall events are identified and estimated in Section 4.1. In Section 4.2, the change in return values estimated from observed discharge is evaluated to study how much statistics have changed in the last eight years. The predictive capabilities of the aST, uST, and NS models are compared in Section 4.3. Lastly, in Section 4.4, results of a stationary EVA and NEVA on observed discharge are compared to results of stationary EVA on 16 synthetic time series with the same length as observations.

4.1. TRENDS IN ANNUAL MAXIMA DISCHARGE AND RAINFALL

As previously mentioned, modelled discharge is stationary, however, before performing an EVA on observations, it is necessary to test the data for nonstationarity. A time series is considered nonstationary if it contains trends, shifts, or cyclicity. In this section the MK test is used to identify trends in AM discharge and corresponding rainfall events.

4.1.1. DATA AND METHODOLOGY

The MK test, introduced in Section 2.5.1, is a hypothesis-based test that will be used to identify trends in AM discharge throughout the Meuse and determine with what confidence there is a trend. Sen's slope is calculated to estimate the magnitude of observed trends. To compare the magnitude of trends across the Meuse, Sen's slope is calculated after normalizing the discharge by dividing by the subcatchment area and is converted to mm/day; in this report this is referred to as the normalized Sen's slope.

To determine if observed trends in AM discharge result from trends in rainfall, the MK test is used to identify trends in 1-day, 3-day, and 5-day cumulative rainfall for subcatchments upstream of stations with trends in AM discharge. Thiessen polygons are used to obtain the weighted average rainfall for each subcatchment. To ensure the rainfall contributing to extreme discharge is obtained, the maximum rainfall between the day corresponding to the AM discharge and the day prior to the AM discharge is selected. A summary of the methodology used to identify statistically significant trends in observed AM discharge and corresponding rainfall events is provided in Figure 14.

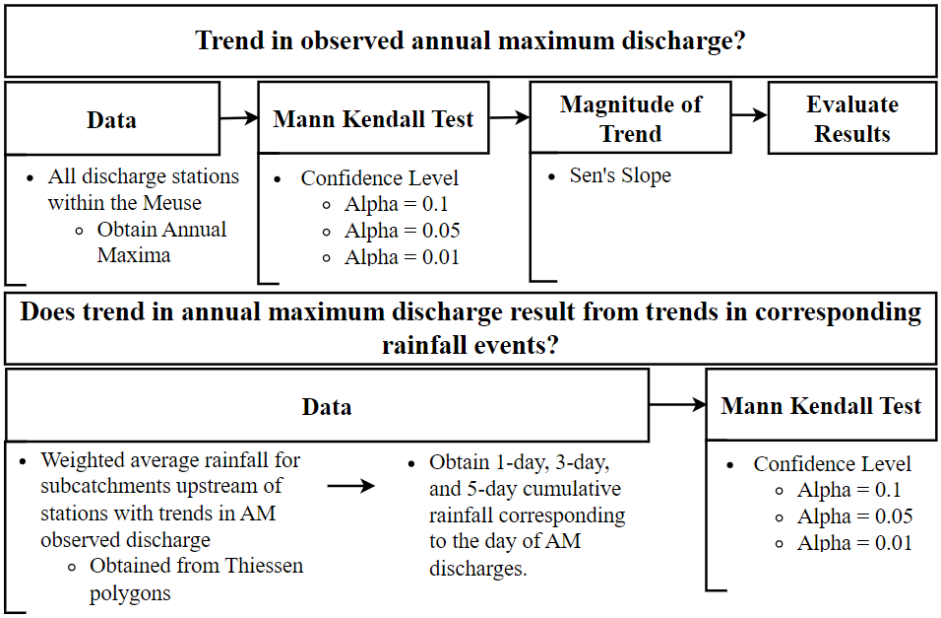


Figure 14: Trends in AM Discharge and Corresponding Rainfall Events Methodology

4.1.2. RESULTS

4.1.2.1. TREND IN ANNUAL MAXIMUM OBSERVED DISCHARGE

Results are presented in Figure 15. Stations with increasing trends are shown in blue and stations with decreasing trends are shown in red. A larger symbol indicates a larger normalized Sen's slope, larger trend in AM discharge, and a darker shade indicates increasing confidence in the identified trend.

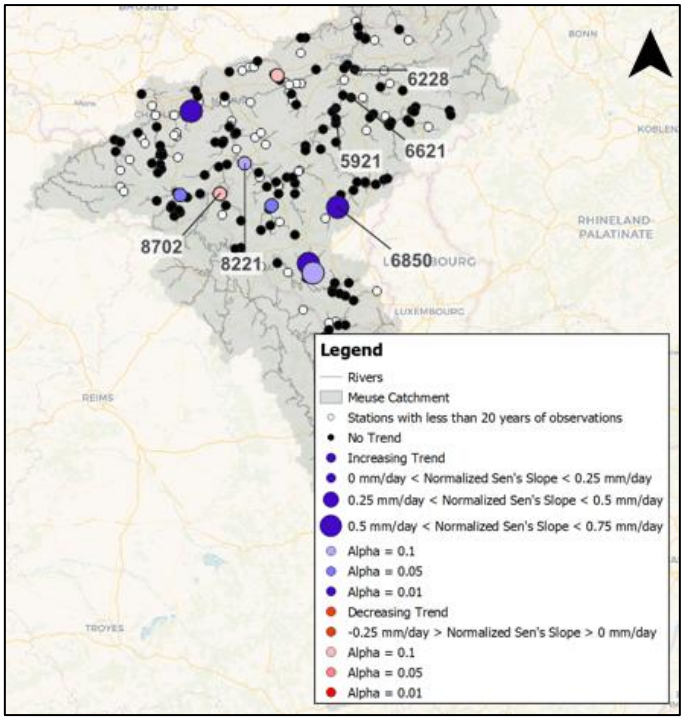


Figure 15: Trend in AM Observed Discharge (The shade of the symbol indicates the confidence of the trend determined from the Mann Kendall test, the size of the symbol indicates the magnitude of the trend determined by the Normalized Sen's Slope)

MK indicated with 90 percent confidence that 9 of the 130 stations exhibited a trend in AM discharge. As shown in Figure 15, more stations exhibit increasing trends than decreasing trends which can be attributed to effects of climate and land use change. Decreasing trends in AM discharge may be due to flood control measures, climate change, or long-term water storage (Slater et al., 2021).

MK and Sen's slope results for observed discharge are included in Appendix B.1. The time series with the largest increasing and largest decreasing trend in AM discharge are shown in Figure 16 and Figure 17, respectively. Due to the several data gaps at the beginning of the historical record at Station 8702, normalized Sen's slope is also calculated from data after the last data gap in 1990.

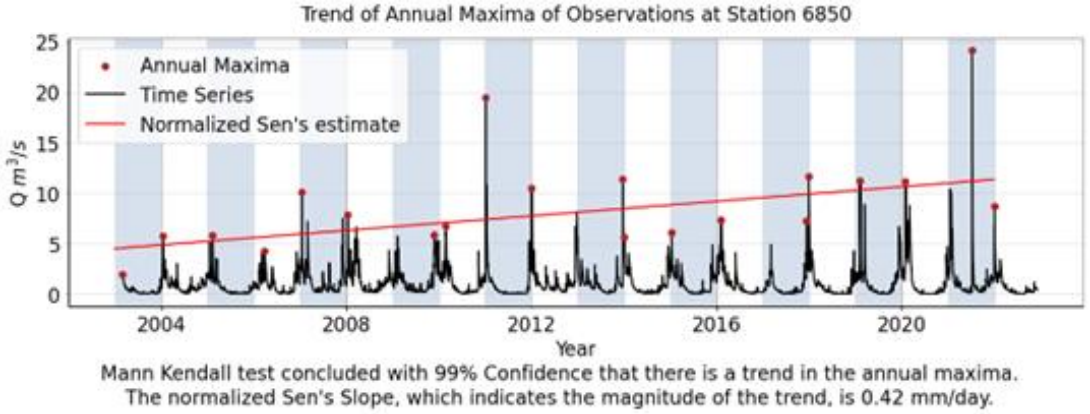


Figure 16: Station with Largest Increasing Trend in AM Discharge

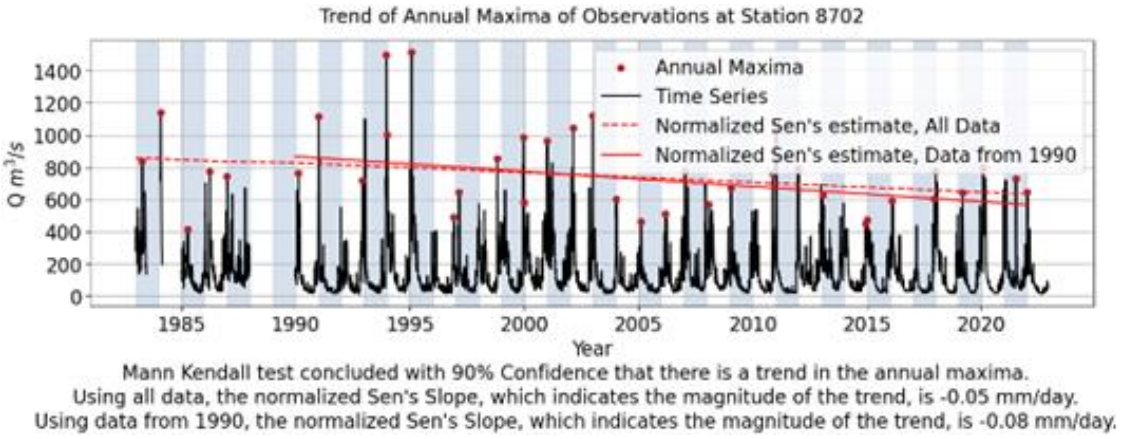


Figure 17: Station with Largest Decreasing Trend in AM Discharge

4.1.2.2. TRENDS IN OBSERVED RAINFALL

To evaluate whether trends in AM discharges result from increase in precipitation due to climate change, MK is used to identify possible trends in observed rainfall.

As a preliminary analysis, AM 1-day rainfall at each rain gauge is studied. As shown in Figure 18, the most common season for AM discharges and AM rainfall within the Meuse do not correspond. Most AM discharges occur in the winter while most AM 1-day rainfall events occur in the summer. Therefore, rainfall corresponding to the same day as AM observed discharges will be selected to analyze trends in observed rainfall. To ensure rainfall contributing to extreme discharge is obtained, maximum rainfall between the day corresponding to the AM discharge and the day prior will be compared and the highest value will be selected for this analysis.

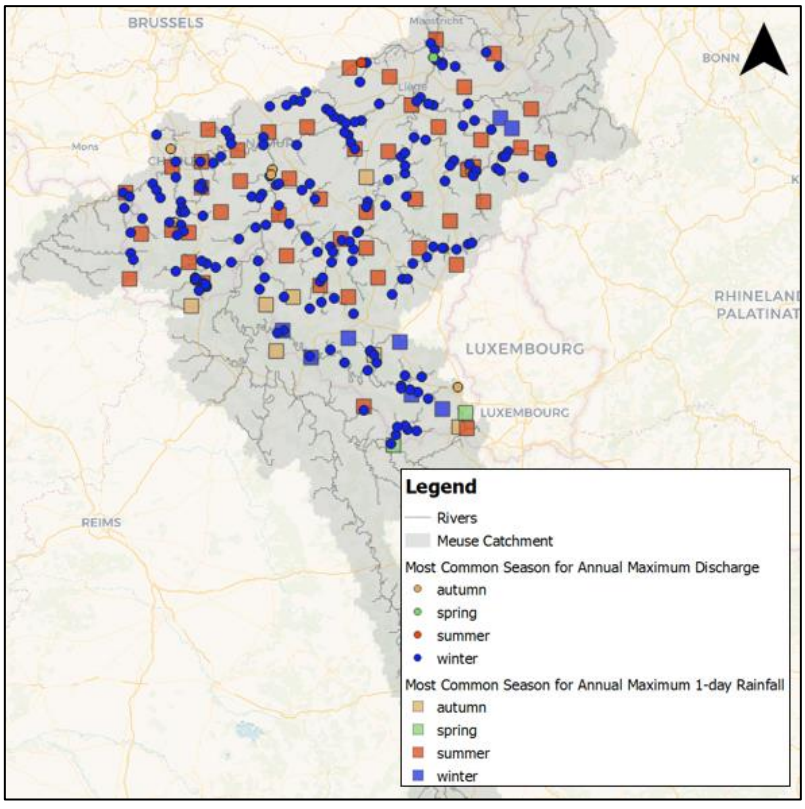


Figure 18: Most Common Season for AM Discharge and AM 1-day Cumulative Rainfall at all Stations

Weighted average rainfall for subcatchments upstream of stations with trends in AM discharge is obtained using Thiessen polygons. Thiessen polygons are created for each timestep so that the weighted average only includes rain gauges with data available for that timestep. An example of a Thiessen polygon produced for one timestep is shown in Figure 19. After obtaining the rainfall for the Thiessen polygons the weighted average rainfall is obtained for each subcatchment upstream of stations with trends in AM discharge.

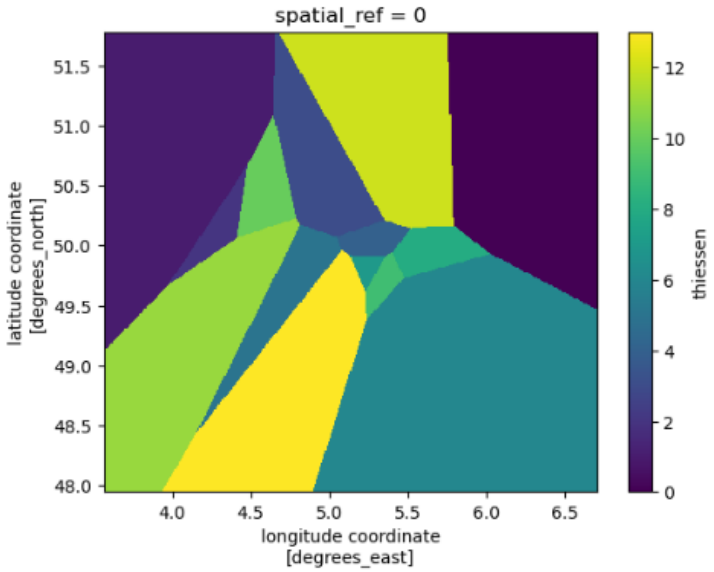


Figure 19: Example of Thiessen Polygon for One Timestep

Te Booij (2022) performed trend analysis on 1-day, 3-day, and 5-day rainfall within the Meuse, therefore, it was decided to use the same rainfall durations and put more emphasis on results depending on the sizes of the upstream subcatchments. The hydrologic response in small subcatchments is shorter and more impacted by short duration rainfall events, while the response in large subcatchments is longer and more impacted by long duration rainfall events. Therefore, 1-day rainfall is emphasized for small subcatchments (less than 500 km^2), 3-day rainfall is emphasized for medium sized subcatchments (between 500 km^2 and 2,000 km^2), and 5-day rainfall is emphasized for large subcatchments (larger than 2,000 km^2).

A summary of MK results for AM discharge and corresponding rainfall events is provided in Table 5. Results are sorted by catchment size with the station with the largest subcatchment area at the top. An emphasis on rainfall trend results for the various rainfall durations based on subcatchment size, as previously described, is shown in bold.

From Table 5, it can be seen that not all stations with trends in AM discharge have statistically significant trends in corresponding rainfall events at the evaluated confidence levels. However, this does not mean that trends do not exist but that they could not be verified at the specified confidence levels with available observations. Overall, the direction of statistically significant trends, increasing versus decreasing, for observed rainfall and discharge correspond which indicates that rainfall likely contributes to trends in discharge at those stations. It should also be emphasized that the smallest rainfall duration considered in this analysis was 1-day rainfall which may be too long of a duration for subcatchments with very small areas. If a smaller rainfall duration was applied for stations with smaller subcatchment areas, more statistically significant trends may have been found.

Table 5: Trends in AM Discharge and Corresponding Rainfall Events

StationID	Subcatchment Area (km^2)	Average Discharge (m^3/s)	Average AM Discharge (m^3/s)	Trend in AM Discharge			Trend in Max Rainfall corresponding to AM Discharge Events		
				Trend	Confidence	Sen's Slope (mm/day)	Trend, 1-day	Trend, 3-day	Trend, 5-day
8702	9633	150	780	decreasing	90%	-0.08	99% decreasing	90% decreasing	**No Trend**
8221	1320	20	160	increasing	95%	0.07	No Trend	**No Trend**	No Trend
9081	453	5	35	increasing	90%	0.07	**No Trend**	No Trend	No Trend
7244	306	2	15	decreasing	95%	-0.04	**99% decreasing**	95% decreasing	No Trend
8341	302	5	45	increasing	99%	0.12	**No Trend**	No Trend	No Trend
7140	190	5	30	increasing	90%	0.46	**No Trend**	No Trend	No Trend
6980	190	1	10	increasing	99%	0.46	**No Trend**	No Trend	No Trend
6850	91	1	10	increasing	95%	0.60	**90% increasing**	99% increasing	95% increasing
6930	57	1	5	increasing	99%	0.69	**No Trend**	99% increasing	No Trend

4.2. OBSERVED CHANGES IN RIVER DISCHARGE SINCE 2015

4.2.1. DATA AND METHODOLOGY

As previously mentioned, the stochastic weather generator used in GRADE is limited to resampling from observations and is, therefore, not capable of producing daily rainfall exceeding the range of historical data. However, since the implementation of GRADE in 2015, events like what occurred in July 2021 add additional information about the statistics of extreme discharges. To evaluate the change in estimated discharges since 2015, observed AM are fit to Gumbel using MLE for data up until 2015 through 2022. Therefore, eight sets of AM are fit to Gumbel for data through 2015, 2016, 2017, 2018, 2019, 2020, 2021, and 2022 at each station.

In addition, the change in GEV estimates is studied to emphasize how estimates differ between extreme value models. Station 6228 is chosen since it was the station of interest most impacted by the July 2021 event. This is a very extreme case; therefore, Station 8221 is selected for comparison.

The methodology for this analysis is summarized in Figure 20.

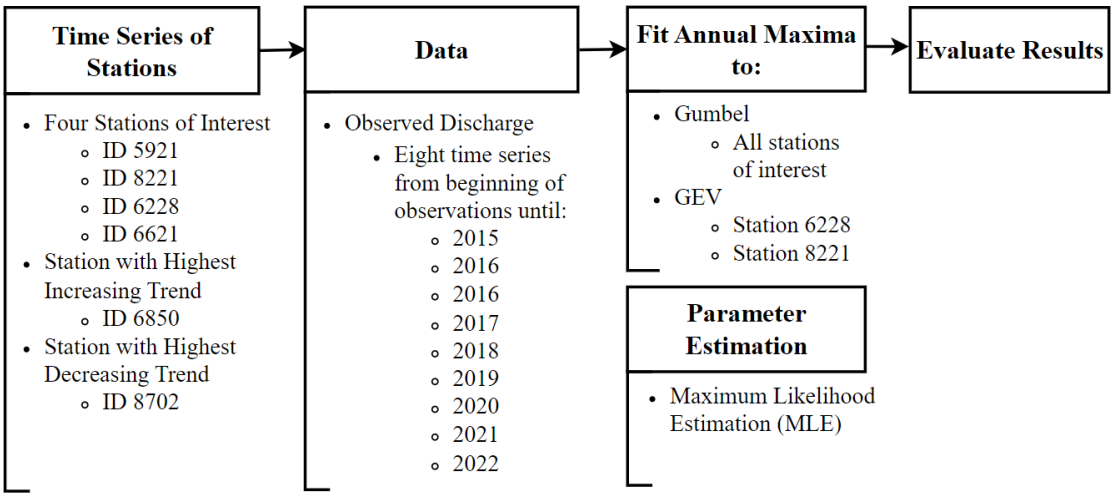


Figure 20: Observed Changes in River Discharge Since 2015

4.2.2. RESULTS

Observed changes in Gumbel estimates for Station 5921 are presented in Figure 21; results for all stations of interest are included in Appendix B.1. The eight sets of AM fit to Gumbel

and corresponding AM are shown on the left side of Figure 21. For example, the curve using data until 2015 and the 2015 AM are shown in the lightest blue and the curve using data until 2022 and the 2022 AM are shown in the darkest blue. The 2021 event is shown in red.

The change in the 10-year RP from 2015 to 2022 is shown on the right side of Figure 21. The confidence interval shown in blue is created by only resampling values before 2020 and the confidence interval shown in red is created by resampling data until 2022. This is done to highlight the impact the July 2021 event had on the confidence of estimates.

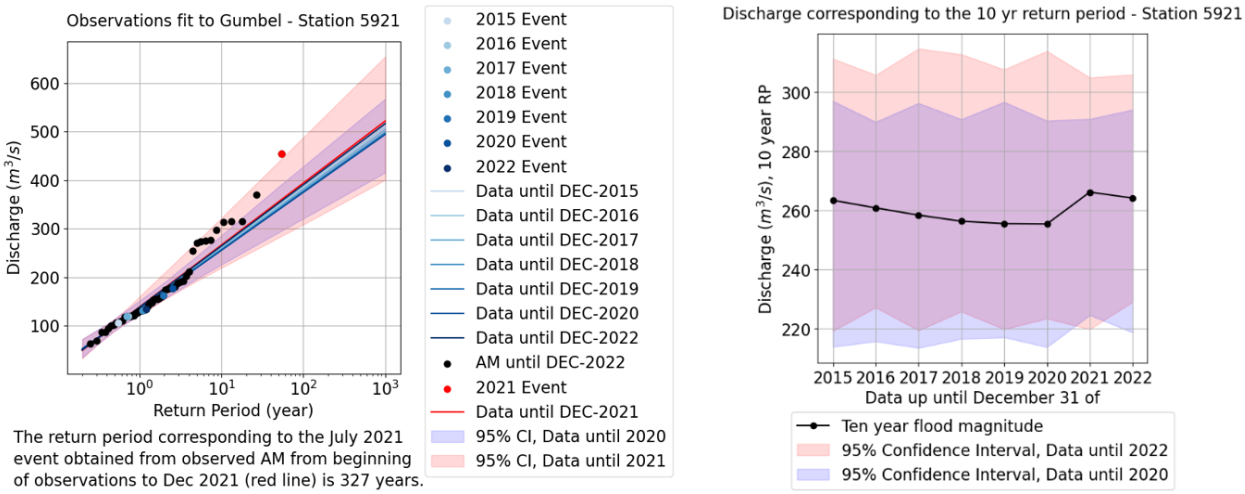


Figure 21: Observed Changes in Gumbel Estimates Since 2015 - Station 5921

GEV curves for Stations 6228 and 8221 are presented on the left and right of Figure 22, respectively.

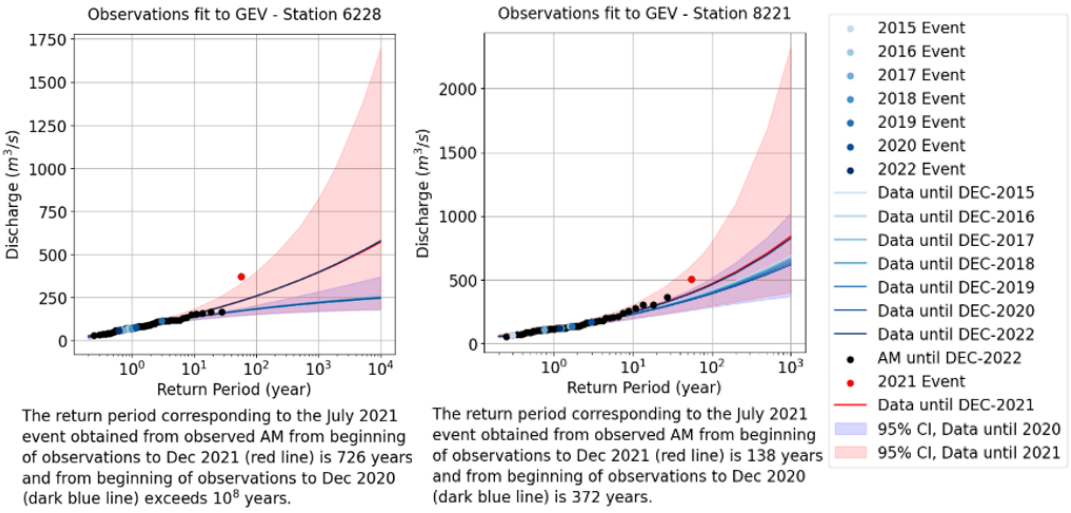


Figure 22: Discharge Frequency Curve for GEV (Left: Station 6228, Right: Station 8221)

A summary of the impact the 2021 event had at each station is presented in Table 6. The ratio of the 100-year discharge estimated by fitting Gumbel and GEV before and after the 2021 event, $Q_{100,\ post2021}/Q_{100,\ pre2021}$ are shown to highlight the impact this extreme event had on estimated discharges. Additionally, the last column of this table presents the ratio between the magnitude of the 2021 event and the previously recorded highest AM discharge. The 2021 event exceeded all previously recorded AM at all stations except Station 8702.

Table 6: Impact of 2021 Event at Stations of Interest

Station	Record Length	Gumbel $Q_{100,\ post2021}/Q_{100,\ pre2021}$	GEV $Q_{100,\ post2021}/Q_{100,\ pre2021}$	2021 Event was XX times higher than previously recorded highest AM
5921	54	1.05	1.13	1.2
8221	55	1.07	1.20	1.4
6228	56	1.08	1.20	2.2
6621	49	1.08	1.22	1.4
6850	20	1.10	1.35	1.3
8702 [1]	38	0.99	0.99	--

Note:
1. 2021 Event did not exceed the previously recorded highest annual maximum discharge.

There are three main findings from these results:

- Short historical records make it difficult to obtain reliable estimates of discharges.
- Selection of the extreme value model impacts estimated discharges and RPs. This concept will be further explored in Section 5.1.
- Addition of one extreme event can significantly impact estimated discharges obtained from observations.

First, results illustrate that estimated quantiles vary year to year due to limited historical data. This is seen by the change in discharge frequency curves fit using data until 2015 through 2022 at each station. Additionally, as shown in Table 6, the ratio between estimated 100-year discharges before and after the 2021 event tends to increase when less data is available. For example, Station 6850 has the shortest record length, 20 years, and has the largest ratios between estimated 100-year discharges before and after the 2021 event indicating there is a larger difference between $Q_{100,\ post2021}$ and $Q_{100,\ pre2021}$. This highlights the difficulty of obtaining robust estimates from limited historical records. However, many other factors, other than record length, impact these ratios including various modelling assumptions and the magnitude of the 2021 event relative to historical AM.

Second, comparison between the RP corresponding to the 2021 event for Gumbel and GEV demonstrate the extreme value model can significantly impact RPs. As mentioned in Figure 22, the RP corresponding to the 2021 event estimated fitting data until the end of 2021 to GEV are 726 years and 138 years at stations 6228 and 8221, respectively. When fit to Gumbel the RPs are 9,232 years and 2,683 years at stations 6228 and 8221, respectively. The estimated RPs using Gumbel are 13 and 19 times as large as those estimated using GEV. This demonstrates how difficult it can be to estimate very extreme events and how different models can result in significantly different results. The impact of various extreme value models is further explored in Section 5.1.

Third, adding one very extreme event, like what occurred in 2021, can have a very large impact on estimated discharges. The ratio of the estimated 100-year discharges before and after the 2021 event indicate the 100-year Gumbel (GEV) estimated discharges increased between factors of 1.05 and 1.10 (1.13 and 1.35) at the five stations of interest heavily impacted by the July 2021 event. While the addition of the 2021 event impacted discharges

estimated from both Gumbel and GEV, the ratio of GEV estimates is likely higher due to the flexibility provided by the GEV shape parameter.

The impact the July 2021 event had on GEV estimates is also illustrated in Figure 22. At Station 6228 the GEV tail behavior changed from a GEV type III exhibiting a light tail behavior prior to the 2021 event to a GEV type II exhibiting heavy tail behavior after the 2021 event. From Table 6 it can be seen that the magnitude of the July 2021 event exceeded the previously recorded highest AM by a factor of 2.2 making it a particularly unique event at this station. This could suggest that the July 2021 event may come from a separate distribution of rare, extreme events (Ludwig et al., 2023).

As mentioned at the bottom of Figure 22, at Station 6228 the 2021 event had a RP of 726 years and 10^8 years when including and excluding the 2021 event, respectively. RP of 10^8 years seems unrealistic since removing only the highest AM, the July 2021 event, would result in RP over 1.38^5 (10^8 years/726 years) times lower than if it were included.

Station 8221 was also heavily impacted by the July 2021 event with a magnitude of 1.4 times the previously recorded highest AM. However, as shown in Figure 22 this event did not impact the limiting type of the GEV distribution. All curves exhibit heavy tailed behavior. When the 2021 event was excluded from the analysis, the RP of this event was estimated to be over 2.7 (372 years/138 years) times as high as when it was included.

4.3. COMPARISON OF PREDICTIVE CAPABILITY OF NEVA METHODS

Sections 4.1 and 4.2 identified trends in observed AM discharge and demonstrated the vulnerability of applying EVA to short historical records with rare events. While many authors, including Milly et al. (2008), argue that stationarity should not be the default assumption when estimating flood risk due to the changing climate, results of statistical trend tests on short historical records should be interpreted carefully and not be solely relied on to infer the persistence of nonstationarity. Therefore, in this section, predictive capabilities of the nonstationary linear (NS), approximated stationary (aST), and updated stationary (uST) models are compared to evaluate their performance in estimating discharges from observed data.

4.3.1. DATA AND METHODOLOGY

Before comparing the predictive capabilities of the nonstationary models, the evolution of estimated parameters and discharges are studied using the NS linear model. This approach assumes the change in distribution parameters can be modelled as a linear function in time using the equations introduced in Section 2.6.1. Gumbel is used for this analysis because extreme discharge at Borgharen, where the Meuse enters the Netherlands, is closely modelled by Gumbel. Gumbel parameters are estimated from observations using MLE with sliding windows. A sliding window length of 30 years is chosen, which covers approximately one climate period, so that the data within each window can be considered stationary (Mudersbach & Jensen, 2010). Therefore, it is assumed that data within each 30-year window is stationary. Stations with a trend in AM discharge and at least 45 years of observed data are considered for this analysis such that the evolution of distribution parameters and estimated

return levels can be studied over a sufficient period. Out of the nine stations with a trend, four stations have at least 45 years of observed discharge and are considered for this analysis.

Station 8221 is chosen to evaluate the predictive capabilities of the nonstationary models described in Section 2.6 because it has the longest historical record of the stations with trends in observed AM discharge with data available from 1968 through 2021. To evaluate the predictive capabilities of each NS method only data up until 2009 are used to fit the Gumbel parameters of the linear NS model using a moving window of 30 years. This will result in 25 estimates of parameters and return levels for Station 8221. The period from 1968 through 2009 will be referred to as the fitting period and the period from 2009 through 2021 will be referred to as the evaluation period.

A summary of the methodology used to this analysis is presented in Figure 23.

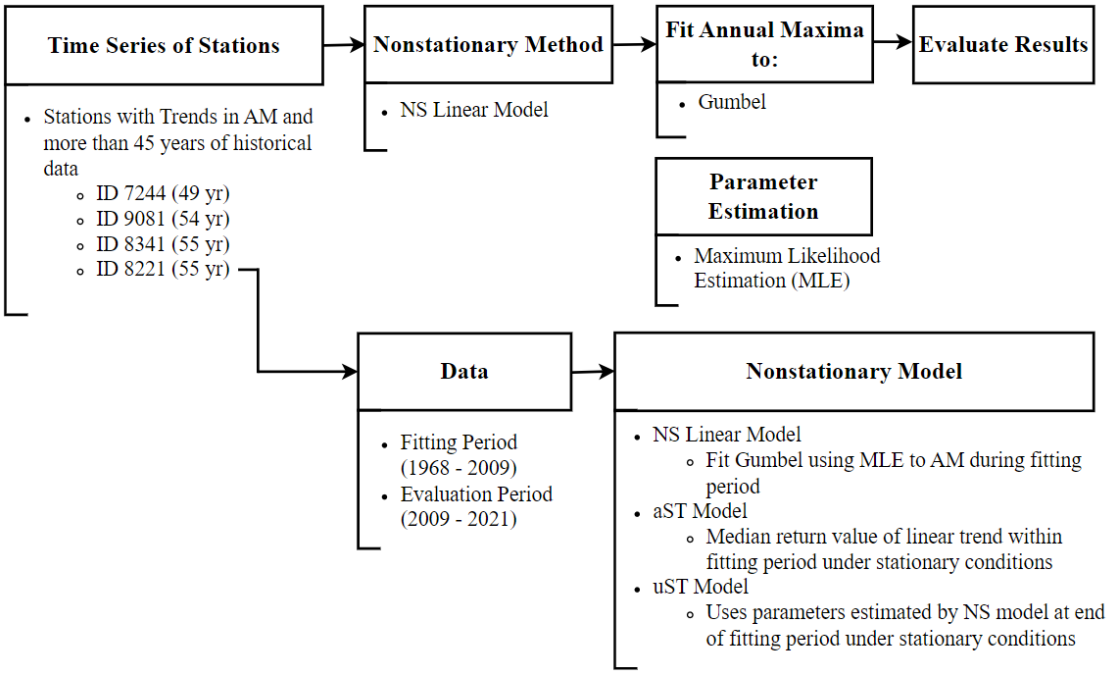


Figure 23: Methodology to Evaluate Predictive Capability of Nonstationary Methods

4.3.2. RESULTS

Gumbel parameters and return levels estimated using the linear model for Station 8221, Lesse at Gendron, are presented in Figure 24 and Figure 25. Results for the other three stations are included in Appendix B.3. There are 55 years of observations available at Station 8221, therefore, using a sliding window of 30 years results in 25 estimates of the Gumbel parameters and discharges, shown in black. A linear trendline is fit to the estimated parameters and return levels and is show in red. The dotted black lines indicate the 95 percent confidence interval of the first estimate; for Station 8221 this includes the 30 years from the beginning of observations in 1968 until 1997. The confidence interval is created by resampling the observed annual maxima between 1968 and 1997 100 times and calculating the 2.5^{th} and 97.5^{th} percentile of the estimates. For simplicity, when describing results obtained using the sliding windows the last year of the sliding window is mentioned when discussing results. For example, when discussing results obtained using the sliding window from 1968 until 1997, ‘estimate in 1997’ is mentioned.

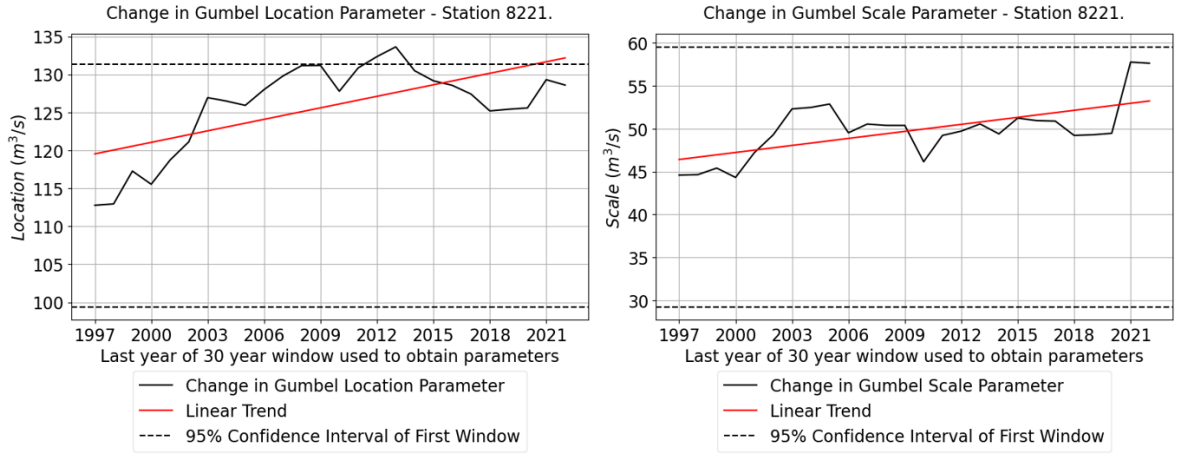


Figure 24: Evolution of Gumbel Parameters - Station 8221 using 30-year Sliding Window

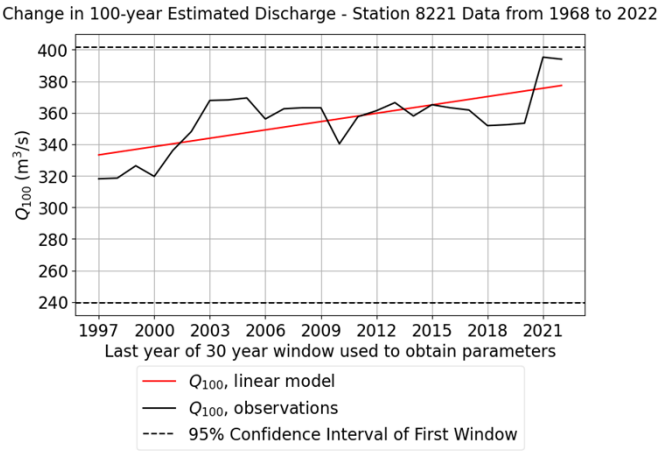


Figure 25: Evolution of Estimated 100-year Discharge - Station 8221 using 30-year Sliding Window

As shown in Figure 24, an increasing linear trendline is fit to the estimated Gumbel location parameter which increases 15 m^3/s throughout the fitting period. The location parameter describes the location or shift in the distribution; therefore, it is expected that there is an increasing tendency in the location parameter at stations with an increasing trend in AM discharge. An increasing linear trendline is also fit to the Gumbel scale parameter. The scale parameter describes the spread of the distribution; therefore, an increase in the scale parameter indicates increased variability in the extremes used to obtain the parameters. Only 30-years is considered to obtain each estimate, because 30-years covers approximately one climate period. Therefore, there is some variability of each estimate shown in the solid black line. Although a linear trendline can be fit to the estimates, it may not be the best model to characterize this relationship.

Gumbel parameters and estimated discharges at all stations, except for Station 9081 increase (decrease) during the fitting period where increasing (decreasing) trends in AM are identified. At Station 9081 the MK test indicated with 95% confidence that there was an increasing trend in AM and the Sen's Slope estimated the magnitude of the trend to be 0.07 mm/day. As shown in Appendix B.1, the largest recorded AM at this station was recorded in 1980 and is nearly twice as large as the majority of observed AM. Therefore, when the 1980 AM is included in the 30-year window used to estimate the Gumbel parameters, the scale parameter is high and relatively stable but starts to lower in 2010 when it is no longer included in the estimation. Similarly, when the 1980 AM is included in the 30-year window used to estimate

the location parameter, the location parameter increases and slowly flattens after 2010 when it is no longer included in the estimation. The estimated return level also slightly increases until 2010 and flattens out. This demonstrates the difficulties of fitting a linear model with very extreme events. Thirty-year windows are used to obtain the distribution parameters, because that covers approximately one climate period and can be considered stationary. However, as seen at Station 9081, very extreme events can have an impact on the NS linear model.

The estimated 100-year discharges for Station 8221 using the NS, aST, and uST approaches inferred from the fitting period are shown in Figure 26.

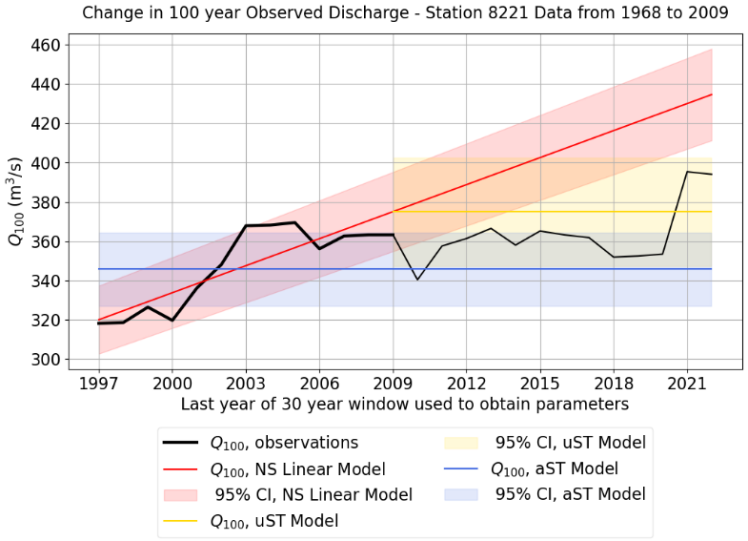


Figure 26: Change in 100-year Observed Discharge Estimated using the NS, aST, and uST models - Station 8221

AIC values for the models for the fitting period are provided in Table 7. The AIC values indicate there is little evidence that the aST model performs better than the NS and uST models; the difference between AIC values is less than five.

Table 7: NS and aST Comparison of Model Fit

Model	AIC
NS Linear Model	468
uST Model	468
aST Model	464

However, the three models vary significantly throughout the evaluation period. As shown in Table 8, Q_{100} estimated for 2022 from the NS Linear, uST, and aST models were 10 percent higher, 5 percent lower, and 12 percent lower, respectively, than those estimated from observations. When these models are extrapolated to the year 2050, the NS linear model results in 100-year discharges 1.5 and 1.6 times higher than those obtained from the uST and aST models, respectively. This demonstrates the challenges and risks associated with applying nonstationary models, especially when limited historical data is available.

Table 8: Comparison of Q100 for the Various NS Statistical Methods

Model	Percent Difference between $Q_{100, 2022}$ Estimated from Observations and the Three Nonstationary Statistical Methods[1]	$Q_{100, 2050}$ $(m^3/s)^2$
NS Linear Model	10%	563
uST Model	-5%	375
aST Model	-12%	346

Notes:
1. Positive percent difference indicates estimate obtained from nonstationary statistical method was higher than what was obtained from observations.
2. 100-year discharge extrapolated to the year 2050.

4.4. VARIABILITY OF CURRENT CLIMATE

In Section 4.1, the MK test concluded that nine stations in the Meuse have a trend in observed AM discharge. However, modelled discharge is stationary. Therefore, it is important to reflect on the assumption of stationarity in the generation of synthetic discharge when observed trends are identified.

KNMI provides 1,040 years of stationary meteorological data generated from the RACMO climate model using 16 climate scenarios, 65 years each, with different possible initial climate conditions. One reason for currently using stationary time series is the large amount of uncertainty in realizations of the current climate. In this section, the range of possible estimates for a NEVA is compared to the range of possible estimates estimated using the synthetic data generated from realizations of the current climate.

4.4.1. DATA AND METHODOLOGY

Stationary and nonstationary EVA on observations are performed and compared to a stationary EVA on 16 synthetic time series from each of the 16 modelled members. The methodology introduced in Section 2.6.1 is used to perform the NEVA on observed discharge. Observed AM are detrended under the assumption that nonstationarity can be described by a linear trend before performing a stationary EVA. Station 8221 is chosen for this analysis because it has the longest historical record of the stations with trends in observed AM discharge. A comparison of the uncertainty of estimates between the methodologies is quantified by comparing the generated confidence intervals. A summary of the methodology used for this analysis is presented in Figure 27.

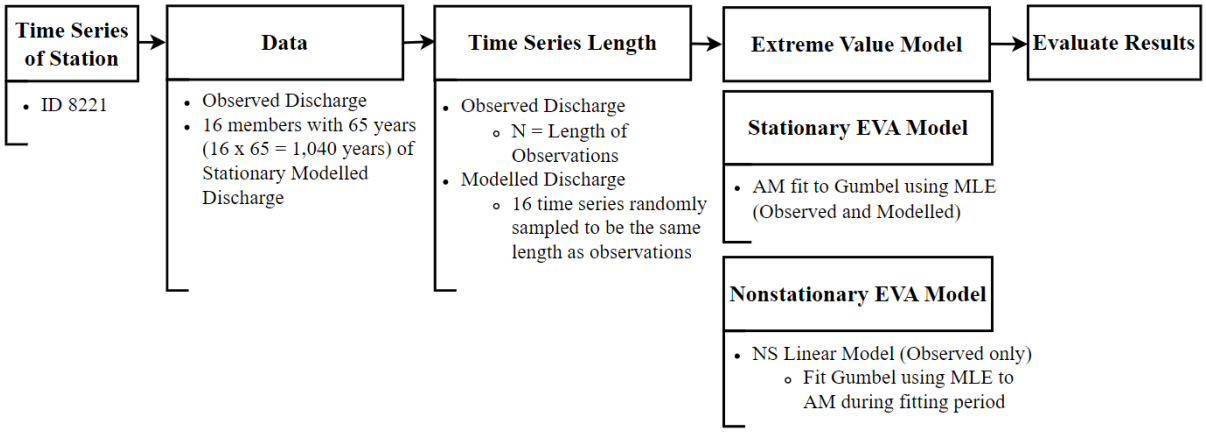


Figure 27: Methodology to Compare Uncertainty in Estimates using Nonstationary Methods to Uncertainty resulting from the Variability of Current Climate

4.4.2. RESULTS

Results are presented in Figure 28. Results of the stationary EVA of the 16 synthetic series randomly sampled from each of the 16 modelled members to be the same length as observations, 55 years, are shown in black. Results of the stationary EVA on the detrended AM are shown in red and results of the NEVA, performed using the methodology introduced in Section 2.6.1, are shown in blue.

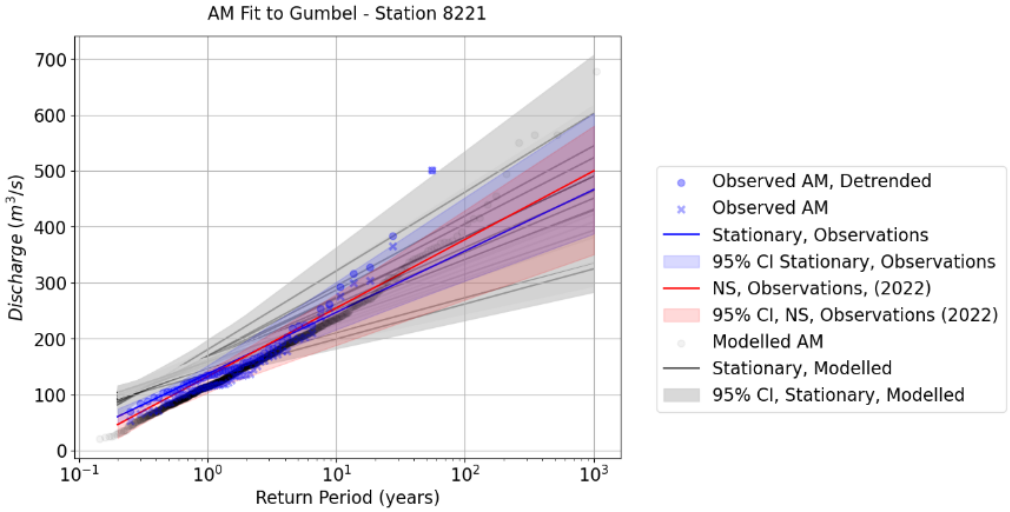


Figure 28: Stationary and Nonstationary EVA on Observations compared to Stationary EVA on 16 Time Series Randomly Sampled to be the same length as observations from the 65 years available at each of the 16 Members of Stationary Modelled Discharge

Results indicate that the large variability in climate simulations leads to more uncertainty in discharge estimates than the NEVA on observations. The confidence intervals, which indicate the range of possible estimates, is wider for the 16 synthetic series, shown in black, than the NEVA on observations, shown in red. It should be mentioned that the confidence interval of the NEVA does not account for the uncertainty resulting from the assumption of a linear trend in the Gumbel parameters. Therefore, there is additional uncertainty in the assumption of modelling a linear trend. Similarly, there is additional uncertainty in assumptions made in the climate and hydrological models not accounted for in the confidence intervals for the modelled data. The confidence interval for the 1,000-year RP for the stationary EVA of the 16 climate scenarios is still nearly twice as wide as the confidence interval generated for the nonstationary EVA of observations. This demonstrates the 16 members result in a wide range of discharge estimates when analyzed individually due to the variability in climate prediction and the shorter dataset.

The 16 synthetic series were resampled to be the same length as observations, 55 years, so results could be compared to results obtained from observations. Each of the 16 members have slightly different initial conditions to represent different realizations of the current climate. Therefore, the members were analyzed separately to demonstrate that estimates obtained from each member can vary due to the variability in simulations of the current climate. However, if the 16 members were stacked to obtain 1,040 years of modelled data the variability in the estimates would be much smaller. More extremes would be available to fit the distribution leading to more precise estimates. The impact of record length is further explored in Section 6.1.

5. EXTREME VALUE MODELS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

This chapter gives an answer to the research question "*How much influence do extreme value models have on discharge estimates and their uncertainty?*" To answer this question, discharge estimates and their uncertainty are compared using various extreme value distributions and parameter estimation methods.

5.1. EXTREME VALUE DISTRIBUTIONS

Statistical models are commonly used to assess the RPs of extreme events by fitting observed extremes to a distribution and extrapolating to estimate extreme discharges. This method assumes the full distribution of extremes is described by the distribution. Different distributions can result in significantly different estimates for the same RP; therefore, it is important to carefully select the best fit distribution (Coulson, 1991). In this section the performance of GP, GEV, and Gumbel are compared for three datasets: observations, modelled data the same length as observations, and 1,040 years of modelled data.

5.1.1. DATA AND METHODOLOGY

Stations of Interest
The six stations of interest include the four major outlets in the Ardennes and the stations with the highest increasing and highest decreasing trends in AM discharge.

Data and Time Series Length
Observed hourly discharge is available at each station of interest. Average daily observed discharge is obtained before performing the EVA so results may be compared with modelled discharge. Daily modelled discharge of 1,040 years is available for each station of interest. In addition, to evaluate the impact record length has on the fit of the extreme models, two different record lengths are considered: same length as observations and 1,040 years. Modelled data is randomly sampled from the stacked 1,040 years of modelled data at each station to obtain modelled data that is the same length as observations.

Extreme Value Model
AM are fit to either Gumbel or GEV using MLE and peaks are fit to GP using MLE. To ensure events are independent in the POT approach a declustering time, minimum separation distance between extremes, of 48 hours is chosen. Thresholds are determined by studying the MRLP plots and comparing the results GOF tests for different thresholds.

Evaluation of Results
Confidence intervals of the discharge frequency curves are obtained using parametric bootstrapping (Caires, 2007). Parametric bootstrapping assumes the data follows a specific distribution function. First, estimated discharges are obtained by fitting AM or peaks to the desired distribution. The estimated discharges are randomly sampled, with replacement, 100 times, each the same size as the original dataset, and fit to the desired distribution.

Resampling with replacement indicates that after a value is randomly sampled it is replaced before the next sample is taken; this ensures the resampled datasets vary, because duplicate values are possible, and the estimates can differ. The quantiles with probability of 0.025 and 0.975 of the empirical distribution of the sample of bootstrap estimates are calculated to obtain the 95 percent confidence intervals. Percentile confidence intervals tend to be asymmetric and are typically more realistic than confidence intervals obtained assuming the sample is normally distributed (Caires, 2007).

Akaike's criterion, described in Section 2.4, is used to select the best fit distribution. According to this criterion, the distribution with the lowest AIC score best fits the data. In addition, discharges estimated from fitting observations are compared to evaluate the performance of the synthetic data. Reported discharge provided by Service Public de Wallonie (SPW), obtained from the two-parameter lognormal distribution, is also compared to the estimated return levels.

A summary of this methodology is described in Figure 29.

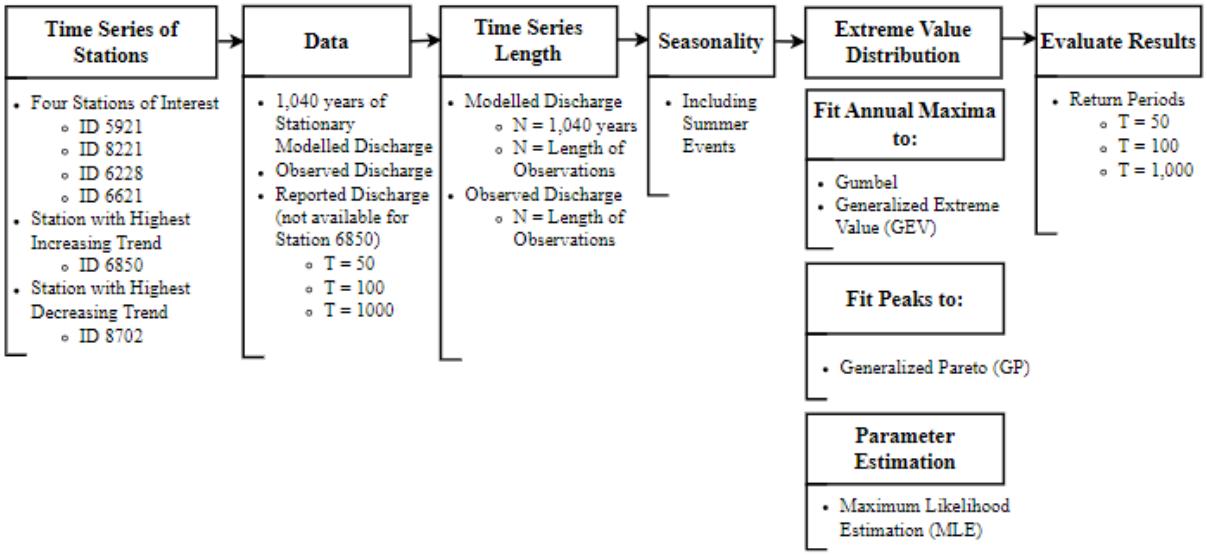


Figure 29: Extreme Value Model EVA Methodology

5.1.2. RESULTS

The 50-year, 100-year, and 1,000-year discharges from the GEV (blue), GP (green), and Gumbel (red) distributions at Station 5921 are shown in Figure 30, Figure 31, and Figure 32. The error bars show the estimates and confidence intervals obtained from the modelled data; lighter shaded bars on the left are for modelled data the same length as observations and the darker shaded bars on the right are for 1,040 years of modelled data. Estimates obtained from observations are shown in the dashed blue (GEV), green (GP), and red (Gumbel) lines; confidence intervals for estimates obtained from observations are not shown. Reported discharge provided by SPW, obtained from the two-parameter lognormal distribution, is shown in the solid pink line. Results for the remaining stations are provided in Appendix C.1

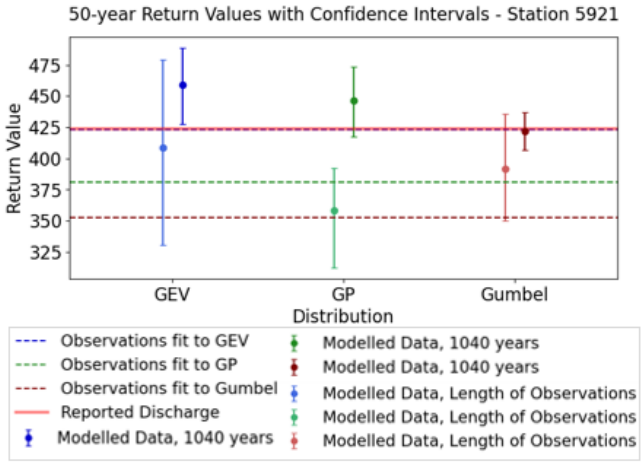


Figure 30: 50-year Discharges Estimated from GEV, GP, and Gumbel - Station 5921

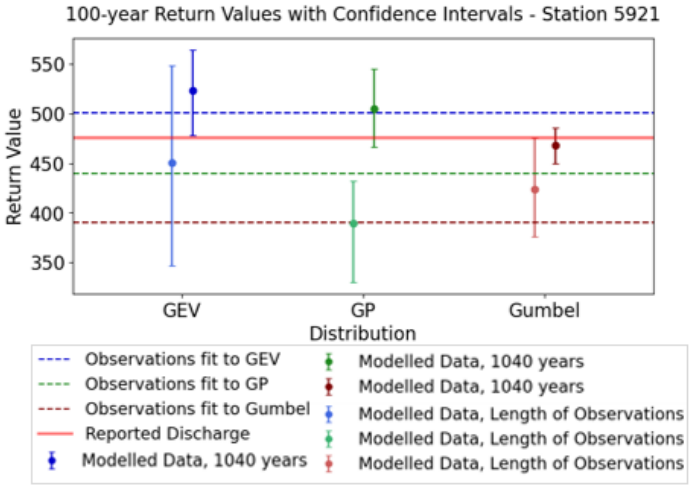


Figure 31: 100-year Discharges Estimated from GEV, GP, and Gumbel - Station 5921

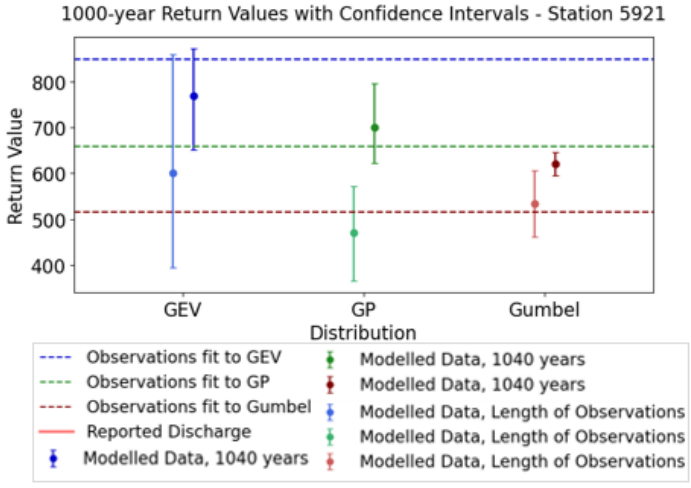


Figure 32: 1,000-year Discharges Estimated from GEV, GP, and Gumbel - Station 5921

Comparing results between the six stations of interest the following is discovered:

- The tail behavior of extremes varies at different locations in the Meuse.

The five stations where GEV provides the highest estimate exhibit heavy tails and belong to the Fréchet distribution. Station 8702 has a light tail and belongs to the Reverse Weibull

distribution. Figure 33 shows PDFs that illustrate the heavy and light-tailed behavior of these two distributions. Extreme discharges that exhibit heavy tail behavior, as shown by the Fréchet distribution, indicate that extreme discharges are more likely to occur than extreme discharges that exhibit light tailed behavior, such as those that belong to the Reverse Weibull distribution. Therefore, it is important to consider heavy tailed behavior, if it exists, to prevent underestimating the probability of extremes.

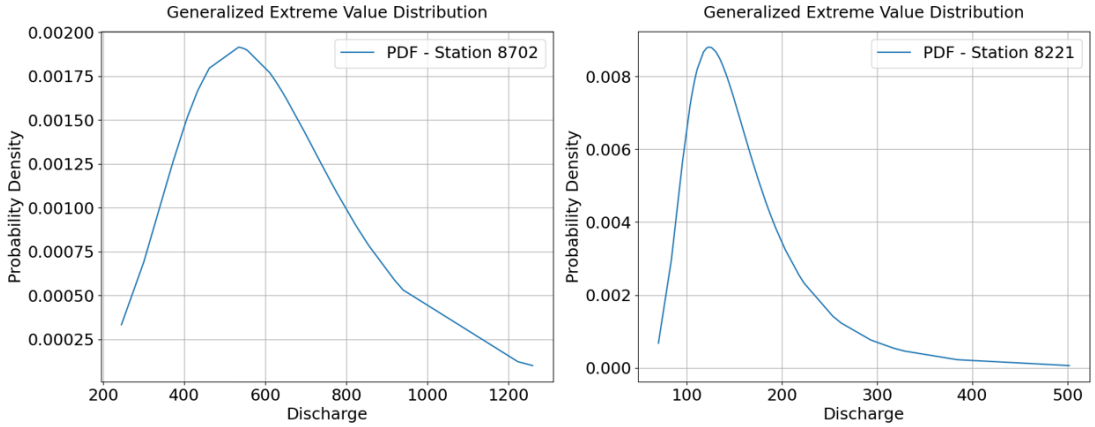


Figure 33: PDF GEV Type III (Light-Tailed) on Left & PDF GEV Type II (Heavy-Tailed) on Right

The three main findings from comparing discharges estimated from observed and modelled data for the three distributions, as shown in Figure 30, Figure 31, and Figure 32 for Station 5921 include:

- Estimates obtained from observations are fairly comparable to those obtained from synthetic data. However, there are some differences discrepancies between observed and modelled discharges.
- Extreme value distributions can result in significantly different estimated discharges for the same RP. The difference between estimates obtained from different distributions increases with increasing RP.
- Increasing the record length resulted in narrower confidence intervals which indicates that longer datasets can be used to obtain more precise estimates.

Regarding the first point, GEV, GP, and Gumbel estimates obtained from observations are fairly comparable to those obtained from modelled data. However, larger differences between estimates obtained from modelled and observed data at some stations of interest indicate there are some discrepancies between observed and modelled discharges. It is discovered that at the stations of interest estimates of the best fit model obtained from 1,040-years of modelled data (modelled data the length of observations) are between 31% (43%) lower and 12% (16%) higher than those obtained from observations. Additionally, the largest difference, 43 percent, in GEV estimates obtained from observed and modelled data are Station 6850. This could be a result of the statistical uncertainty in fitting the GEV shape parameter with limited historical data at this station; As presented in Table 4, Station 6850 has only 20 years of observations compared to the other five stations of interest with between 38 and 56 years. The percent differences in 100-year estimated discharges obtained from observed and modelled data at each station are presented in Table 18 in Appendix C.1.

Second, results of this analysis demonstrate extreme value distributions can result in significantly different estimated discharges for the same RP. At the stations of interest GEV

and Gumbel estimates of the 100-year discharge vary between 3 and 14 percent and GP and Gumbel estimates vary between 4 and 9 percent. These percentages do not include those obtained at Station 6228 because Gumbel was not a good fit for the 1,040 years of modelled data. Differences between GEV and Gumbel are smallest at Station 8702 indicating the shape parameter is very close to zero. The percent difference between GEV (GP) and Gumbel estimates for the 50-, 100-, and 1,000-year RPs at each station are shown in Table 19 in Appendix C.1.

Differences between GEV and Gumbel are smallest at Station 8702 indicating near exponential tail behavior. The remaining five stations of interest all exhibit heavy tailed behavior and, as previously presented in Table 4, have much smaller upstream catchment areas than Station 8702. This indicates the possibility of very large local extreme discharges.

Additionally, the difference between estimates increases with increasing RP due to the different tail behaviors of the distributions. GEV and Gumbel estimates of the 1000-year discharge vary between 5 and 26 percent and GP and Gumbel estimates vary between 6 and 16 percent. The difference between estimates obtained from different distributions demonstrates the importance of selecting the best distribution.

Regarding the third point, increasing the record length resulted in narrower confidence intervals. GEV is widely used to model extremes due to the flexibility provided by the shape parameter. However, as shown in Figure 30, Figure 31, and Figure 32, this flexibility results in high uncertainty in discharge estimates when limited data is available. As expected, when additional data is available this uncertainty decreases for all three distributions. However, as shown in Figure 32, there is not a lot of overlap in the confidence intervals of estimates obtained from the 1,040 years of data. This demonstrates that when using longer datasets, it is especially important to find the most suitable model for the data to prevent a false perception of accuracy. The impact of record length is further explored in Section 6.1.

A summary of the best fit distribution based on AIC is provided in Table 9. At all stations included in this analysis, the Q-Q plot and KS tests, introduced in Section 2.4, indicated that all three distributions were a good fit for all three datasets except for the 1,040 years of modelled data at Station 6228. A summary of the GOF tests for this station is provided in Appendix C.1.2. As shown in Table 9, AIC indicates that GEV is the best fit distribution for most datasets, however, the best fit model shifts to a different distribution at three of the six stations when more synthetic data is added. Two possible explanations are:

- Differences in sampling methods; less extremes are fit to GEV model.
- Calculation in distribution selection criteria, AIC, penalizes distributions that require estimation of additional parameters.

While the AM approach is commonly preferred due to its straightforward sampling process, it is often critiqued since it neglects useful information such as the second highest event in a year which could be higher than other extremes in an AM series. POT extracts a larger number of extremes, therefore, providing more information about the tail behavior of extremes.

Additionally, AIC penalizes distributions that require more parameters to be estimated. As previously mentioned, the flexibility provided by the GEV shape parameter results in

significant uncertainty when limited data is available. When more data is added the uncertainty in estimates decreases; this concept will be further explored in Section 6.1. However, the AIC penalizes the GEV distribution because it requires estimation of this additional parameter.

At the three stations where GP was the most suitable model for the 1,040 years of synthetic data, Stations 8221, 6228, and 8702, the average difference between 100-year discharges obtained from GEV and GP is three percent. Therefore, while AIC indicates that GP is a better fit for the 1,040 years of modelled data than GEV at three of the six stations of interest, estimated discharges obtained from the two distributions do not significantly vary.

Table 9: Best Fit Distribution Based on AIC

Station	Record Length	Observed or Modelled	Best Fit Distribution(s) based on AIC
8221	55 years	Observed	GEV
	55 years	Modelled	GEV
	1,040 years	Modelled	GP
5921	54 years	Observed	GEV
	54 years	Modelled	GEV
	1,040 years	Modelled	GEV
6228	56 years	Observed	GEV
	56 years	Modelled	GEV
	1,040 years[1]	Modelled	GP
6621	49 years	Observed	GEV
	49 years	Modelled	GEV
	1,040 years	Modelled	GEV
8702	33 years	Observed	GEV
	33 years	Modelled	GEV
	1,040 years	Modelled	GP
6850	20 years	Observed	GP
	20 years	Modelled	GEV
	1,040 years	Modelled	GEV

Notes:
1. KS test indicated the Gumbel Distribution was not a good fit.

5.2. PARAMETER ESTIMATION METHODS

This section investigates the impact parameter estimation techniques used to fit data to Gumbel have on estimated discharges and their uncertainty.

5.2.1. DATA AND METHODOLOGY

Stations of Interest

The six stations of interest include the four major outlets in the Ardennes and the stations with the highest increasing and highest decreasing trends in AM discharge.

Data and Time Series Length

Observed hourly discharge is available at each station of interest. Average daily observed discharge is obtained before performing the EVA so that results may be compared with modelled discharge. Daily modelled discharge of 1,040 years is available for each station of interest. In addition, the record length can have a considerable impact on distribution parameters, therefore, two different record lengths are considered: same length as observations and 1,040 years. Modelled data is randomly sampled from the stacked 1,040 years of modelled data to obtain modelled data that is the same length as observations.

Extreme Value Model

Gumbel is used for this analysis because extreme discharges at Borgharen, where the Meuse enters the Netherlands, can be closely modelled by Gumbel. MoM and MLE are among the most widely known and used methods for estimating the Gumbel parameters (Aydin & Şenoğlu, 2015a). Therefore, to evaluate the influence of parameter estimation methods, AM modelled discharge is fit to Gumbel using MoM and MLE.

Evaluation of Results

As in Section 5.1, parametric bootstrapping is used to obtain the confidence intervals and Akaike's criterion is used to select the best fit model.

A summary of the methodology used in this analysis is described in Figure 34.

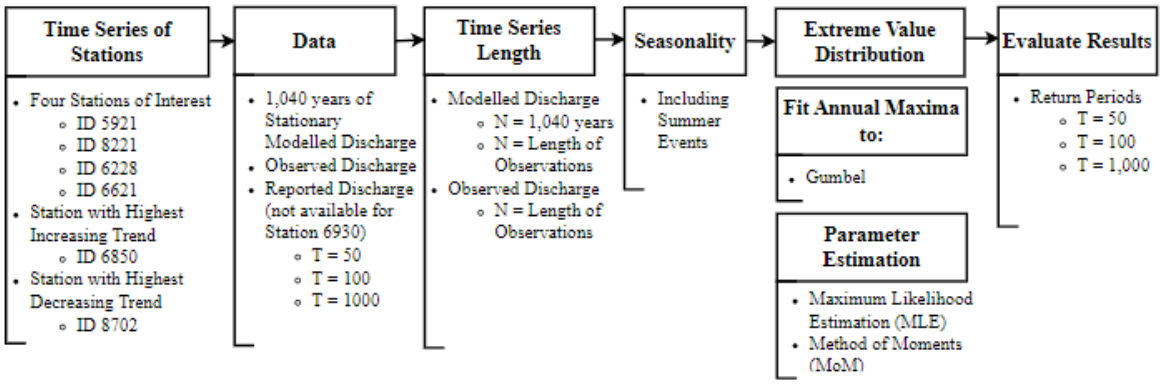


Figure 34: Parameter Estimation EVA Methodology

5.2.2. RESULTS

A box plot showing the estimated discharges for Station 5921 estimated by fitting the data to Gumbel using MoM and MLE is shown in Figure 35. Estimates are grouped by model, for example "MLE Obs" shown on the x-axis shows the results of observed AM fit to Gumbel using MLE; further explanation is provided in the caption. The light, medium, and dark blue bars, shown for each model, specify the estimate and confidence intervals corresponding to the 50-, 100-, and 1000-year RPs. The red shaded lines show the reported discharges obtained from SPW. Results for the remaining stations are provided in Appendix C.2.

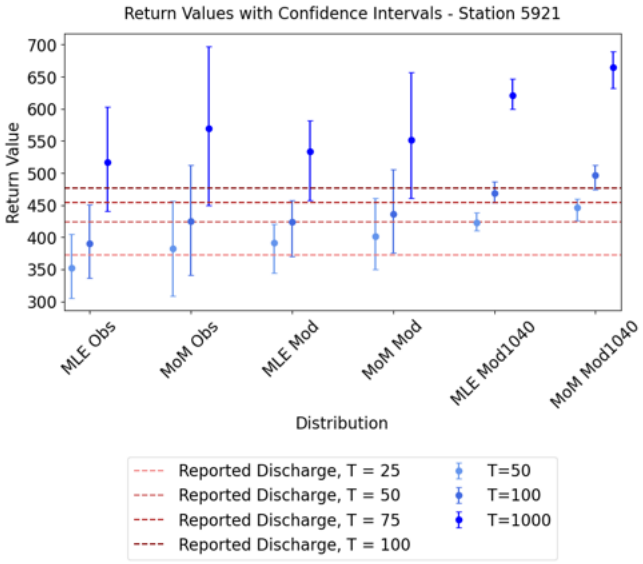


Figure 35: Station 5921 - Comparison of Discharges Estimated Using Different Parameter Estimation Techniques (Obs, Mod, and Mod1000 refers to observations, modelled data the same length of observations, and modelled data 1,040 years long; The reported discharge is the discharge provided by SPW and was obtained by fitting the two parameter lognormal distribution using MLE.)

The two main findings from comparing of estimates obtained using MoM and MLE include:

- Estimated discharges vary depending on the parameter estimation method.
- Estimates obtained from MLE are more precise than those obtained from MoM.

First, results of this analysis indicate estimated discharges vary depending on the parameter estimation method. Estimates at the five stations heavily impacted by the 2021 event obtained using MLE and MoM to fit observed extremes varied between 8 and 14 percent; Station 8702, that was not heavily impacted by the 2021 event had a 2 percent difference between estimates obtained using MLE and MoM. The 2021 event may contribute to the difference between estimates obtained using MLE and MoM. The presence of one very extreme event likely makes it more difficult to obtain accurate estimates of the distribution parameters.

The difference between estimates of the 100-year discharge obtained using MLE and MoM to fit 1,040 years (same length of observations) of modelled extremes varied between 1 and 9 (0 and 5) percent. These values do not include the percent differences at Station 6228 because Gumbel was not a good fit for the 1,040 years of modelled data. These results suggest that estimates obtained using MLE and MoM can vary up to 9 percent. The percent difference between estimates of the 100-year discharge obtained using MLE and MoM at all stations is provided in Table 21 in Appendix C.2.

Second, estimates obtained using MLE have narrower confidence intervals than those obtained using MoM indicating estimates obtained using MLE are more precise.

The AIC values for Station 5921 are presented in Table 10. Results of the remaining stations are included in Appendix C.1.2. The AIC indicated the model fit using MLE fit the data better than the model fit using MoM for all datasets at all stations.

Table 10: Comparison of Model Fit for MLE and MoM - Station 5921

Distribution - Fit Method	Record Length	Observed or Modelled	AIC
Gumbel - MLE	54 years	Observed	613
Gumbel - MoM	54 years	Observed	614
Gumbel - MLE	54 years	Modelled	593
Gumbel - MoM	54 years	Modelled	594
Gumbel - MLE	1,040 years	Modelled	12071
Gumbel - MoM	1,040 years	Modelled	12090

6. EVENT SETS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

This chapter gives an answer to the research question "*How much influence do event sets have on discharge estimates and their uncertainty?*" Event sets describe the time series used to obtain discharge estimates and, therefore, affect the amount of data included in the analysis. The sensitivity of discharge estimates is compared for two types of event sets: record length and seasonality.

6.1. RECORD LENGTH

The disadvantage of the commonly adopted approach of fitting observed AM or peak discharges to extreme value distributions is that extrapolation is necessary to estimate extreme quantiles. The asymptotic assumption in EVT assumes the tail of the distribution of extremes follows a specific distribution as the sample size approaches infinity. However, limited historical data makes it difficult to confidently model the tail behavior of extremes. RACMO makes it possible to generate long synthetic time series to provide more information about the tail behavior of extreme discharges. This section studies the impact record length has on the estimation of discharges and their uncertainty.

6.1.1. DATA AND METHODOLOGY

Stations of Interest
The six stations of interest include the four major outlets in the Ardennes and the stations with the highest increasing and highest decreasing trends in AM discharge.

Data and Time Series Length
The impact record length has on the uncertainty of estimated discharge is assessed by sequentially stacking all 16 members of the RACMO data so that return levels are estimated using various record lengths. Modelled data is randomly sampled from the stacked 1,040 years of modelled data to obtain modelled data the same length as observations.

Extreme Value Model
AM are fit to Gumbel using MLE. Gumbel is used because extreme discharges at Borgharen, where the Meuse enters the Netherlands, are closely modelled by Gumbel. In addition, AM are fit to GEV using MLE to evaluate the impact record length has on the shape parameter.

Evaluation of Results
As in Section 5.1, parametric bootstrapping is used to obtain the confidence intervals. The evolution of distribution parameters with increasing record length is shown using parameters obtained from the Scipy Stats Python package which uses the opposite sign convention of the shape parameter than what was introduced in Section 2.2.1 (Virtanen et al., 2020).

A summary of the methodology used in this analysis is described in Figure 36.

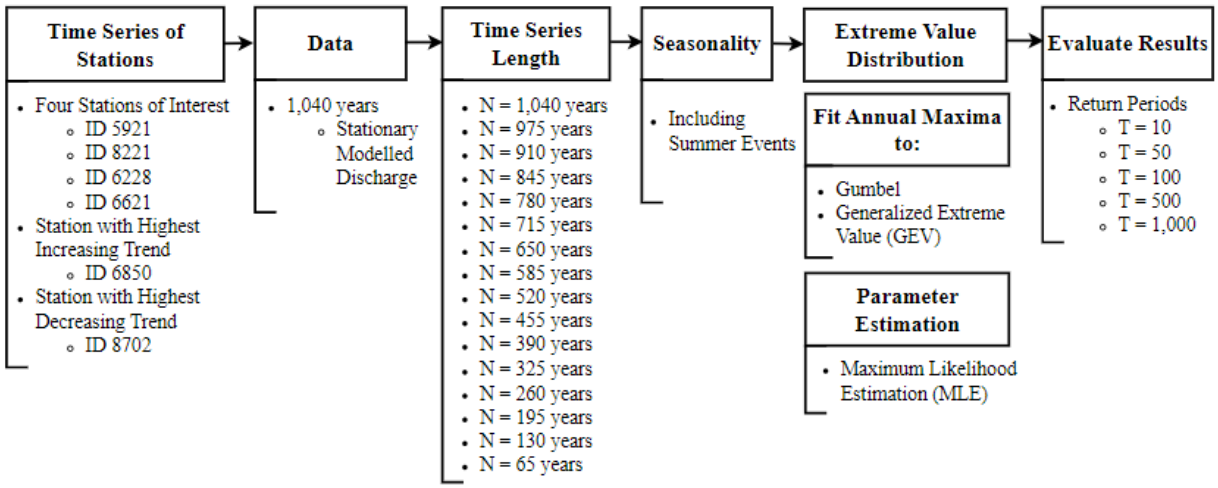


Figure 36: Length of Time Series EVA Methodology

6.1.2. RESULTS

Estimations of discharges and parameters for GEV and Gumbel at Station 5921 are shown for various record lengths in Figure 37 and Figure 38. Results for the remaining stations are included in Appendix D.1.

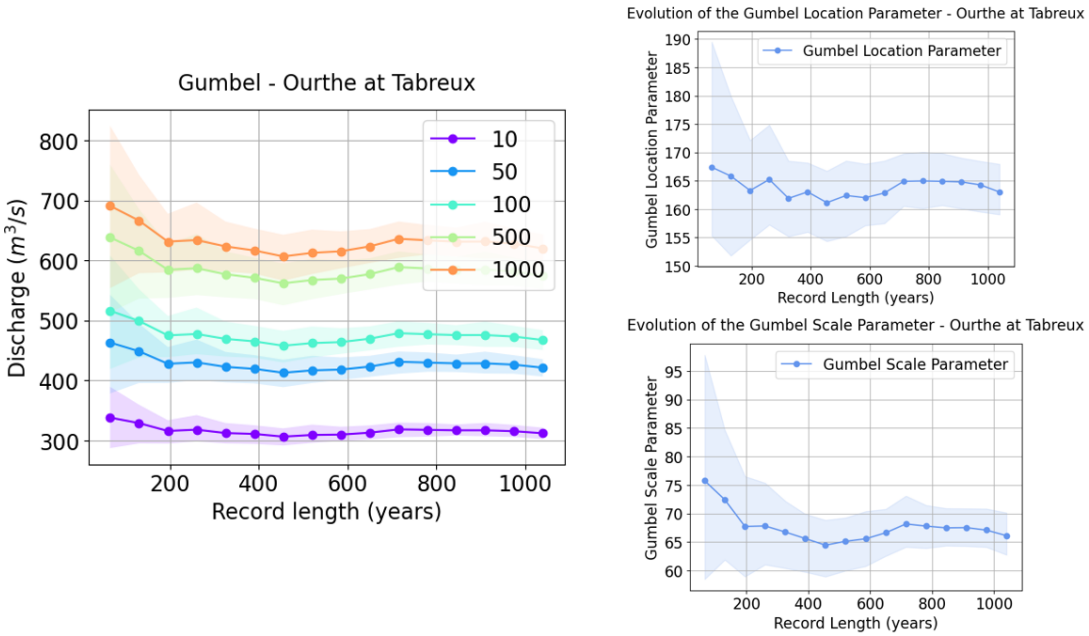


Figure 37: Discharges Estimated from the Gumbel Distribution for varying Record Lengths - Station 5921

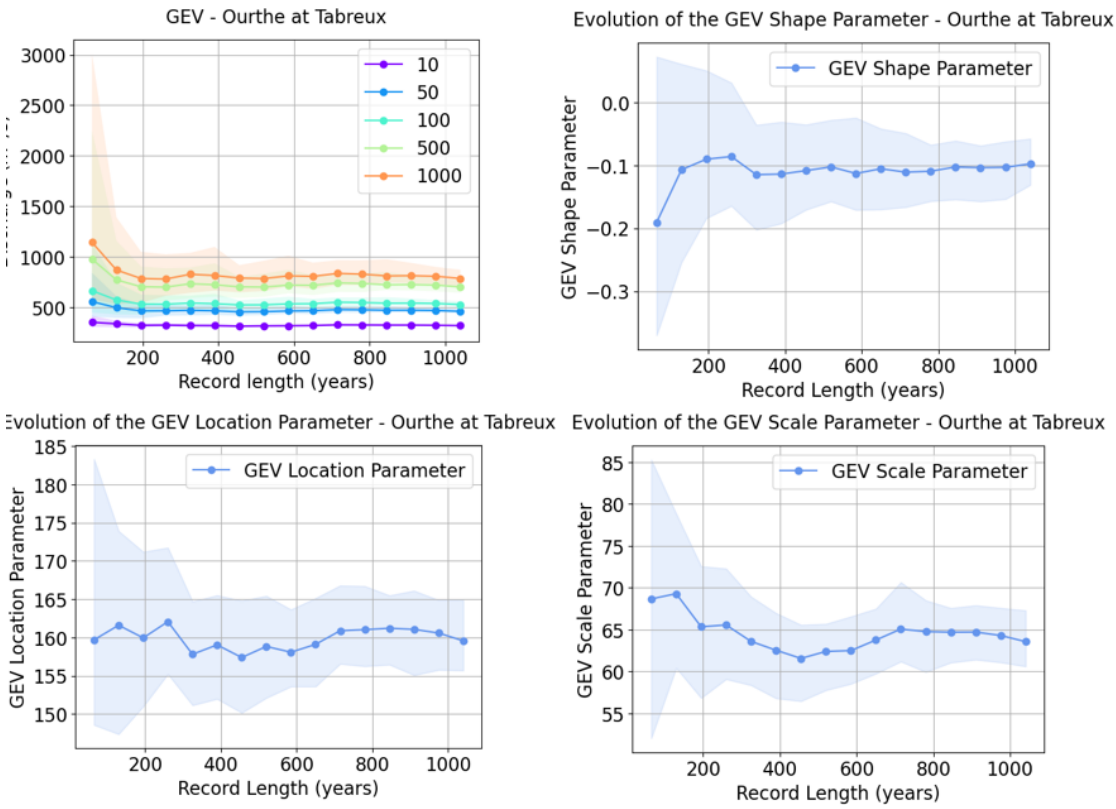


Figure 38: Discharges Estimated from the GEV Distribution for varying Record Lengths - Station 5921

The main findings from this analysis include:

- The confidence of estimated discharges increases with increasing record length.
- The flexibility of the GEV shape parameter results in significant uncertainty in discharge estimates for short records.

First, from Figure 37 and Figure 38 it can be seen that the confidence of estimates for all RPs increases with increasing record length. The range of percent differences in the width of confidence intervals of the 100-year discharge obtained for various record lengths from those obtained from a record length of 65 years is presented in Table 11; results for individual stations are provided in Table 22 in Appendix D.1. For example, at the six stations of interest the width of the confidence interval of 100-year discharges obtained using Gumbel from a record length of 130 years decreased between 20 and 45 percent from estimates obtained from a record length of 65 years. GEV estimates obtained from a record length of 130 years decreased between 15 and 61 percent from estimates obtained from a record length of 65 years. The width of the confidence interval decreases further for increasing record length which demonstrates that estimates become more precise. The width of the confidence interval of 100-year discharges obtained using Gumbel and GEV from a record length of 1,040 years decreased between 75 and 85 percent and 70 and 89 percent, respectively, from estimates obtained from a record length of 65 years.

Table 11: Range of Percent Differences in Width of Confidence Intervals of 100-year Discharges Obtained from Various Record Lengths from those Obtained from a Record Length of 65 years

Record Length	Range of Percent Differences in Width of Confidence Intervals of 100-year Discharges Obtained from Various Record Lengths from those Obtained from a Record Length of 65 years	
	Gumbel	GEV
130 years	20% - 45%	15% - 61%
195 years	38% - 65%	26% - 74%
260 years	50% - 72%	52% - 81%
1040 years	75% - 84%	70% - 89%

Second, as shown by the confidence intervals in Figure 38, there is significant uncertainty in the GEV estimates when limited data is available. This is a result of the flexibility of the GEV shape parameter. For shorter record lengths, the uncertainty in the GEV shape parameter is wide enough to cover the three possible GEV types: Reverse Weibull, Gumbel, and Fréchet. Therefore, the tail behavior of extremes cannot be confidently modelled with short records. However, with increasing record length the shape parameter converges to one GEV type. While GEV is widely used to model hydrologic extremes due to the flexibility provided by the shape parameter, it can also result in high uncertainty when few extremes are fit the distribution. However, the long time series can improve and reduce the uncertainty of extreme discharge estimates.

6.2. SEASONALITY

Application of stationary EVA assumes events are i.i.d and there are no trends, shifts, or cyclicity in the time series. However, presence of seasonality in a hydrologic time series can invalidate these two assumptions. In the presence of seasonality, AM selected from different seasons could belong to different distributions contradicting the assumption of i.i.d. events. Additionally, the assumption of stationarity assumes there is no cyclicity, however the presence of seasonality implies cyclicity is present and parameters are not constant in time. Therefore, it is important to identify whether seasonality is present and understand the potential impact it can have on discharge estimates.

First, a brief introduction of the seasonality in the Meuse is included in Section 6.2.1. Section 6.2.2 describes the methodology used to study the impact of seasonality on extreme discharge estimates and results are provided in Section 6.2.3.

6.2.1. OBSERVED SEASONALITY IN THE MEUSE

As mentioned in Section 4.1, AM discharges and AM 1-day rainfall events do not correspond. To gain a better understanding of seasonality within the Meuse, the seasonality of AM discharge and 1-day rainfall is studied. Figure 39 shows the percent of AM for each season; discharge is shown in the left column and 1-day rainfall is shown in the right column.

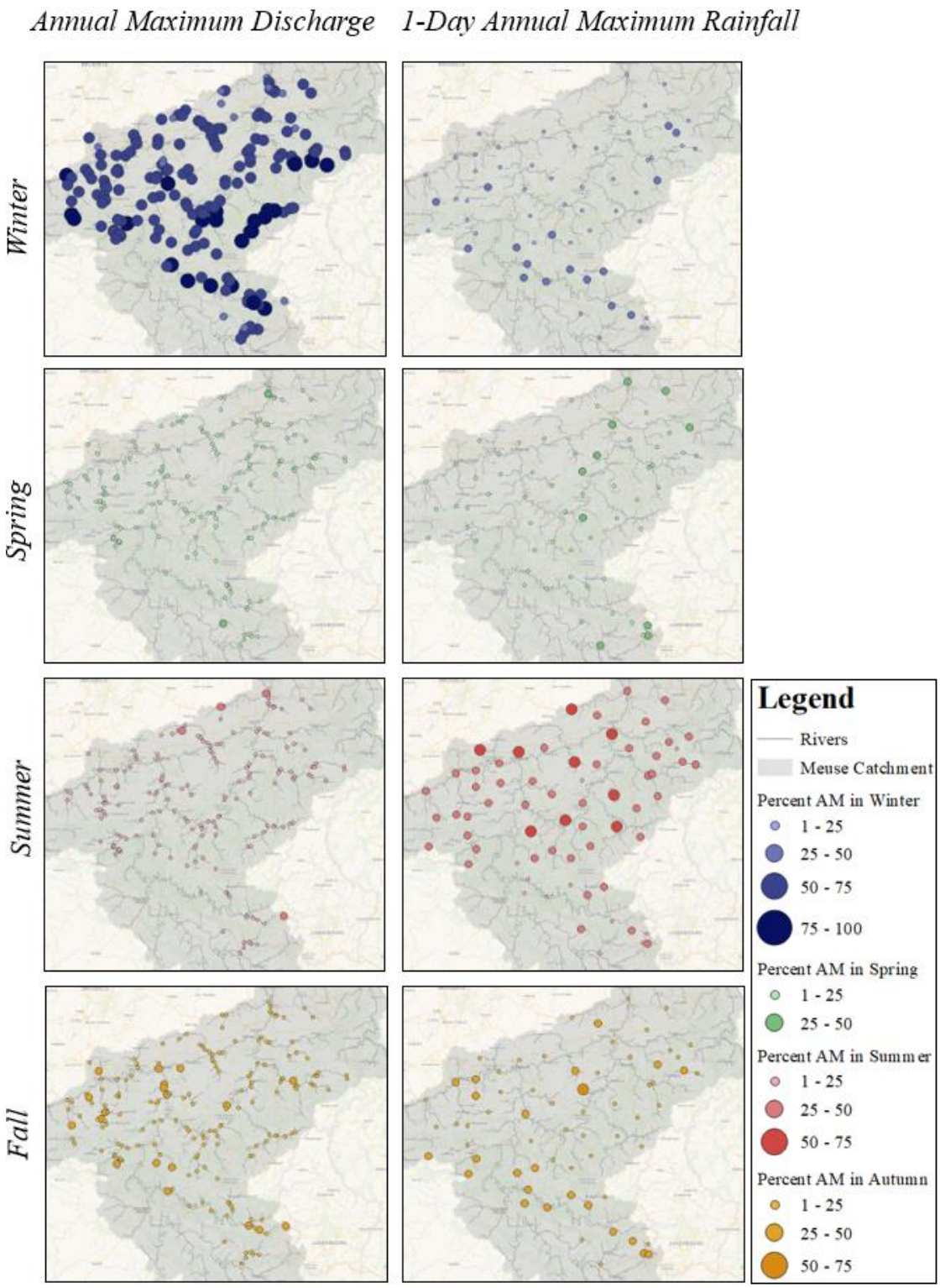


Figure 39: Percent of Annual Maximum Discharge and 1-Day Annual Maximum Rainfall Events in the Meuse

From this figure it is clear the majority of AM 1-day rainfall events occur during summer and the majority of AM discharge events occur during the winter. This agrees with Berghuijs et al. (2011) who studied various flood generating mechanisms across Europe and found maximum daily precipitation tends to occur in summer for most of central and (north) eastern Europe. It also agrees with Villarini et al. (2011) who analyzed seasonal and AM daily discharge records for central Europe and found a large fraction of AM flood peaks occur in winter.

The season of highest AM discharge and 1-day rainfall is shown in Figure 40.

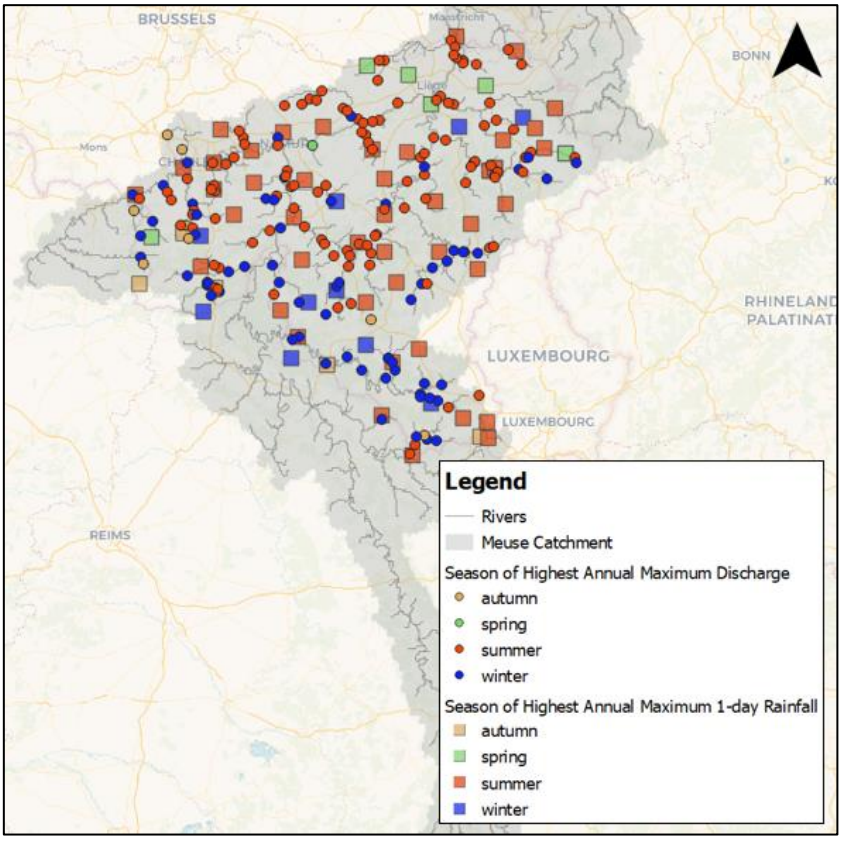


Figure 40: Season of Highest Annual Maximum Discharge and 1-Day Annual Maximum Rainfall

The majority of the highest AM rainfall and discharge in the Lorraine occur in winter while the majority of highest AM rainfall and discharge in the Ardennes occur in the summer. However, as previously shown in Figure 18 of Section 4.1.2, most AM 1-day rainfall in the Meuse occur in the winter. This illustrates that the 2021 event had the most impact in the Ardennes; the highest summer AM resulted from the 2021 event. This matches the results shown in Table 6 that were discussed previously in Section 4.2; Station 8702 is the only station of interest not in the Ardennes and the magnitude of the 2021 event did not exceed the previously recorded highest AM.

This preliminary analysis demonstrates the seasonal variability present throughout the Meuse which is important to consider when estimating extreme discharges.

6.2.2. DATA AND METHODOLOGY

Two analyses are performed to study the impact of seasonality in the Meuse: including versus excluding summer events and considering winter and summer events separately before combining the results. The variables involved in the analyses are briefly described below.

Stations of Interest
The six stations of interest include the four major outlets in the Ardennes and the stations with the highest increasing and highest decreasing trends in AM discharge.

Data and Time Series Length
Observed hourly discharge is available at each station of interest. Average daily observed discharge is obtained before performing the EVA so that results may be compared with modelled discharge. Daily modelled discharge of 1,040 years is available for each station.

To account for the variation of modelled discharge when obtaining modelled data the same length as observations, the 1,040 years of stacked modelled discharge is randomly sampled 100 times to obtain 100 time series the same length of observations. To ensure seasonality is intact, the sampled modelled discharge is randomly sampled from the same month for which it was modelled. The median of the 100 discharge frequency curves will be used to evaluate the modelled data the same length as observations. The confidence interval can then be determined by taking the 2.5 and 97.5 quantiles of the 100 discharge frequency curves.

Seasonality
To evaluate the influence seasonality has on return level estimates, two analyses are performed. In the first analysis estimates obtained using datasets including summer events is compared to estimates obtained excluding summer events. Summer months, April through September, are removed before obtaining AM.

In the second analysis “summer” and “winter” months are partitioned and individual EVA are performed before combining the results as described in Section 2.7. Summer months include April through September and winter months include October through March. EVT requires the assumption that events are i.i.d, therefore, this analysis is based on the assumption that events that occur in the months from April to September and October to March belong to the same distribution.

Extreme Value Model
AM are fit to Gumbel and GEV, using MLE.

Evaluation of Results
As in Section 5.1, parametric bootstrapping is used to obtain the confidence intervals.

A summary of the methodology used to evaluate the impact of seasonality is described in Figure 41.

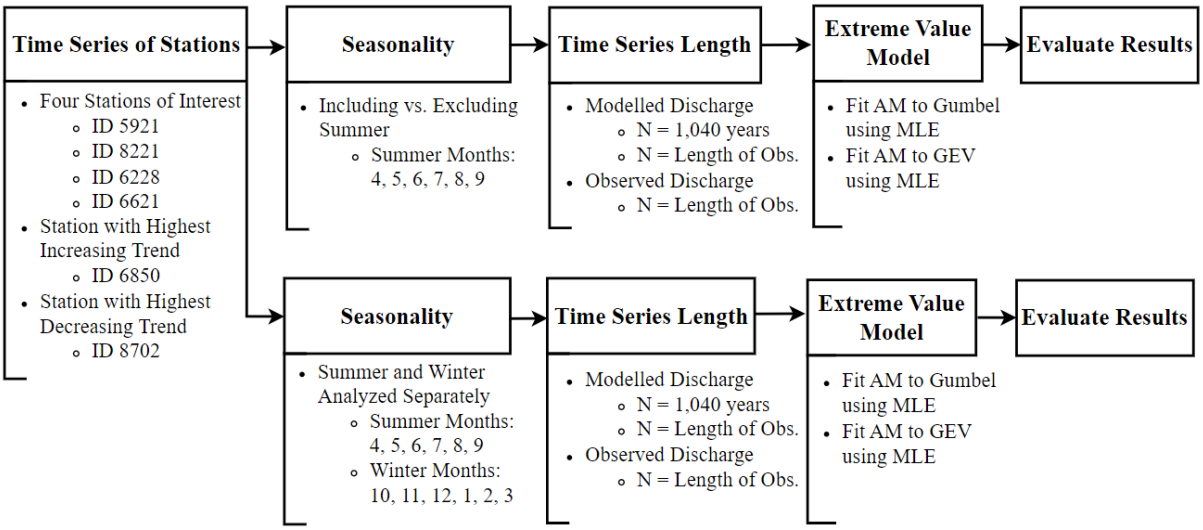


Figure 41: Seasonality EVA Methodology

6.2.3. RESULTS

This section presents the results for Station 5921; results for the remaining stations are included in Appendix D.2.

Seasonality of Annual Maximum Discharge

The seasonality of AM discharge at Station 5921 is shown in Figure 42. The upper left and right figures of Figure 42 show median discharge, grouped by month, for the stacked 1,040 years of modelled discharge and observed discharge, respectively. The lower right figure of Figure 42 shows the months of observed and modelled AM discharge. The lower left figure of Figure 42 shows the 95 percent confidence interval of median discharge, grouped by month, of 100 randomly sampled modelled discharge the same length of observations. This is shown to demonstrate the possible variability in magnitude of AM each month when resampled from the stacked modelled members. The 95% confidence interval of minimum and maximum discharge is also shown to highlight the variation in the sampled discharge throughout the year. The dashed lines show the minimum and maximum discharge available each month from the stacked 1,040 years of modelled discharge.

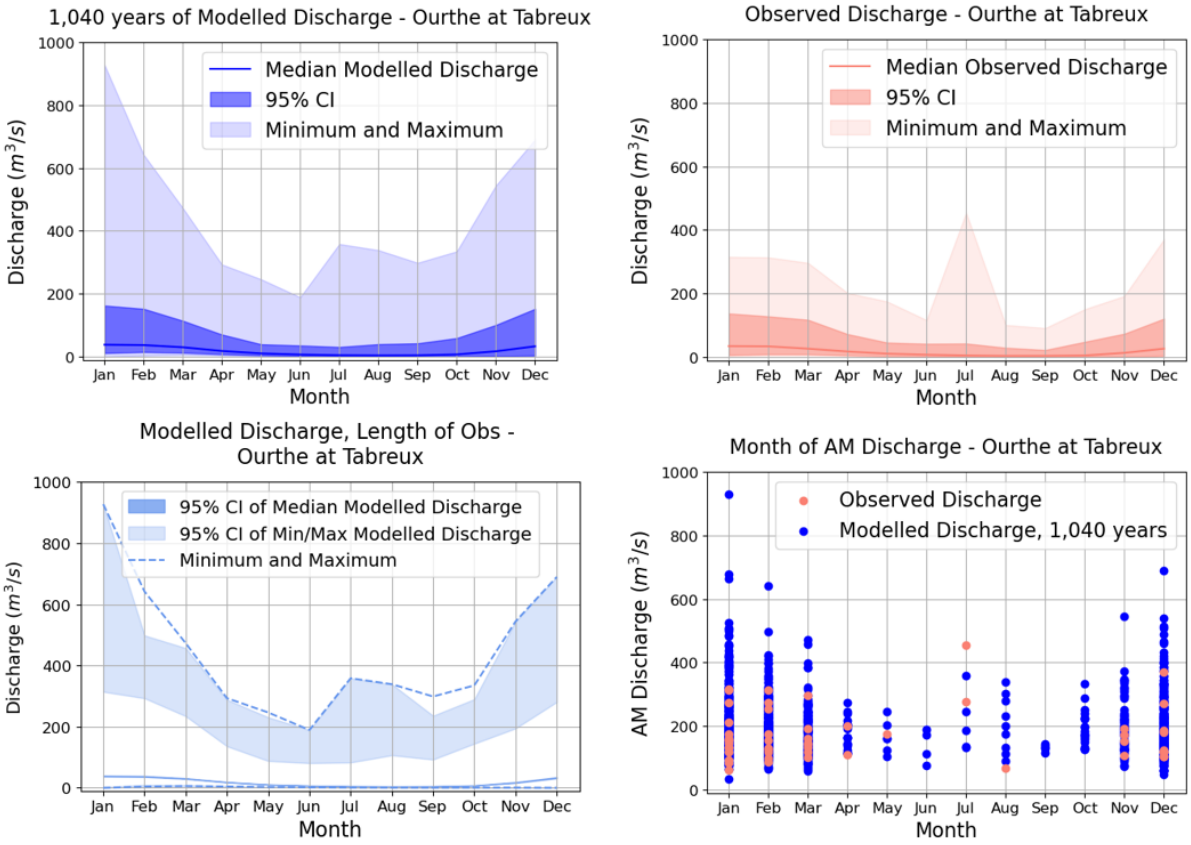


Figure 42: Seasonality of Annual Maximum Discharge - Station 5921

Three main findings from studying seasonality of observed and modelled discharge shown in Figure 42 and Appendix D.2 include:

- The Meuse is dominated by winter events.
- Summer peaks are present in both observed and modelled discharge. However, at four of the six stations included in this analysis the modelled summer peaks do not reach the magnitude of the July 2021 event.
- Less summer AM are present in the modelled data compared to what has been observed.

First, from this analysis it is clear that the Meuse is dominated by winter events. The majority of observed and modelled AM occur in the winter. However, extreme events like what was observed in July 2021 not only demonstrate that the Meuse is susceptible to summer events, but summer events can exceed the magnitude of extreme winter events.

Second, while AM summer events are present in both modelled and observed discharge, the highest summer events in the 1,040 years of synthetic data did not reach the magnitude of the July 2021 event. The ratio of highest modelled summer AM to highest observed summer AM ranged between 0.7 and 1.7 as shown in Table 12. At Stations 6228 and 6621 this ratio is above one indicating that the highest modelled summer AM exceed the highest observed summer AM by factors of 1.7 and 1.3, respectively. However, there is only one summer AM that exceeds the magnitude of the July 2021 event in the 1,040 years of synthetic data at Stations 6228 and 6621. Therefore, the average ratio between highest modelled summer AM of 100 synthetic series randomly sampled to be the same length of observations and highest observed summer AM at Stations 6228 and 6621 is 0.5 and 0.4, respectively. This

demonstrates that although summer AM are present in the 1,040 years of synthetic data, the magnitude of the July 2021 event is not reproduced in most series of modelled discharges.

Station 8702 was the only station included in this study where the highest observed AM did not occur during summer; the July 2021 event is not the highest recorded observed AM and was the only observed summer AM. There are only 2 percent of summer AM present in the 1,040 years of synthetic discharges at this station. As previously shown in Figure 40, winter is the most common season of AM in the Lorraine, which is upstream of Station 8702. In addition, Station 8702 has a significantly larger upstream area compared to the other five stations of interest; Station 8702 has approximately 10,120 km^2 area upstream while the other five stations of interest vary between 70 km^2 and 1,610 km^2. Therefore, Station 8702 is likely impacted by multiple hydrological responses caused by various meteorological factors and soil characteristics of upstream tributaries. These factors influence the seasonal variability at locations throughout the Meuse that is seen by comparing the results of the six stations.

Third, there are less synthetic summer AM than what has been observed. As shown in Table 12, between 5 and 27 percent of observed AM and between 2 and 11 percent of modelled AM occur during the summer. The average percentage of summer AM in 100 randomly sampled series the same length as observations is between 1 and 6 percent. This indicates that more summer AM have been observed than what is present in the modelled data.

Table 12: Comparison Between Observed and Modelled Summer Annual Maxima

Station	Observed Record Length	Ratio of Highest Modelled Summer AM to Highest Observed Summer AM[1]		Percent of AM that Occur during Summer in:		
		Modelled Data (1,040 yr.)	Modelled Data (Length of Obs.)[3]	Observed Data	Modelled Data (1,040 yr.)	Modelled Data (Length of Obs.)[3]
8221	55	0.7	0.3	11%	5%	3%
5921	54	0.8	0.4	11%	5%	3%
6228	56	1.7	0.5	27%	11%	6%
6621	49	1.3	0.4	12%	5%	3%
8702 [2]	38	0.9	0.3	11%	2%	1%
6850	20	0.8	0.2	5%	5%	3%

Note:

1. Ratio above 1 indicates highest modelled summer AM exceeded highest observed summer AM.
2. At all stations of interest, except Station 8702, the July 2021 event was the highest observed summer AM.
3. Average value of 100 synthetic series randomly sampled to be the same length as observations.

Extreme Value Analysis Including versus Excluding Summer Events

Results of the EVA including and excluding summer for Station 5921 are presented in Figure 43. The discharge frequency plot for observations, modelled data the same length of observations, and stacked 1,040 years of modelled data are shown on the left, middle, and right, respectively. For clarity, confidence intervals are only shown for GEV.

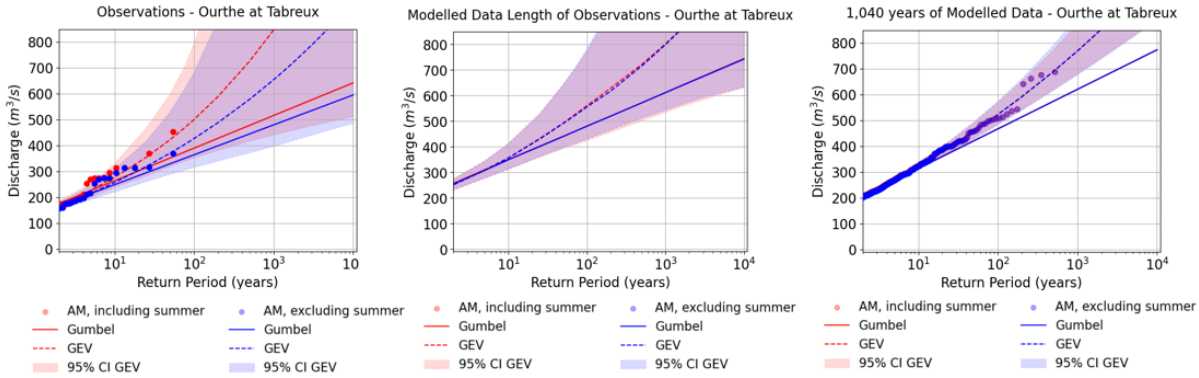


Figure 43: Results of EVA Including and Excluding Summer Events (April - September) - Station 5921

The main findings from this analysis include:

- Gumbel and GEV estimated discharges decreased when summer was removed from observed data in the Ardennes.
- Removing summer AM had minimal impact on discharges estimated from modelled data.

First, 100-year Gumbel (GEV) estimates decreased between 6 and 16 (15 and 48) percent when observed summer AM were removed at the five stations where the July 2021 event exceeded all previously recorded AM. Percent differences of 100-year discharges are presented in Table 13. Although the July 2021 event was an AM at Station 8702, it did not exceed previously recorded AM. In addition, the highest summer AM was only half the magnitude of the highest winter AM. Therefore, removing summer events increased 100-year Gumbel and GEV estimates by 1 and 7 percent, respectively.

As previously seen in Section 4.2, the July 2021 event impacted the tail behavior of extreme discharges at Station 6228 when fitting observed AM to GEV. Therefore, when removing observed summer events from this station the GEV tail behavior changed from a heavy tail to a light tail. Additionally, removing observed summer AM had the most impact at this station with 100-year Gumbel and GEV estimates decreasing 16 and 34 percent, respectively.

Table 13: Percent Difference Between Q_{100} when Including versus Excluding Summer

Station	Percent Difference Between Q_{100} when Including versus Excluding Summer Events[1]					
	Observations		Modelled Data (Length of Obs.)[2]		Modelled Data (1,040 years)	
	Gumbel	GEV	Gumbel	GEV	Gumbel	GEV
8221	-11%	-20%	-0.3%	-0.2%	-0.3%	-0%
5921	-6%	-15%	-0%	-1%	-0.2%	-0.2%
6228	-16%	-34%	-1%	-6%	-2%	-2%
6621	-10%	-32%	-1%	-1%	-1%	-1%
8702	1%	7%	0%	0.1%	0.1%	0.3%
6850	-13%	-48%	0%	0%	0%	0%

Note:
1. A negative percent difference indicates Q_{100} was lower when excluding summer events.
2. Average value of 100 synthetic series randomly sampled to be the same length as observations.

Second, removing summer from the 1,040 years of modelled discharge changed 100-year discharges between -2 and 0.3 percent. The average percent difference for 100 randomly sampled series the same length as observations estimated for Gumbel and GEV was between -1 and 0 percent and -6 and 0.1 percent, respectively. Stations with a higher percentage of modelled summer AM, as previously shown in Table 12, appear to have a larger difference in estimated discharges when summer is excluded. For example, 11 percent of the 1,040 years of modelled data at Station 6228 occurred during the summer and when removing summer, Gumbel and GEV estimates decreased 2 percent. While this difference is much less than the difference in estimates for observations, Table 12 also indicates that there is a less percentage of summer AM in all six modelled series compared to what has been observed.

Extreme Value Analysis Partitioning Summer and Winter Events

The extreme value analyses obtained by considering winter and summer months separately before combining results for Station 5921 is presented in Figure 44.

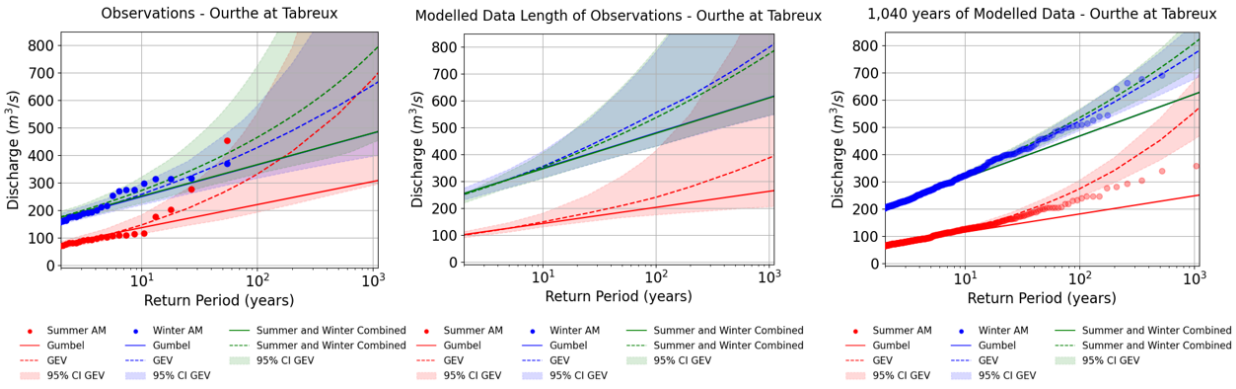


Figure 44: Combined Winter and Summer (GEV and Gumbel) - Station 5921

The main findings from partitioning summer and winter months include:

- Summer events exhibit heavier tails than winter events.
- While winter events dominate the Meuse, at some locations, summer has a strong influence on the tail behavior.

First, for both observed and modelled data, summer extremes exhibit heavier tails than winter extremes. This means that at a certain RP summer AM can exceed winter AM which could indicate summer and winter extremes are generated by different flood mechanisms and belong to different statistical distributions. Therefore, the assumption of i.i.d. events is no longer valid, and summer and winter extremes should be considered separately.

Second, winter events dominate the distribution of extremes. Estimated discharges when only considering winter maxima are close to those obtained when considering the full year of extremes. However, locations with more summer AM exhibit heavier tailed behavior indicating very extreme discharges are more common. For example, Station 6228 has 27% observed summer AM, as previously mentioned in Table 12, and results in a strong heavy tailed behavior as shown in the bottom left corner of Figure 107. Therefore, at some locations summer events have a strong influence on the tail behavior of the extreme value distribution.

As shown in Table 14, GEV is a better fit for both summer and winter EVA at all six stations.

Table 14: AIC Results for EVA for Summer and Winter (dark green indicates strong evidence that model is better fit, medium green indicates model is better fit, light green indicates weak evidence that model is better fit)

Station ID	Distribution	Observations		1,040-year Modelled Data	
		Summer AIC	Winter AIC	Summer AIC	Winter AIC
5921	Gumbel	572	602	10407	12083
	GEV	**560**	**598**	**10295**	**12063**
8221	Gumbel	579	589	9804	11570
	GEV	**559**	**585**	**9698**	**11556**
6228	Gumbel	558	548	8686	10518
	GEV	**535**	**546**	**8423**	**10442**
6621	Gumbel	510	531	9639	11440
	GEV	**488**	**527**	**9486**	**11406**
6850	Gumbel	99	100	4073	5918
	GEV	**78**	**96**	**3965**	**5910**
8702	Gumbel	488	527	12626	14548
	GEV	**485**	**524**	**12600**	**14545**

A comparison between the combined winter and summer EVA and an EVA neglecting seasonality is presented in Figure 45.

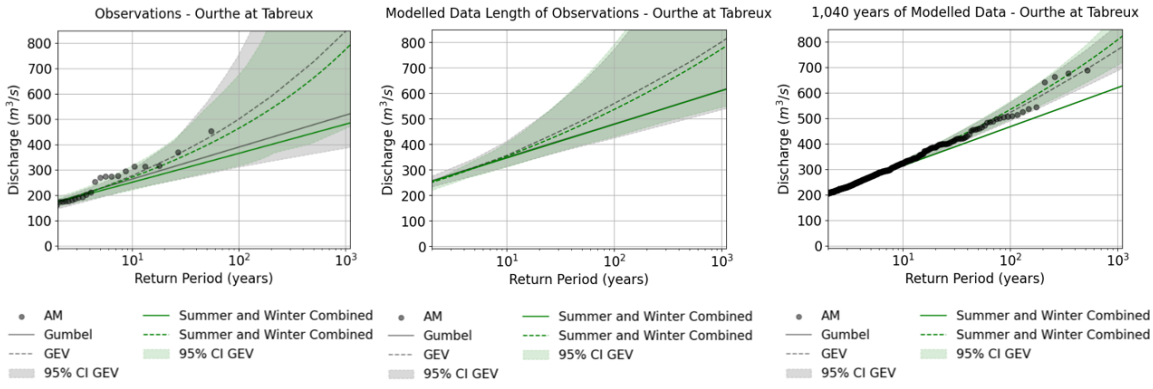


Figure 45: EVA of full series of AM (neglecting seasonality shown in grey) Compared to Combined Summer and Winter EVA (shown in green)

The main findings of this analysis include:

- Neglecting seasonality can result in over or underestimations of extreme discharges.
- Over and underestimations of discharges obtained from modelled discharge are smaller than those obtained from observed discharges.

First, the differences between the discharge frequency curves for the combined summer and winter and the full series, neglecting seasonality, illustrate that neglecting seasonality can result in under or overestimations of extreme discharges. The percent difference between the 100-year discharges obtained from the full year of AM (grey dashed line in Figure 45) and the winter AM (blue line in Figure 44) or the combined summer and winter analysis (green line in Figure 45) is presented in Table 15. At the six stations, neglecting seasonality in historical data underestimated 100-year discharges up to 24 percent and overestimated discharges up to 22 percent.

Second, the over and underestimations of discharges obtained from modelled discharge are smaller than those obtained from observed data. Modelled discharges the same length as observations (1,040 years) were underestimated up to 5 (0) percent and overestimated up to 2 (2) percent. Smaller differences could be a result of limited synthetic summer events. As shown in Table 15, 100-year discharges estimated from modelled winter AM the same length as observations (1,040 years) differed between 0 and 6 (0 and 2) percent. Therefore, the smaller difference between the estimates obtained from the combined summer and winter and the full series, neglecting seasonality, could also be due to the limited summer events in the synthetic data as previously discussed in this section and shown in Table 12.

Table 15: Percent Difference in 100-year Discharges Obtained from GEV for the Full Year of AM and Winter, Combined Analysis

Station	Percent Difference in Q_{100} Obtained from GEV for the Full Year and Winter, Combined Analysis[1]					
	Observations		Modelled Data (Length of Obs.)		Modelled Data (1,040 years)	
	Winter	Combined	Winter	Combined	Winter	Combined
8221	-20%	-9%	0%	1%	0%	2%
5921	-15%	-7%	-1%	-4%	0%	2%
6228	-34%	15%	-6%	-5%	-2%	1%
6621	-32%	-24%	-1%	-4%	-1%	1%
8702	21%	22%	0%	2%	0%	0%
6850	-48%	-17%	4%	0%	0%	0%

1. Negative percentage indicates Q_{100} for winter or combined analysis was lower than what was obtained considering the full year of AM.

7. DISCUSSION

The objective of this thesis is to evaluate the modelling assumptions affecting extreme discharge estimates and their uncertainty. Several factors influencing discharge estimates and their uncertainty were identified in the development of the research questions presented in Section 1.2. Each research question was individually addressed in Chapters 4 through 6 of this report through application of the case study which focuses uses synthetic data generated by RACMO for the Meuse. Within each chapter several analyses were performed to evaluate the impact various modelling assumptions have on extreme discharges and their uncertainty. This chapter provides a high-level discussion of the results presented in this thesis.

7.1. NONSTATIONARITY OF OBSERVED DISCHARGE

7.1.1. TRENDS IN ANNUAL MAXIMA DISCHARGE AND RAINFALL

What is the interpretation of the results:
Statistically significant trends in observed AM discharge were identified at nine of 130 stations in the Meuse: seven increasing trends and two decreasing trends. Of those nine stations, four had statistically significant trends in corresponding rainfall events. While land use change was not explored in this thesis, Tu et al. (2005b) investigated the change in flood peaks of the Meuse and concluded that while flood peaks have increased since the 1980s, land cover has remained relatively stable in the last century. Therefore, increasing trends should be attributed to climate change rather than changes in land use. Decreasing trends may be due to flood control measures, climate change, or long-term water storage.

What are the limitations of these results:
The main limitations of these results arise in the application of a statistical trend test whose results depend on record length and strength of the trend. The longest available record of discharge and rainfall were only 56 and 40 years long, respectively; however, results of the MK test become more powerful with increasing record length. Additionally, Kundzewicz et al. (2004) mention that it is difficult to identify weak trends using statistical tests. Therefore, limited record length and weak trends make it difficult to accurately identify trends.

What did other researchers find, do my results agree:
Diermanse et al. (2010) performed four statistical tests and found there is no statistically significant increasing trend in AM discharges of the Meuse River at Borgharen, where the Meuse enters the Netherlands. However, if the original data was extended from 92 to 130 years by repeating the last years of observed AM, the probability of detecting a statistically significant trend increased. In addition, results of the tests were close to the significance level therefore they concluded that if hypothesized trends were to continue over the next few decades statistical tests would detect an increasing trend. Although Borgharen was not a station of interest in this thesis, results from the study by Diermanse et al. (2010) indicate record length is likely a limiting factor of identifying statistically significant trends.

What are the implications, how does it fit in the bigger picture:
While statistically significant trends were identified at only 9 of the 130 stations, short historical records and the limited power of statistical tests likely limit the number of

statistically significant trends that can be identified. Many researchers, such as Milly et al. (2008), argue against the use of stationary methods when estimating flood risk due to the changing climate. However, results presented in this thesis, later discussed in Section 7.1.3, also demonstrate the potential implications of solely relying on statistical trend tests to infer the persistence of nonstationarity. Regardless, these conclusions should serve as a warning for flood risk management practitioners to reevaluate the assumption of stationarity as more observations become available in the coming years.

7.1.2. OBSERVED CHANGES IN RIVER DISCHARGE SINCE 2015

What is the interpretation of the results:
Since the implementation of GRADE in 2015, events like what occurred in July 2021 have added additional information about the tail behavior of extreme discharges. At Station 6228, the RP of the 2021 event was estimated to be 726 years when including the 2021 AM and 10^8 years when excluding the 2021 AM; At this station the 2021 event was 2.2 times the magnitude of the previously recorded highest AM. Additionally, the 2021 event changed the tail behavior of extremes at this station from a light to heavy tail. This suggests that the addition of just one very extreme event can change the estimated RP of an event by a factor of 1.38^5 (10^8 years/726 years) which demonstrates the vulnerability of applying an EVA to limited historical data.

What are the limitations of these results:
Two limitations of these results include limited record length and potential for measurement errors. Station 6228 had the longest historical record out of the 130 stations in the Meuse, however there are only 56 years of discharges. Additionally, observed discharge extremes are susceptible to measurement inaccuracies; During very extreme floods river gauges may get lost or stop working altogether. Based on these limitations, it should be acknowledged that historical records may not offer a complete picture of extreme discharge events.

What did other researchers find, do my results agree:
Vorogushyn et al. (2022) performed a similar analysis but fit AM to GEV at a German gauge in the Ahr subcatchment, also heavily impacted by the July 2021 event, and had similar findings. They estimated that the July 2021 event corresponded to a RP over 10^8 years when only considering data from 1946 to 2019 and concluded this estimated RP was unrealistic because events of this magnitude occurred in 1804 and 1910.

What are the implications, how does it fit in the bigger picture:
While the 2021 event provides valuable information about extreme discharges in the Meuse, results suggest that the addition of this one very extreme event is likely not enough to justify use of a stochastic weather generator over a physically based climate model. Even with the addition of the 2021 event, there is still significant uncertainty in the tail behavior of extremes; Adding just one extreme event at Station 6228 changed the estimated RP of the 2021 event by a factor of 1.38^5. Additionally, observed records are susceptible to potential measurement accuracies which adds additional uncertainty to estimates. However, comparison between the performance of the stochastic weather generator and RACMO, with the additional observations now available, could be an interesting topic for further research.

7.1.3. COMPARISON OF PREDICTIVE CAPABILITY OF NEVA METHODS

What is the interpretation of the results:
The 100-year discharges estimated for the year 2022 from the NS Linear, uST, and aST models were 10 percent higher, 5 percent lower, and 12 percent lower, respectively, than those estimated from observations. Additionally, when these models were extrapolated to the year 2050, the NS linear model resulted in 100-year discharges 1.5 and 1.6 times higher than those obtained from the uST and aST models, respectively. These results demonstrate there are significant differences between different NS methods and highlights the possibility of significantly overestimating discharges when projecting NS linear trends.

What are the limitations of these results:
The main limitation of these results is that only one NS model was considered in this study, however, there are many approaches of NS modelling that will result in different predictions making it even more difficult to account for nonstationarity. For simplicity, it was assumed that the change in AM can be modelled as a linear function of time, however, it is difficult to predict how nonstationarity will persist, especially with limited historical records; changes after the fitting period are not considered. Time was the only covariate considered, however, several physical processes impact hydrological extremes such as temperature or land use changes. Additionally, Gumbel was the only distribution considered for this analysis, however, distributions can result in significantly different estimates. Lastly, only one station was analyzed to demonstrate the limitations of NS models as a predictive tool for flood risk.

What did other researchers find, do my results agree:
Luke et al. (2017) compared the performance of NS, aST, and uST models and found that when a trend is present in the fitting period, the AIC typically favors the NS model. However, they argue against NS methods because extrapolation of a trend based on limited historical data can lead to unrealistic discharges. They concluded that the uST model is preferred over the NS model when a detected trend can be attributed to a known physical alteration in the watershed. Otherwise, the aST model is preferred.

What are the implications, how does it fit in the bigger picture:
Application of NS models implies the time varying model continues to persist which is often difficult to confirm from limited historical records. While stationarity should not be the default assumption when estimating hydrological extremes, statistical tests should not be solely relied upon to infer the persistence of nonstationarity. Practitioners should investigate possible sources of nonstationarity before deciding how best to estimate future flood risk. Based on the conclusions made by Luke et al. (2017), if changes can be attributed to physical changes in the watershed the uST model is preferred; otherwise, the aST model is preferred.

7.1.4. VARIABILITY OF CURRENT CLIMATE

What is the interpretation of the results:
The 16 members of RACMO, each with 65 years, demonstrate that for limited record length, the uncertainty in estimates from simulations of the current climate is larger than the uncertainty in estimates from a NEVA.

What are the limitations of these results:
One main limitation of these results is the selection of the NS linear model; The assumptions made in this model were discussed in Section 7.1.3. The assumption of a linear trend results

in additional uncertainty not accounted for in the confidence interval. Similarly, the confidence interval of the 16 members does not account for the uncertainty from assumptions made to model the climate. Lastly, since the synthetic discharge is bias corrected to be stationary this analysis could only be performed on series the same length as observations.

What did other researchers find, do my results agree:
Slingo et al. (2011) studied uncertainty in climate models and concluded the natural variability of the climate will always lead to a level of uncertainty in climate models no matter how much the model uncertainty is reduced. Due to the natural variability, even the current observed climate is just one possible scenario of many that could have occurred. Therefore, simulations of the current climate can result in significant differences based on the initial conditions.

What are the implications, how does it fit in the bigger picture:
These results demonstrate that, for limited record length, there is a wider range of possible discharge estimates from the 16 climate scenarios than estimates obtained from a NS linear model. This implies the variability of the current climate generates more uncertainty in estimated discharges than NEVA. Therefore, for limited record length, application of NS models offers little value due to the large variability of the current climate. However, while the long series of synthetic discharges available for this study were stationary it would be interesting to reperform this analysis using longer datasets after incorporating nonstationarity in the generation of synthetic discharges.

7.2. EXTREME VALUE MODELS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

7.2.1. EXTREME VALUE DISTRIBUTION

What is the interpretation of the results:
At five of the six stations, GEV and Gumbel (GP and Gumbel) estimates of the 100-year discharge obtained from the 1,040 years of synthetic data vary between 3 and 14 (4 and 9) percent; KS indicated Gumbel was not a good fit for the 1,040 years of synthetic data at Station 6228, therefore, those results are not included in the previous statement. The difference between estimates increased up to 26 percent for the 1,000-year RP. Therefore, although models can pass statistical GOF tests, different distributions can result in different estimates for the same RP. Additionally, the tail behavior of extremes varies at different locations further emphasizing the importance of comparing the performance of different distributions; heavy tail behavior was found at five of the six stations.

What are the limitations of these results:
The primary limitation of these results is that only three distributions were considered. GEV and GP are widely used, however several distributions have been proposed to fit hydrological extremes. For example, the United States, uses the Log-Pearson Type-3 (LP3) distribution. Additionally, for simplicity the commonly used AIC was the only distribution selection criteria considered, however, comparing results from several tests could provide more insight.

What did other researchers find, do my results agree:
Many researchers have found that while POT is often underemployed due to the difficulty of threshold selection, it makes use of extremes more efficiently than AM because it allows

more than one extreme per year. For example, Bezak et al. (2014) concluded POT led to better results than AM when estimating extreme discharges from 116 years of observations. In this study, the most suitable model for synthetic data shifted from GEV to GP at three of the six stations when more synthetic data was added; possible explanations were discussed in Section 5.1. However, the average difference between 100-year discharges obtained from GEV and GP at these stations was three percent indicating minor differences.

In literature there is limited understanding on the processes that cause heavy tailed behavior in extreme distributions. However, Merz et al. (2022) studied heavy tails of flood peak distributions and drew several hypotheses to possible mechanisms. A few hypotheses included heavy tails of rainfall, type of flood generation process, and mixture of flood types. It is relevant to emphasize the importance of modelling heavy tailed flood distributions, when they exist, to prevent underestimation of extreme discharges.

What are the implications, how does it fit in the bigger picture:
While a fully distributed hydrological model can provide discharge information at locations where no observations are available, the same distribution cannot be applied at all locations within a catchment. Based on these results, it is recommended to compare the fit of various distributions at each location to avoid over or underestimation of estimated discharges.

7.2.2. PARAMETER ESTIMATION METHODS

What is the interpretation of the results:
Estimated discharges obtained using MLE and MoM to fit AM from 1,040 years of synthetic data differed up to nine percent. This demonstrates that selection of the parameter estimation method alone can strongly influence discharge estimates.

What are the limitations of these results:
The primary limitations of these results are that only two parameter estimation methods are compared using the Gumbel distribution. However, many methods are discussed in literature.

What did other researchers find, do my results agree:
Aydin et al. (2015b) used Monte Carlo simulations to compare seven parameter estimation methods, including MoM, MLE, modified maximum likelihood (MML), method of least squares (LS), method of weighted least squares (WLS), method of percentile (PE), and probability weighted moments (PWM), for the Gumbel distribution using sample sizes from 5 to 1000. Comparing the bias, PWM demonstrated the best performance for both the location and scale parameters for all sample sizes. This agreed with results obtained by Mahdi et al. (2005) who used simulations to compare MoM, MLE, and PWM for Gumbel using sample sizes from 5 to 100. Mahdi et al. (2005) concluded that, in terms of accuracy, PWM outperformed MLE and MoM for all sample sizes and MLE outperformed MoM for all sample sizes. While PWM was not considered in this study, the results also showed that, based on AIC, MLE outperformed MoM.

Additionally, Vivekanandan (2015) stated that MLE is a more precise method than MoM which agrees with the results obtained in this research; The width of the confidence interval for estimates obtained using MLE is smaller than MoM.

What are the implications, how does it fit in the bigger picture:
These results highlight the importance of selecting the most appropriate parameter estimation method and demonstrated that the selected method directly impacts the obtained estimates. Therefore, flood risk management practitioners should select a robust method to prevent under or overestimations.

7.3. EVENT SETS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

7.3.1. RECORD LENGTH

What is the interpretation of the results:
The width of the confidence interval of 100-year discharges obtained using Gumbel (GEV) for a record length of 130 years decreased between 20 and 45 percent (15 and 61 percent) from that obtained from a record length of 65 years. When a record length of 1,040 years was used, the width of the confidence interval decreased up to 89 percent from that obtained from a record length of 65 years. Longer series are particularly valuable when estimating the GEV shape parameter which converges for longer record lengths.

What are the limitations of these results:
The main limitation of these results is that only Gumbel and GEV were considered. As presented in Section 5.1, GP was the most suitable distribution for the 1,040 years of synthetic data at three of the six stations. However, threshold selection is time consuming and GEV estimates differed on average three percent at the three stations where GP was most suitable. Therefore, only Gumbel and GEV were considered. Additionally, a common rule of thumb is that estimated return levels are most reasonable for RPs up to one-third the record length (Ludwig et al., 2023). Based on this rule, 1,040 years of synthetic data can provide accurate estimates up to approximately the 340-year RP.

What did other researchers find, do my results agree:
Slater et al. (2021) studied global changes in the 20, 50, and 100-year floods and found that uncertainties tend to be smaller at locations with records longer than 100 years; similar results are found in this case study. Results presented in this thesis are also in agreement with Papalexiou and Koutsoyiannis (2013) who studied over 15 thousand rainfall records with varying record lengths. They concluded that the record length has a large effect on the estimation of the GEV shape parameter which determines the limiting type. Similar results were found in this research; the GEV shape parameter required longer record lengths to converge.

Wiel et al. (2019) demonstrated the added value of large ensemble simulations to study extreme hydrological events in some of the major rivers in the world. They compared GEV estimates obtained from 100 years of data, GEV estimates obtained from 2,000 years of data, and empirical distribution estimates from 2,000 years of data. The 100-year GEV fit did not provide statistically significant estimates of changes in extreme floods. The 2,000-year GEV fit was comparable to the approach using the empirical distribution. The empirical distribution approach is advantageous in the presence of multiple flood generating mechanisms since it is capable of capturing the double distribution. However, there are still large uncertainties for high RPs using the empirical approach.

What are the implications, how does it fit in the bigger picture:
Application of EVT requires the assumption that a large enough number of events occur each year, so the distribution of extremes is asymptotic. However, often only a few decades of observations are available. The longest historical record of the 130 stations in the Meuse is only 56 years meaning that reasonable return levels can be obtained up until the 18-year RP, one-third the record length. However, higher RPs are needed for the design and maintenance of flood defenses; In the Netherlands the discharge corresponding to the 1250-year (10,000-year) RP is used for the design of flood defenses in riverine areas (coastal areas). Based on the results presented in this thesis, generating long series of synthetic data can be used to improve estimates of extreme events. Therefore, to obtain reasonable estimates for higher RPs, it is recommended to investigate large ensemble techniques.

7.3.2. SEASONALITY

What is the interpretation of the results:
Both observed and modelled summer AM exhibited heavier tail behavior than winter AM which suggests that extreme summer and winter events are generated by different flood mechanisms. This contradicts the assumption of i.i.d. events and indicates summer and winter extremes should be analyzed separately. Comparison between the 100-year discharge estimated from the combined winter and summer approach and the full year of AM, demonstrated that neglecting seasonality in historical data (1,040 years of modelled data) led to under and overestimations up to 24 (0) and 22 (2) percent, respectively. Differences are likely smaller for modelled data due to the limited extreme synthetic summer events.

What are the limitations of these results:
The main limitation of these results is that both the number and intensity of summer events is limited in the synthetic discharges. Modelled summer AM exceeded the magnitude of the July 2021 event at only two of the six stations of interest. Additionally, there are between 5 and 27 percent of historical summer AM at the six stations of interest but only between 2 and 11 (1 and 6) percent of modelled summer AM in the 1,040 years of synthetic data (modelled data the length of observations).

What did other researchers find, do my results agree:
Allamano et al. (2011) performed a stochastic experiment to study the impact of neglecting seasonality in hydroclimatic extremes and concluded that neglecting seasonality induces a downward bias in estimates. They emphasized the importance of accounting seasonality in the estimation of hydroclimatic extremes to avoid over or underestimations.

What are the implications, how does it fit in the bigger picture:
When seasonality is present, the assumption of i.i.d. events is no longer valid. The results shown in this thesis demonstrated that summer and winter events have different tail behaviors and, therefore, belong to different distributions. Therefore, the traditional annual maxima approach cannot be used, and seasonality must be explicitly considered.

8. CONCLUSION AND ANSWER TO RESEARCH

The objective of this thesis is to evaluate the influence various modelling assumptions have on extreme discharge estimates by analyzing the long synthetic discharge series generated using synthetic data obtained from RACMO. Several factors influencing discharge estimates and their uncertainty were identified in the development of the research questions that will be directly answered in this chapter. Recommendations are provided in Sections 8.3 and 8.3.

8.1. CONCLUSION AND ANSWERS TO THE RESEARCH QUESTIONS

How much impact can a very extreme event, such as what occurred in July 2021 in central Europe, have on the statistics of extreme discharge?

Results demonstrated the impact a very extreme event, like what occurred in July 2021, can have on discharge estimates varies depending on the available record length, selection of extreme value model, and magnitude of the extreme event relative to historical AM. The three main conclusions from this analysis are briefly described below:

- It is difficult to obtain reliable discharge estimates from short historical records. The ratio between $Q_{100,\ post2021}$ and $Q_{100,\ pre2021}$ was the largest for the station with the shortest historical record of 20 years and decreased with increasing record length.
- GEV estimates were more impacted by the addition of rare events than Gumbel estimates due to the flexibility provided by the GEV shape parameter. At the five stations of interest where the 2021 event exceeded previously recorded AM, the Gumbel 100-year estimated discharges increased by a factor between 1.05 and 1.10. GEV 100-year estimated discharges increased by a factor between 1.13 and 1.35.
- Lastly, the addition of one very extreme event can significantly impact estimates obtained from relatively short historical records. At Station 6228, the 2021 event was 2.2 times the magnitude of the previously recorded highest AM and changed the tail behavior of extremes from a light to a heavy tail. When the 2021 event was excluded from the analysis, the RP of this event was estimated to be over 1.38^5 (10^8 years/726 years) times as high as when it was included. At Station 8221, the 2021 event was 1.4 times the magnitude of the previously recorded highest AM. When the 2021 event was excluded from the analysis, the RP of this event was estimated to be over 2.7 (372 years/138 years) times as high as when it was included.

To what extent do nonstationary statistical methods impact discharge estimates when a trend in observed discharge is identified?

The predictive capability of the NS, aST, and uST models were compared to evaluate their impact on discharge estimates. From this analysis the following conclusions are made:

- NEVA assumes that the time varying model persists into the future which is difficult to confirm from short records. When the models were extrapolated to the year 2050, the NS linear model resulted in 100-year discharges 1.5 and 1.6 times higher than those obtained from the uST and aST models, respectively. This demonstrates that

while a linear model may fit observations well and pass statistical GOF tests, projecting a linear trend could result in large under or overestimations of estimates.
- The uST and aST models provide estimates much closer to observations.

How does the confidence of discharges estimated from various simulations of the current climate compare to the confidence of estimates from a nonstationary extreme value analysis?

The uncertainty in discharge estimated from performing a NEVA on observations were compared to uncertainty in estimates from various simulations of the current climate. The main conclusions from this analysis are briefly described below:

- For limited record length, the uncertainty in estimates from simulations of the current climate is larger than the uncertainty in estimates from a NEVA.
- Short records, 55 years in this analysis, contributes to the statistical uncertainty of estimated discharges. The impact of record length was further explored in Section 6.1.

How much influence do extreme value models have on discharge estimates and their uncertainty

Extreme Value Distributions
In Section 5.1, the performance of GP, GEV, and Gumbel were compared for three datasets at each station of interest. From this analysis the following conclusions are made:

- The tail behavior of extremes varies at different locations in the Meuse.
- At the six stations estimates of the best fit model obtained from 1,040-years of modelled data (modelled data the length of observations) are between 31% (43%) lower and 12% (16%) higher than those obtained from observations.
- GEV and Gumbel estimates of the 100-year discharge varied between 3 and 14 percent and GP and Gumbel estimates vary between 4 and 9 percent. Additionally, the difference between estimates increases with increasing RP. GEV and Gumbel estimates of the 1000-year discharge vary between 5 and 26 percent and GP and Gumbel estimates vary between 6 and 16 percent.
- Increasing the record length resulted in narrower confidence intervals, however, this could lead to a false sense of accuracy if the most suitable model is not selected.

Parameter Estimation Methods
Section 5.2 compared the impact MLE and MoM have on estimated discharges when used to fit data to Gumbel. The main conclusions of this analysis include:

- Estimates obtained using MLE and MoM to fit the 1,040 years of synthetic data had up to 9 percent difference.
- Estimates obtained from MLE were more precise than those obtained from MoM.
- AIC indicated that models fit using MLE fit the data better than those fit using MoM.

How much influence do event sets have on discharge estimates and their uncertainty?

Length of Time Series

Results presented in Section 6.1 demonstrate that estimates become more precise with increasing record length. The two main conclusions from this analysis are:

- Estimates become more precise with increasing record length. The width of the confidence interval of 100-year discharges obtained using Gumbel (GEV) for a record length of 130 years decreased between 20 and 45 percent (15 and 61 percent) from estimates obtained from a record length of 65 years. The width of the confidence interval of 100-year discharges obtained using Gumbel (GEV) for a record length of 1,040 years decreased between 75 and 85 percent (70 and 89 percent) from estimates obtained from a record length of 65 years.
- The flexibility of the GEV shape parameter results in significant statistical uncertainty in discharge estimates for short records. When a shorter record length is used the confidence in the estimate of the GEV shape parameter is wide enough to cover the three limiting types of GEV. However, with increasing record length, the GEV shape parameter converges to one of the three limiting types.

Seasonality

Results presented in Section 6.2 demonstrate that neglecting seasonality could under or overestimate extreme discharges. From this analysis the following conclusions are made:

- Modelled summer AM exceeded the magnitude of the 2021 event at two of the six stations. Additionally, there are between 5 and 27 percent of historical summer AM at the six stations but only between 2 and 11 (1 and 6) percent of modelled summer AM in the 1,040 years of synthetic data (modelled data the length of observations). Both the magnitude and intensity of modelled summer events is limited.
- Gumbel (GEV) estimates decreased between up to 16 (48) percent when summer was removed from observed data at the five stations of interest where the 2021 event exceeded all previously recorded AM. Gumbel (GEV) estimates changed up to 6 percent when summer was removed from modelled data. The smaller difference when removing summer from modelled data is likely due to the limited extreme summer events present in the synthetic data.
- Analyzing summer and winter events separately revealed that summer extremes exhibit a heavier tail behavior than winter extremes which suggests that summer and winter extremes are generated by different flood mechanisms. This indicates that summer and winter extremes belong to different statistical distributions contradicting the assumption of i.i.d. events. Therefore, seasonality should be accounted for.
- Comparison between the 100-year discharge estimated from the combined winter and summer approach and the full year of AM, demonstrated that neglecting seasonality in historical data led to under and overestimations up to 24 and 22 percent, respectively. Modelled discharges the same length as observations (1,040 years) were underestimated up to 5 (0) percent and overestimated up to 2 (2) percent. Differences

are likely smaller for modelled data due to the limited extreme synthetic summer events.

To what extent do the modelling assumptions in an extreme value analysis affect the estimation of extreme discharge return levels?

Throughout this thesis several modelling assumptions were evaluated to determine their influence on the estimation of extreme discharges using 1,040 years of synthetic data generated by RACMO. While short historical records make it difficult to obtain precise discharge estimates, the results presented in this thesis demonstrate that each decision made throughout an EVA can impact estimates.

Longer records provide considerable value to an EVA. At the six stations of interest, the width of the confidence interval of 100-year discharges obtained using Gumbel (GEV) for a record length of 1,040 years decreased between 75 and 85 percent (70 and 89 percent) from estimates obtained from a record length of 65 years. However, even with the added value of a longer series of synthetic series, results presented in this thesis demonstrate that there is still significant variability in discharge estimates under the various modelling assumptions. For the six stations, the maximum percent difference of the 100-year discharge obtained from the 1,040 years of synthetic data under various modelling assumptions are summarized below:

- Neglecting (full year of AM) versus accounting for seasonality (partitioned summer and winter analysis and combining results) led to estimates that varied up to 2 percent. However, it was also shown that there are a limited number of synthetic summer AM with respect to observed summer AM. Neglecting versus accounting for seasonality led to estimates obtained from observations that varied up to 24 percent.
- GEV (GP) and Gumbel estimates varied up to 14 (9) percent. Percentage does not include results obtained at Station 6228 because Gumbel was not a good fit for the 1,040 years of synthetic data at this station.
- MLE and MoM estimates varied up to 9 percent. Percentage does not include results obtained at Station 6228 because Gumbel was not a good fit for the 1,040 years of synthetic data at this station.

Therefore, while long series of discharges reduce the statistical uncertainty, careful consideration should be given to each modelling assumption to avoid precise but inaccurate estimates.

8.2. RECOMMENDATIONS FOR FLOOD RISK PRACTITIONERS

Based on the results presented in this thesis the following recommendations may be given to flood risk practitioners:

- When limited data is available, one extreme event can significantly impact discharge extremes. Additionally, observed extremes are susceptible to measurement inaccuracies. Therefore, it is recommended to investigate ways to generate longer series of meteorological data that can be used in a hydrological model to obtain long series of discharges.

- Results of statistical tests should not be solely relied on to infer the persistence of nonstationarity. Before applying a NS model, possible mechanisms responsible for nonstationarity should be investigated, such as climate or land use changes.
- Generating stationarity synthetic data for a NS system can result in under or overestimations of extreme discharges. A physically based climate model can be used to obtain long, synthetic NS meteorological data and land use changes can be incorporated into a hydrological model to obtain long, synthetic series of NS discharges. Statistical trend tests for observed and modelled data can be compared to evaluate and improve model performance.
- Large differences between observed and modelled estimates could be due to limited data making it difficult to obtain robust estimates from short historical record or be indicative of poor model performance. Results could be corrected by using the location and scale parameters obtained from observations and the shape parameter estimated from the long series of modelled data, which as shown in Section 6.1, becomes more precise when more data is available.
- While multiple distributions can pass formal GOF tests at one location, each distribution has distinct tail behavior that describes how often extremes occur. Additionally, the tail behavior of extremes varies at different locations in the same system. Therefore, different distributions should be compared at each location of interest when estimating hydrological extremes.
- A robust parameter estimation method should be selected to prevent over or underestimations of hydrological extremes.
- To obtain reliable estimates of high RPs, beyond one-third of the record length, it is recommended to investigate ways to generate longer time series to obtain more precise discharge estimates.
- Results presented in Section 6.2 illustrate that both the number and intensity of summer events is limited in the synthetic discharges. This likely indicates that the RACMO climate model does not capture the physical mechanisms that generate summer events. Therefore, it is recommended to further refine the RACMO climate model and focus on the calibration of summer events.
- Results demonstrated that summer AM have a heavier tail behavior than winter AM, which indicates that at a certain return period summer AM can exceed winter AM. This suggests that summer and winter extremes belong to different distributions and seasonality should be explicitly accounted for in EVA.

8.3. RECOMMENDATIONS FOR FURTHER RESEARCH

Recommendations for further research related to the impact of nonstationarities include:

- Incorporate the climate and land use changes in the hydrological model.

Recommendations for further research related to the uncertainties resulting from extreme value models include:

- Study the performance of additional parameter estimation methods. This thesis was limited to comparison between MLE and MoM but it may be interesting to investigate Bayesian approaches to estimate parameters.
- Evaluate the performance of additional extreme value distributions such as the Metastatistical Extreme Value (MEV) distribution which does not require the

asymptotic assumption required in the application of EVT (Marani & Ignaccolo, 2015). Therefore, it would be interesting to compare the MEV framework EVA.

Recommendations for further research related to the uncertainties in event sets include:

- RACMO climate data could be used as input to the stochastic weather generator to obtain 50,000 years of data that can be put into the hydrological model. Using this approach, could eliminate the need for statistical extrapolation and would be capable of capturing double distribution in the case of multiple flood generating mechanisms.
- Evaluate the impact seasonality has using POT. In this study some datasets were best modelled using the POT method. Allamano et al. (2011) studied the impact of neglecting seasonality and concluded that the annual maxima approach often disguises the seasonality. However, they found that when using the POT method, neglecting seasonality can result in underestimations of discharges. Therefore, this would be interesting to explore for future research.

There are several modelling assumptions that impact extreme value estimates. In the initial stages of this research, several modelling assumptions impacting discharge estimates and their uncertainty were identified and are recommended for future research including:

- Uncertainties related to the hydrologic model.
- Uncertainties related to the temporal resolution applied in the hydrologic model.

BIBLIOGRAPHY

Akaike H. (1974). A new look at the statistical model identification. *IEEE*.

Allamano, P., Laio, F., & Claps, P. (2011). Effects of disregarding seasonality on the distribution of hydrological extremes. *Hydrology and Earth System Sciences*, *15*(10), 3207–3215. https://doi.org/10.5194/hess-15-3207-2011

Aydin, D., & Şenoğlu, B. (2015a). Monte carlo comparison of the parameter estimation methods for the two-parameter gumbel distribution. *Journal of Modern Applied Statistical Methods*, *14*(2), 123–140. https://doi.org/10.22237/jmasm/1446351060

Aydin, D., & Şenoğlu, B. (2015b). Monte carlo comparison of the parameter estimation methods for the two-parameter gumbel distribution. *Journal of Modern Applied Statistical Methods*, *14*(2), 123–140. https://doi.org/10.22237/jmasm/1446351060

Berger, H. E. J. (1992). *Flow Forecasting for the River Meuse*.

Bezak, N., Brilly, M., & Šraj, M. (2014). Comparaison entre les méthodes de dépassement de seuil et du maximum annuel pour les analyses de fréquence des crues. *Hydrological Sciences Journal*, *59*(5), 959–977. https://doi.org/10.1080/02626667.2013.831174

Bouaziz, L., Slomp, R., Lammersen, R., van der Sleen, N., van Voorst, L., van den Brink, H., & Couasnon, A. (2022). *Interreg Euregio Meuse-Rhine Estimation of discharge extremes in the Meuse basin Application of high-resolution climate and hydrological models*.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. In *Behavioral Ecology and Sociobiology* (Vol. 65, Issue 1, pp. 23–35). Springer Verlag. https://doi.org/10.1007/s00265-010-1029-6

Caires. (2007). *WL | delft hydraulics Extreme wave statistics*.

Coles, S. (2001). *An introduction to statistical modelling of extreme values*. Springer-Verlag.

Coulson, C. H. (1991). *Manual of Operational Hydrology in B.C.*

Cousineau, D., Brown, S., & Heathcote, A. (2004). *Fitting distributions using maximum likelihood: Methods and packages*.

Descy, J.-P., Floury, M., Moatar, F., & Sauvage, S. (2022). Chapter 5 - Continental Atlantic Rivers: The Meuse, Loire and Adour-Garonne Basins. *Elsevier Ltd*. https://doi.org/10.1016/B978-0-08-102612-0.00005-5

Diermanse, F. L. M., Kwadijk, J. C. J., Beckers, J. V. L., & Crebas, J. I. (2010). *Statistical trend analysis of annual maximum discharges of the Rhine and Meuse rivers*. http://www.ruimtevoorderivier.nl

Dullaart, J. C. M., Muis, S., Bloemendaal, N., Chertova, M. V., Couasnon, A., & Aerts, J. C. J. H. (2021). Accounting for tropical cyclones more than doubles the global population

exposed to low-probability coastal flooding. *Communications Earth and Environment*, *2*(1). https://doi.org/10.1038/s43247-021-00204-9

El Adlouni, S., Bobée, B., & Ouarda, T. B. M. J. (2008). On the tails of extreme event distributions in hydrology. *Journal of Hydrology*, *355*(1–4), 16–33. https://doi.org/10.1016/j.jhydrol.2008.02.011

Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins-Final report of GRADE 2.0. (n.d.).

Hazelton, M. L. (n.d.). *Method of Moments Estimation*.

Hegnauer M. (2014). *Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins-Final report of GRADE 2.0*.

Hegnauer, M., Veersma, J. J., van den Boogaard, H. F. P., Buishand, T. A., & Passchier, R. H. (2014). *Generator of Rainfall and Discharge Extremes (GRADE) for the Rhine and Meuse basins-Final report of GRADE 2.0*.

Kamal, N., Pachauri, S., & Pradesh, A. (2018). Mann-Kendall Test-A Novel Approach for Statistical Trend Analysis. *International Journal of Computer Trends and Technology*, *63*(1). http://www.ijcttjournal.org

Kendall M.G. (1955). *Further Contributions to the Theory of Paired Comparisons*.

Kundzewicz, Z. W., Graczyk, D., Maurer, T., Przymusińska, I., Radziejewski, M., Svensson, C., & Szwed, M. (2004). *Detection of change in world-wide hydrological time series of maximum annual flow*. http://grdc.bafg.de

Ludwig, P., Ehmele, F., Franca, M. J., Mohr, S., Caldas-Alvarez, A., Daniell, J. E., Ehret, U., Feldmann, H., Hundhausen, M., Knippertz, P., Küpfer, K., Kunz, M., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Seidel, F., & Wisotzky, C. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe - Part 2: Historical context and relation to climate change. *Natural Hazards and Earth System Sciences*, *23*(4), 1287–1311. https://doi.org/10.5194/nhess-23-1287-2023

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water Resources Research*, *53*(7), 5469–5494. https://doi.org/10.1002/2016WR019676

Mann, H. B. (1945). *Nonparametric Tests Against Trend* (Vol. 13, Issue 3). https://www.jstor.org/stable/1907187

Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, *79*, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, *127*, 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., Guse, B., Viglione, A., Vorogushyn, S., Macdonald, E., Wietzke, L., & Schumann, A. (2022). Understanding Heavy Tails of Flood Peak Distributions. In *Water Resources Research* (Vol. 58, Issue 6). John Wiley and Sons Inc. https://doi.org/10.1029/2021WR030506

Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change: Stationarity is dead: Whither water management? In *Science* (Vol. 319, Issue 5863, pp. 573–574). https://doi.org/10.1126/science.1151915

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Scheibel, M., Seidel, F., & Wisotzky, C. (2023). A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe - Part 1: Event description and analysis. *Natural Hazards and Earth System Sciences*, *23*(2), 525–551. https://doi.org/10.5194/nhess-23-525-2023

Mudersbach, C., & Jensen, J. (2010). Nonstationary extreme value analysis of annual maximum water levels for designing coastal structures on the German North Sea coastline. *Journal of Flood Risk Management*, *3*(1), 52–62. https://doi.org/10.1111/j.1753-318X.2009.01054.x

Nerantzaki, S. D., & Papalexiou, S. M. (2022). Assessing extremes in hydroclimatology: A review on probabilistic methods. In *Journal of Hydrology* (Vol. 605). Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2021.127302

Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications*, *6*(2), 119–132. https://doi.org/10.1017/S1350482799001103

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. In *Water Resources Research* (Vol. 49, Issue 1, pp. 187–201). https://doi.org/10.1029/2012WR012557

Parmet, B., & Mulders, R. (1999). *Design discharge of the large rivers in The Netherlands-towards a new methodology* (Issue 255). I AHS Publ.

Salas, J. D., & Obeysekera, J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *Journal of Hydrologic Engineering*, *19*(3), 554–568. https://doi.org/10.1061/(asce)he.1943-5584.0000820

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. *Journal of the American Statistical Association*, 1379–1389.

Serinaldi, F., Kilsby, C. G., & Lombardo, F. (2018). Untenable nonstationarity: An assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, *111*, 132–155. https://doi.org/10.1016/j.advwatres.2017.10.015

Slater, L., Villarini, G., Archfield, S., Faulkner, D., Lamb, R., Khouakhi, A., & Yin, J. (2021). Global Changes in 20-Year, 50-Year, and 100-Year River Floods. *Geophysical Research Letters*, *48*(6). https://doi.org/10.1029/2020GL091824

Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1956), 4751–4767. https://doi.org/10.1098/rsta.2011.0161

Slomp, R. (2021). *Flooding in the Meuse river basin in the Netherlands, 2021 Thursday July 15th - Thursday July 22nd, coping with a summer flood, some first impressions*.

Taboga, M. (2021). *"Empirical Distribution" Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://statlect.com/asymptotic-theory/empirical-distribution*.

te Booij, B.-J. (2022). *Trend analysis of extreme precipitation events in the Meuse catchment, obtained with re-forecasts from the ECMWF*.

Tu, M., Hall, M. J., de Laat, P. J. M., & de Wit, M. J. M. (2005a). Extreme floods in the Meuse river over the past century: Aggravated by land-use changes? *Physics and Chemistry of the Earth*, *30*(4-5 SPEC. ISS.), 267–276. https://doi.org/10.1016/j.pce.2004.10.001

Tu, M., Hall, M. J., de Laat, P. J. M., & de Wit, M. J. M. (2005b). Extreme floods in the Meuse river over the past century: Aggravated by land-use changes? *Physics and Chemistry of the Earth*, *30*(4-5 SPEC. ISS.), 267–276. https://doi.org/10.1016/j.pce.2004.10.001

van der Wiel, K., Wanders, N., Selten, F. M., & Bierkens, M. F. P. (2019). Added Value of Large Ensemble Simulations for Assessing Extreme River Discharge in a 2 °C Warmer World. *Geophysical Research Letters*, *46*(4), 2093–2102. https://doi.org/10.1029/2019GL081967

Villarini, G., Smith, J. A., Serinaldi, F., & Ntelekos, A. A. (2011). Analyses of seasonal and annual maximum daily discharge records for central Europe. *Journal of Hydrology*, *399*(3–4), 299–312. https://doi.org/10.1016/j.jhydrol.2011.01.007

Virtanen, P. and G., Ralf and Oliphant, Travis E. and Haberland, Matt and Reddy, Tyler and Cournapeau, & David and Buroviski. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*.

Vivekanandan, N. (2015). Flood frequency analysis using method of moments and L-moments of probability distributions. *Cogent Engineering*, *2*(1). https://doi.org/10.1080/23311916.2015.1018704

Vorogushyn, S., Apel, H., Kemter, M., & Thieken, A. (2022). *Analyse der Hochwassergefahrdung im Ahrtal unter Berucksichtigung historischer Hochwasser*.

World Health Organization. (2023). *Floods*.

APPENDIX A: DATA CURATION

Three processes were performed to filter out discharge stations inadequate for this study including:

- Filtering out stations with less than 20 years of observations;
- Plotting the time series and inspecting satellite imagery to remove stations that may be impacted by nearby hydraulic structures; and
- Inspection of the name, river, and basin data fields for each station to identify stations along canals.

Filtering out stations with less than 20 years of observations

Out of the 192 discharge stations within the Meuse, 141 have more than 20 years of observations.

Plotting the time series and inspecting satellite imagery

The time series of 141 discharge stations with more than 20 years of observations were plotted to identify any irregularities that would warrant removing the station from the study. A brief summary of the stations removed after plotting the time series is provided below.

The time series of Station 5771, shown in Figure 46, exhibits irregular flow patterns which appeared to be human influenced. The river field for this station, Canal Albert, confirmed that this station measures the discharge along a canal. Therefore, this station was removed from this study. To ensure that no other stations along canals are included in this study, the name, river, and basin fields for each station are carefully examined as mentioned in the following section.

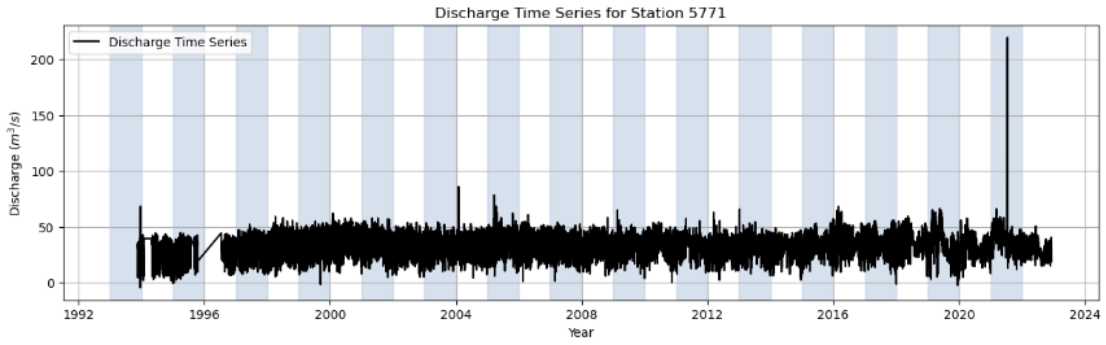


Figure 46: Discharge Time Series for Station 5771

Inspection of satellite imagery showed that stations 7132, 7831, 8017, 8022, 9214 are located nearby hydraulic structures. The satellite imagery and time series of these stations are shown in Figure 47 through Figure 56.

Figure 47: Satellite Imagery near Station 7132

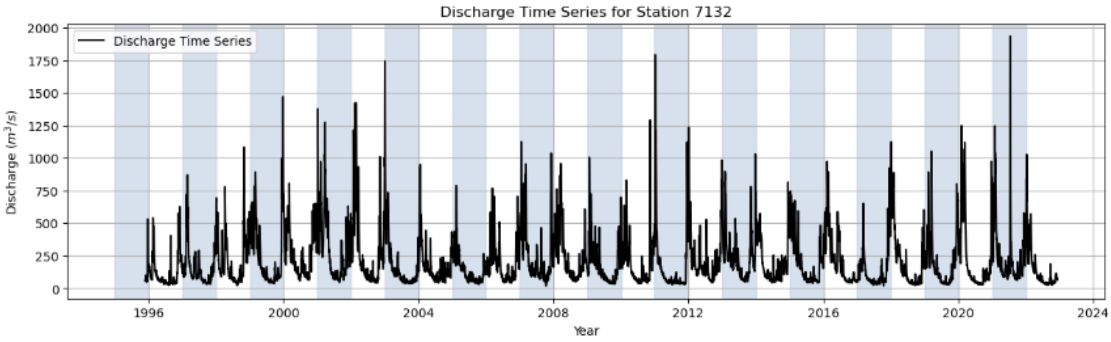


Figure 48: Discharge Time Series for Station 7132

Figure 49: Satellite Imagery for Station 7831

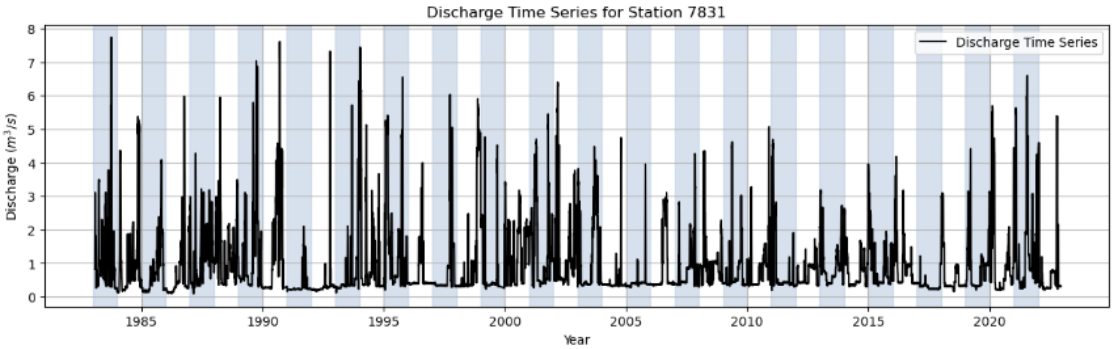


Figure 50: Discharge Time Series for Station 7831

Figure 51: Satellite Imagery for Station 8017

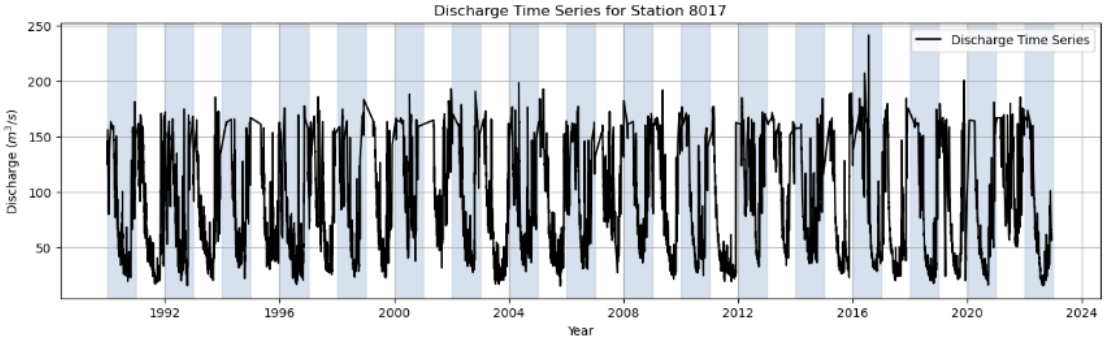


Figure 52: Discharge Time Series for Station 8017

Figure 53: Satellite Imagery for Station 8022

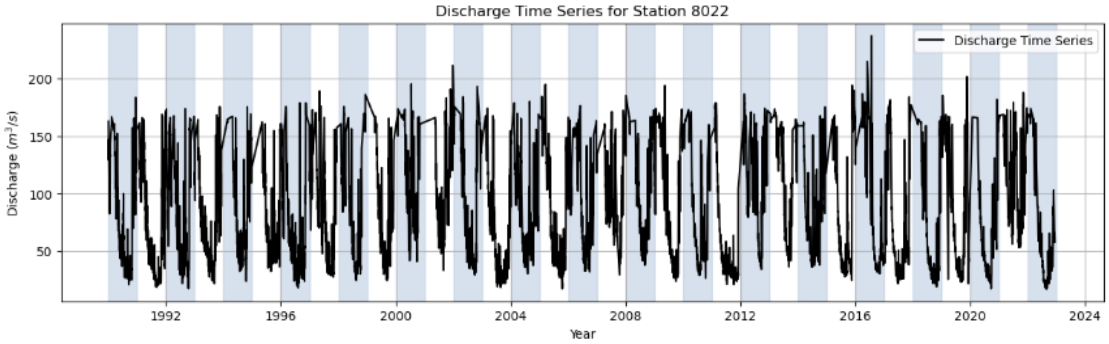


Figure 54: Discharge Time Series for Station 8022

Figure 55: Satellite Imagery for Station 9214

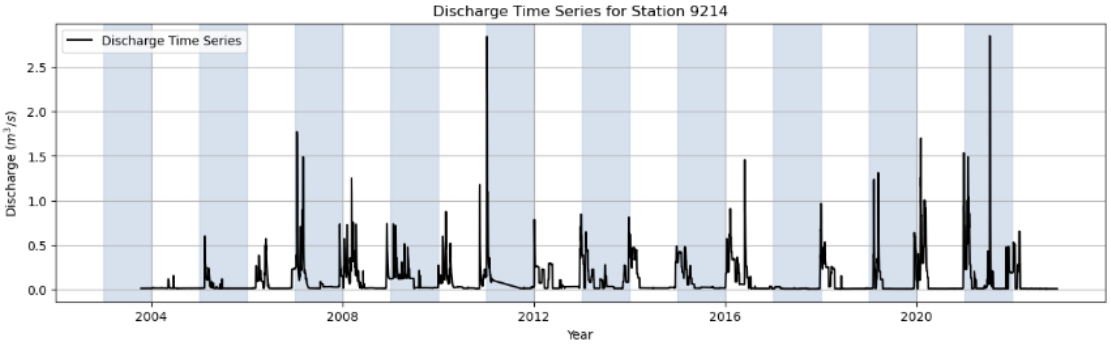


Figure 56: Discharge Time Series for Station 9214

There were a few stations with irregular flow patterns, but trees made it difficult to determine whether or not there were structures in these areas. Due to the irregular flow patterns stations 6220, 6340, and 6440 were removed. The satellite imagery and time series for these stations are provided in Figure 57 through Figure 62.

Figure 57: Satellite Imagery near station 6220

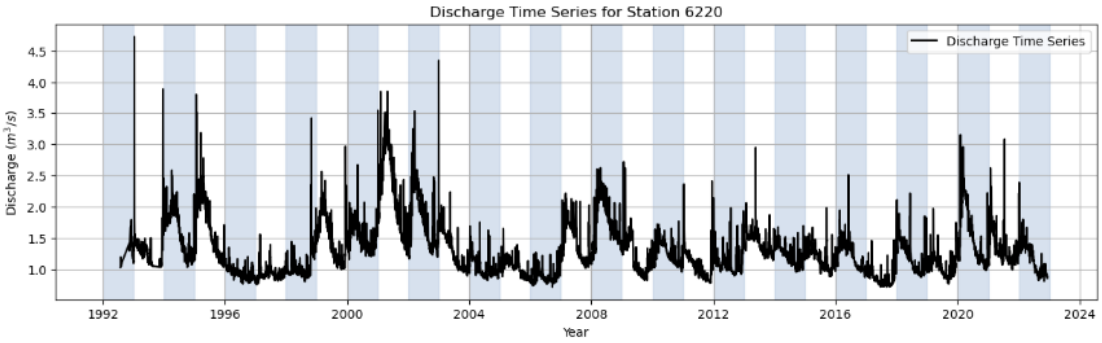


Figure 58: Discharge Time Series for Station 6220

Figure 59: Satellite Imagery near Station 6340

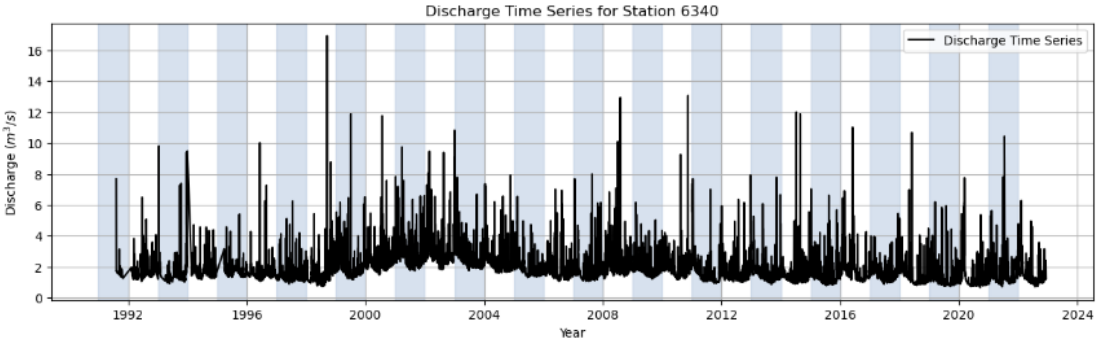


Figure 60: Discharge Time Series for Station 6340

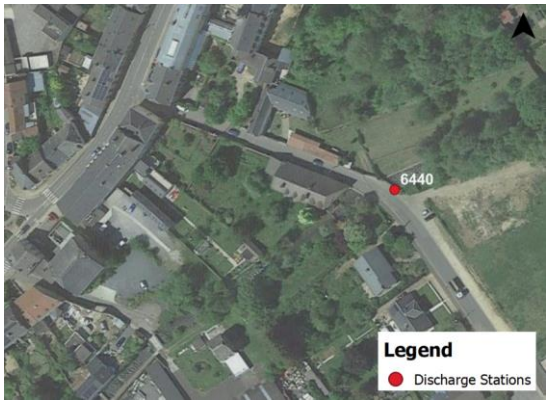


Figure 61: Satellite Imagery near Station 6440

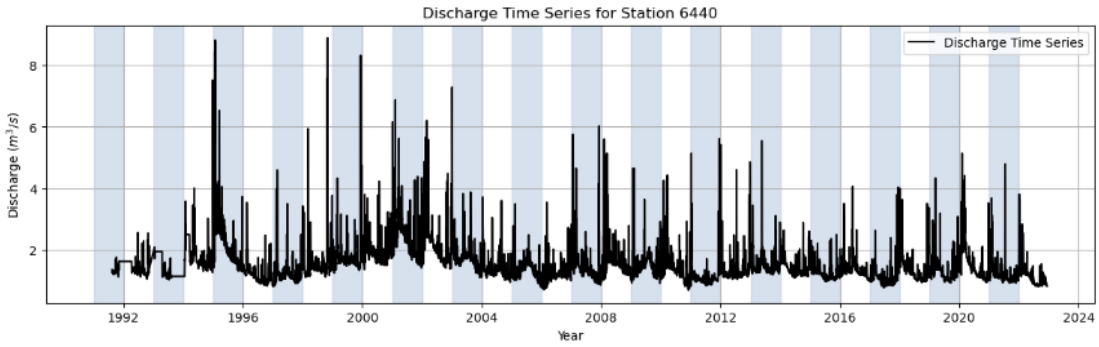


Figure 62: Discharge Time Series for Station 6440

During inspection of the discharge time series, it was discovered that station 5820 had a change in base flow around 2003. The time series for this station, showing this change in base flow is shown in Figure 63.

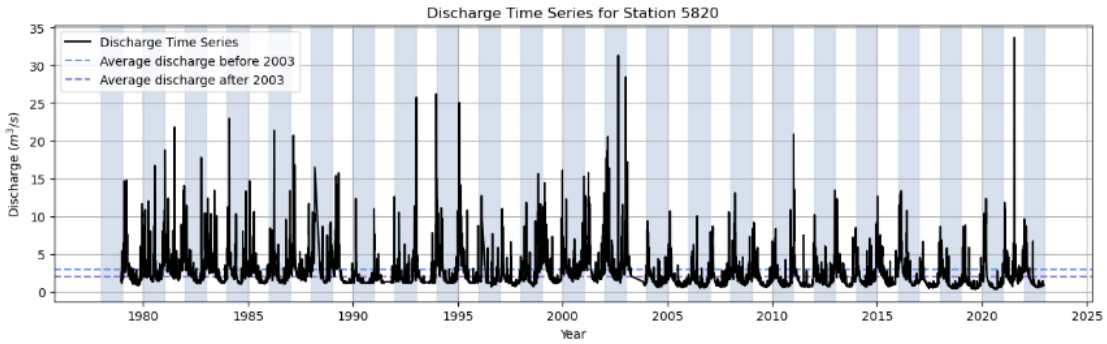


Figure 63: Discharge Time Series for Station 5820

Inspection of the name, river, and basin data fields

To ensure no other canals are included in this study, the name, river, and basin fields for each station are carefully examined. No other stations along canals are found.

Summary

A summary of the justifications for discharge stations removed from this study is provided in Table 16.

Table 16: Summary of Stations (with more than 20 years) Removed from Study

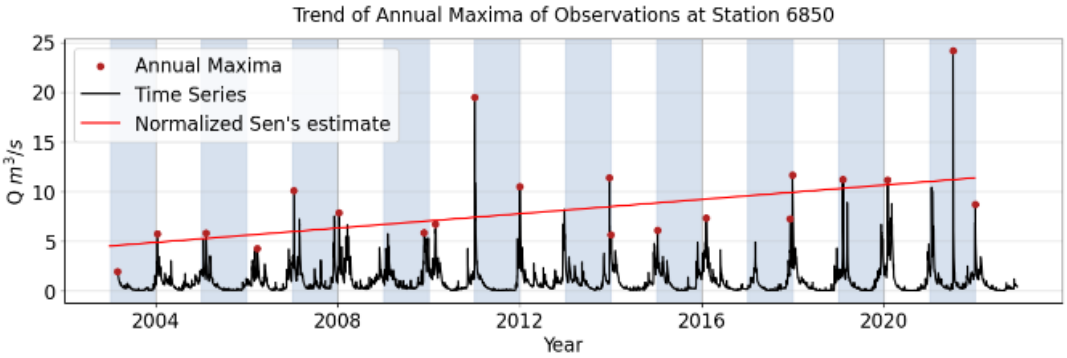
Station ID	Reason for Removing
5771	Irregular flow pattern, Located along canal
7132	Flow influenced by nearby hydraulic structure
7831	Flow influenced by nearby hydraulic structure
8017	Flow influenced by nearby hydraulic structure
8022	Flow influenced by nearby hydraulic structure
9214	Flow influenced by nearby hydraulic structure
6220	Irregular flow pattern
6340	Irregular flow pattern
6440	Irregular flow pattern
5820	Change in base flow

APPENDIX B: NONSTATIONARITY OF OBSERVED DISCHARGE

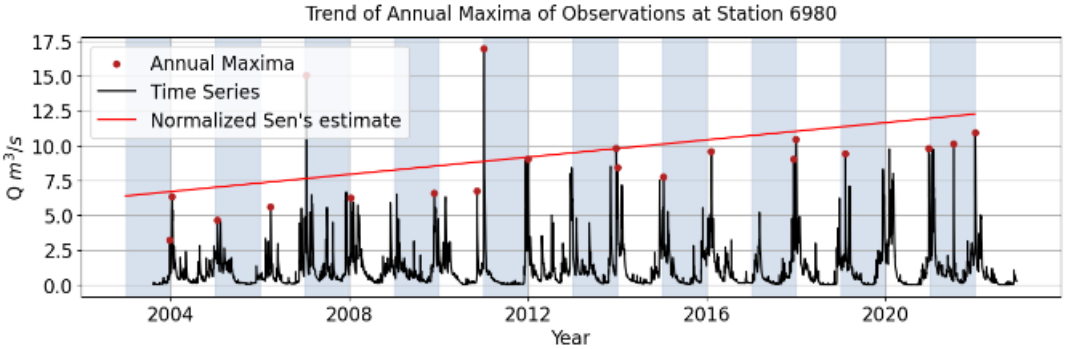
B.1. MANN KENDALL AND SEN'S SLOPE RESULTS FOR DISCHARGE

Table 17: Trend in AM Observed Discharge

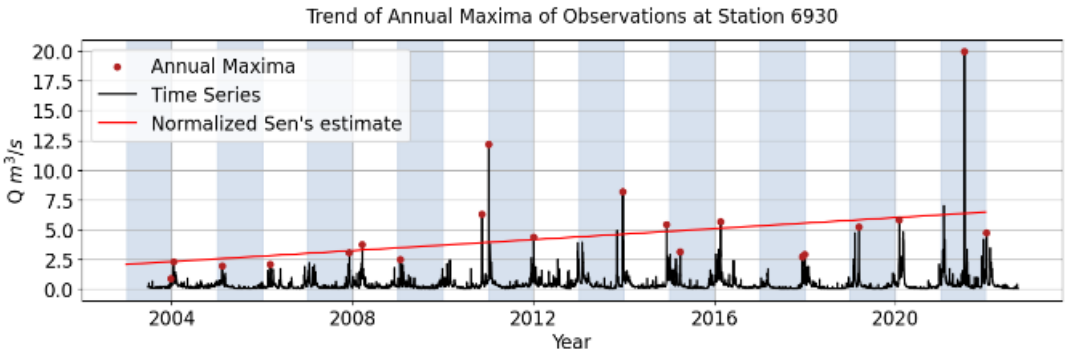
Level of Significance, α = 0.1				
	Number of Stations	Average Years Available at Each Station with this trend	Maximum Years Available at Each Station with this trend	Minimum Years Available at Each Station with this trend
Increasing Trend	7	34	55	20
Decreasing Trend	2	43	49	38
No Trend	121	36	56	20
Level of Significance, α = 0.05				
	Number of Stations	Average Years Available at Each Station with this trend	Maximum Years Available at Each Station with this trend	Minimum Years Available at Each Station with this trend
Increasing Trend	5	34	55	20
Decreasing Trend	0	--	--	--
No Trend	125	36	56	20
Level of Significance, α = 0.01				
	Number of Stations	Average Years Available at Each Station with this trend	Maximum Years Available at Each Station with this trend	Minimum Years Available at Each Station with this trend
Increasing Trend	3	20	20	20
Decreasing Trend	0	--	--	--
No Trend	127	36	56	20



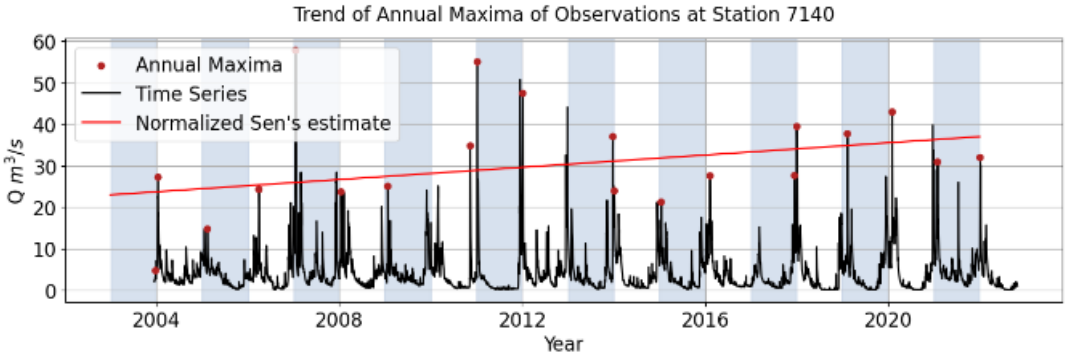
Mann Kendall test concluded with 99% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.42 mm/day.



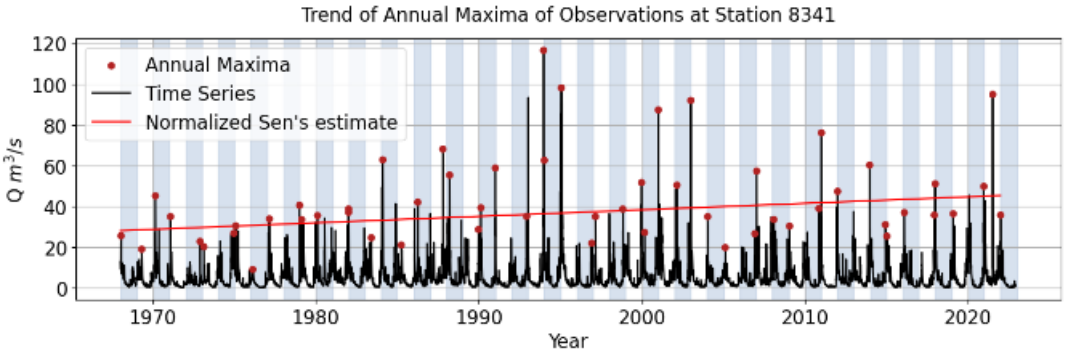
Mann Kendall test concluded with 99% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.41 mm/day.



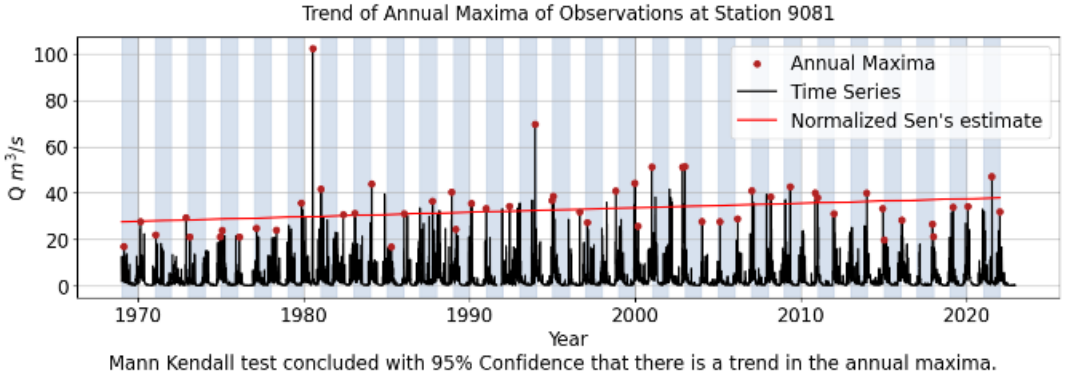
Mann Kendall test concluded with 99% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.35 mm/day.



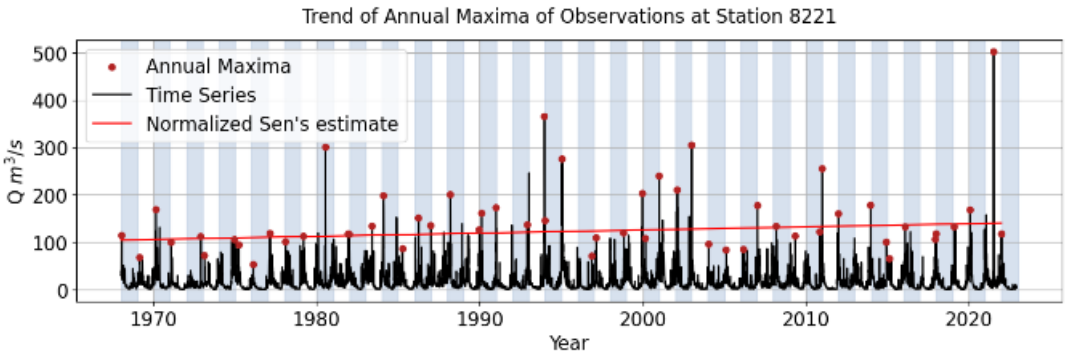
Mann Kendall test concluded with 90% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.29 mm/day.



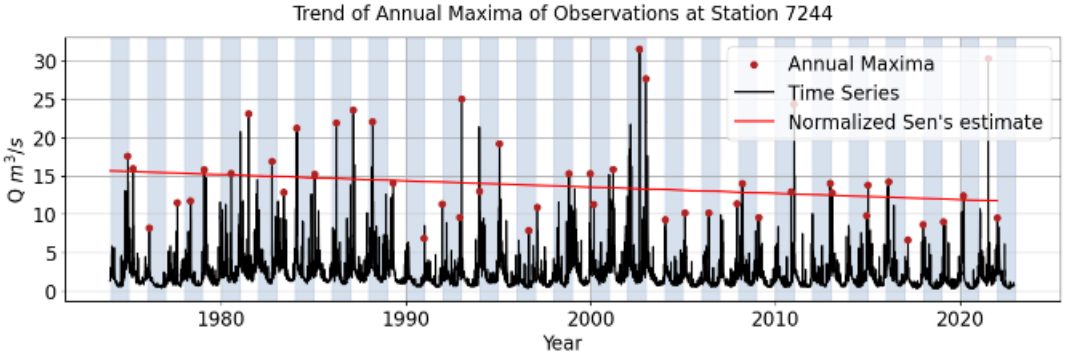
Mann Kendall test concluded with 95% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.09 mm/day.



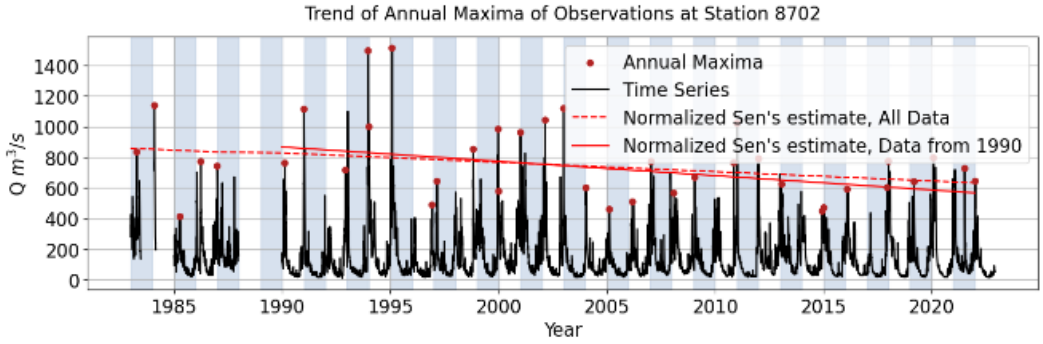
Mann Kendall test concluded with 95% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.07 mm/day.



Mann Kendall test concluded with 90% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is 0.04 mm/day.



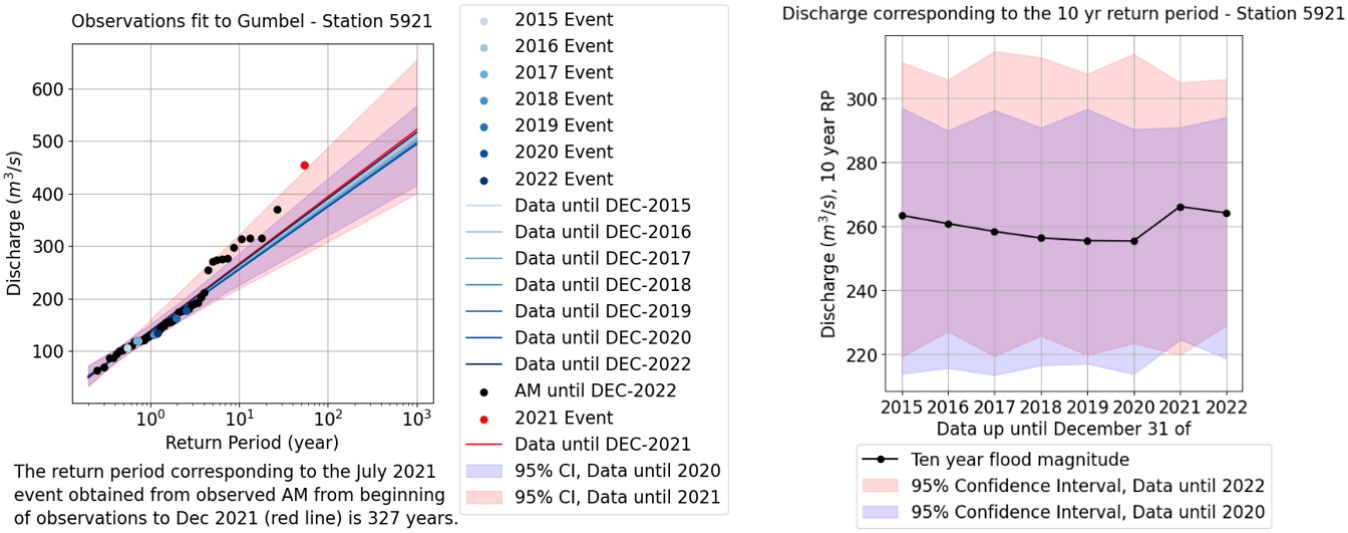
Mann Kendall test concluded with 90% Confidence that there is a trend in the annual maxima.
The normalized Sen's Slope, which indicates the magnitude of the trend, is -0.02 mm/day.



Mann Kendall test concluded with 90% Confidence that there is a trend in the annual maxima.
Using all data, the normalized Sen's Slope, which indicates the magnitude of the trend, is -0.05 mm/day.
Using data from 1990, the normalized Sen's Slope, which indicates the magnitude of the trend, is -0.08 mm/day.

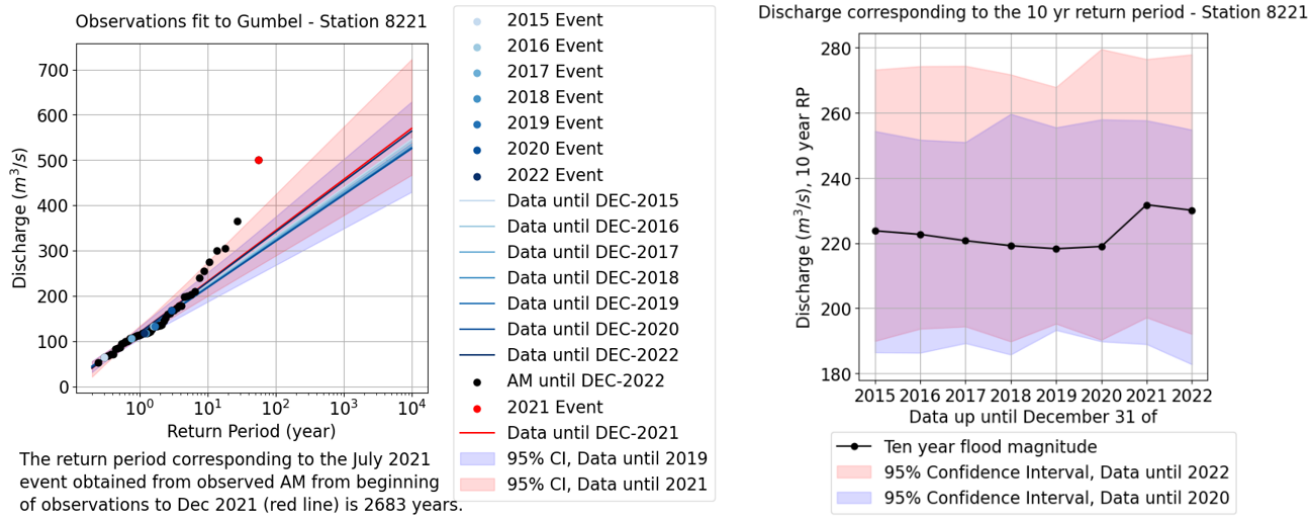
B.2. OBSERVED CHANGES IN RIVER DISCHARGES SINCE 2015

The eight sets of AM fit to the Gumbel distribution for each station of interest are presented in Figure 64 through Figure 69.



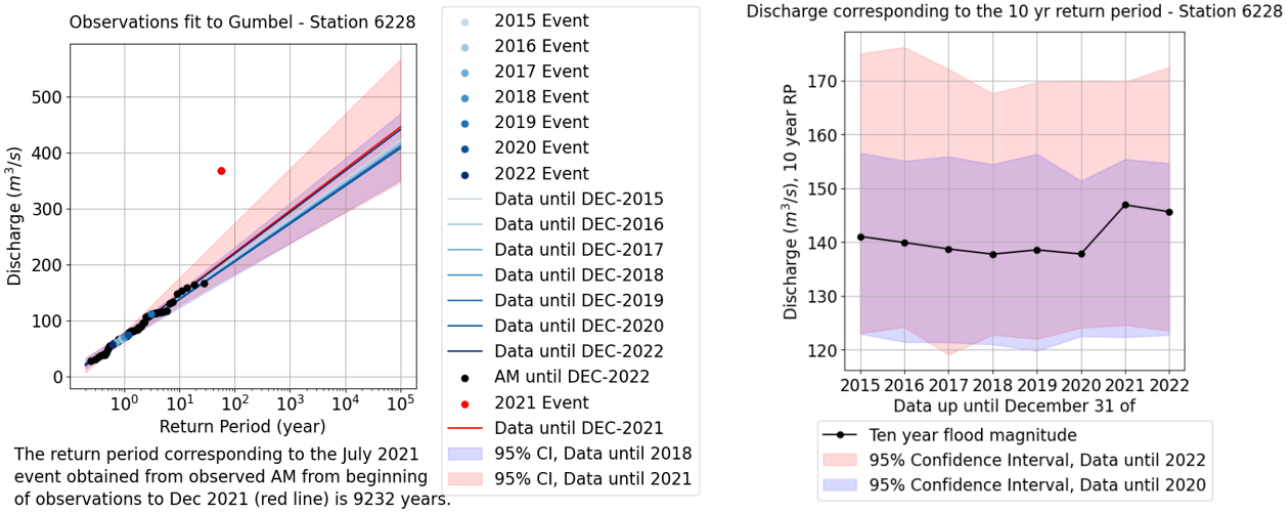
The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 327 years.

Figure 64: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 5921



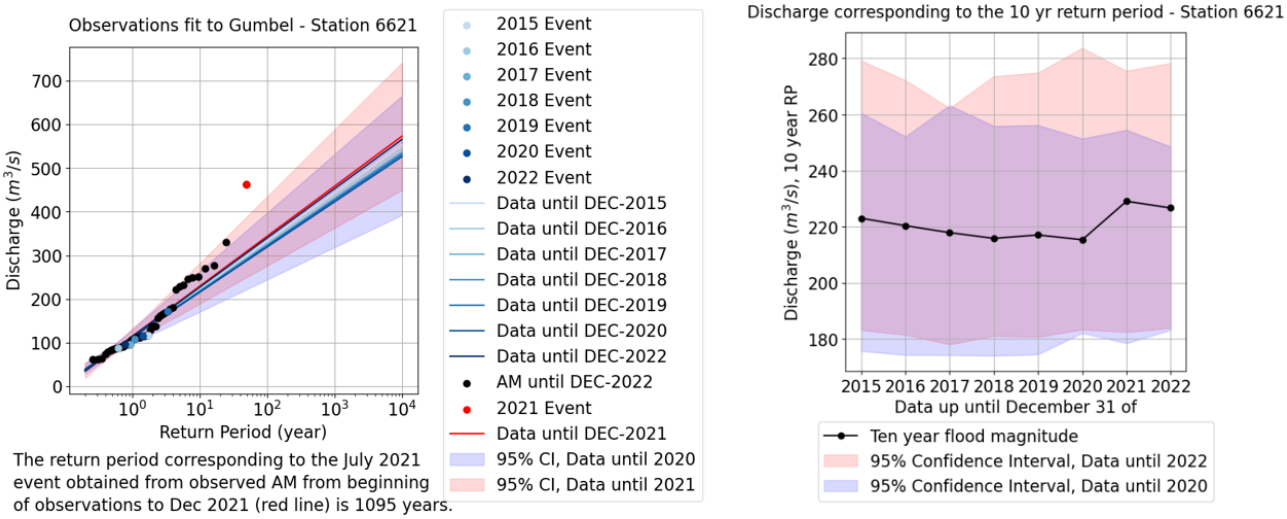
The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 2683 years.

Figure 65: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 8221



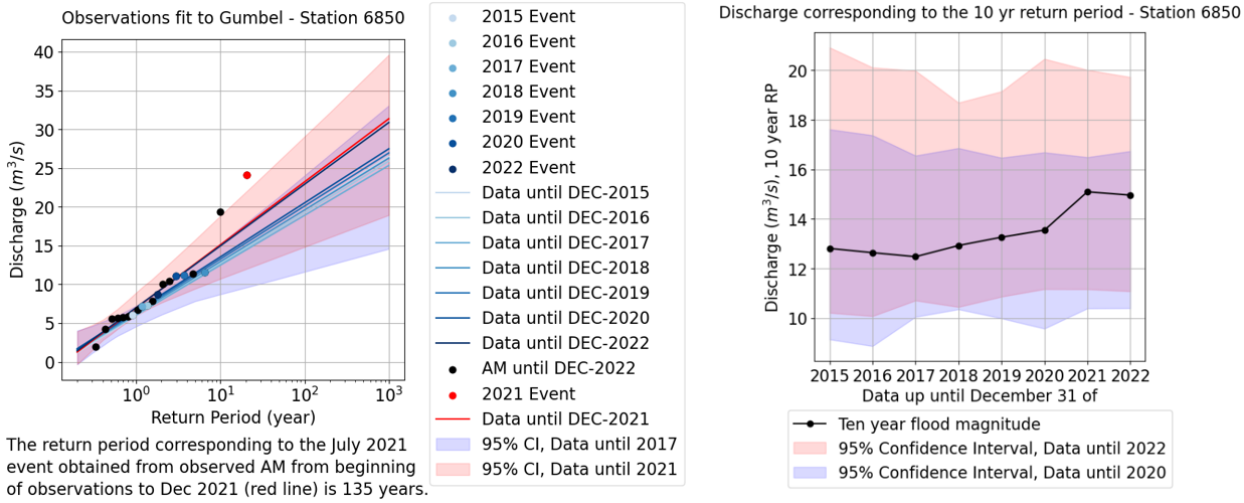
The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 9232 years.

Figure 66: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 6228



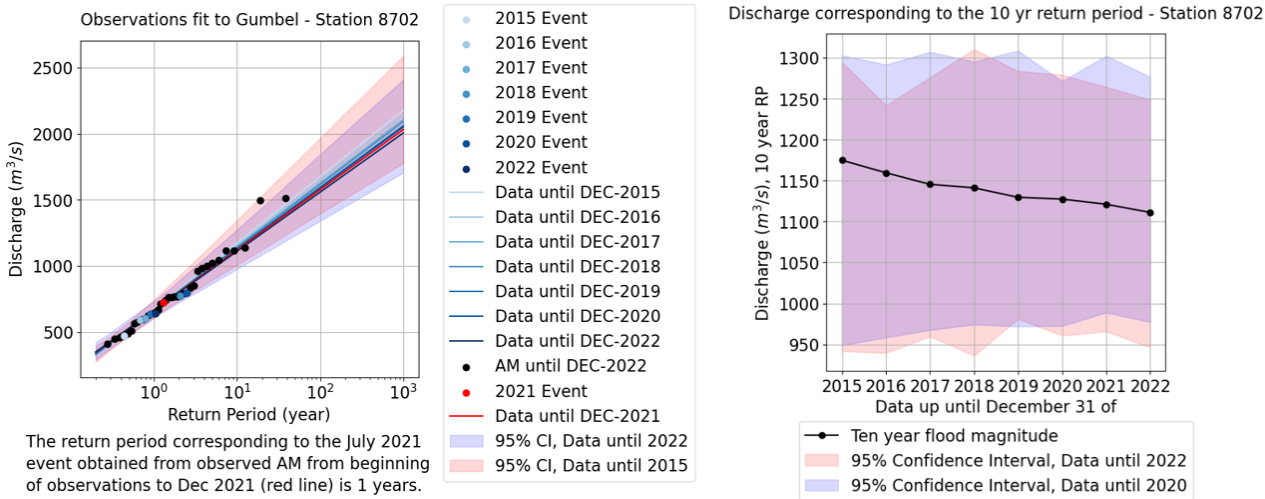
The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 1095 years.

Figure 67: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 6621



The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 135 years.

Figure 68: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 6850



The return period corresponding to the July 2021 event obtained from observed AM from beginning of observations to Dec 2021 (red line) is 1 years.

Figure 69: Discharge Frequency Curve (Left) and 10-year Discharges (Right) for Station 8702

B.3. EVOLUTION OF DISTRIBUTION PARAMETERS AND RETURN LEVELS AT STATIONS WITH A TREND IN OBSERVED ANNUAL MAXIMUM DISCHARGE

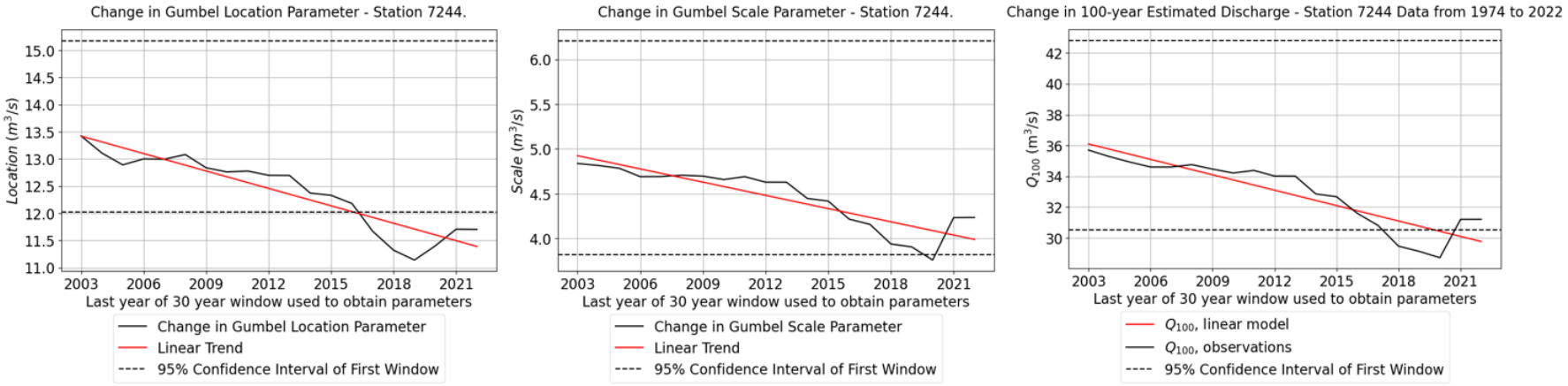


Figure 70: Evolution of Gumbel Parameters and Estimated Return Levels for Station 7244 using 30-year Sliding Window

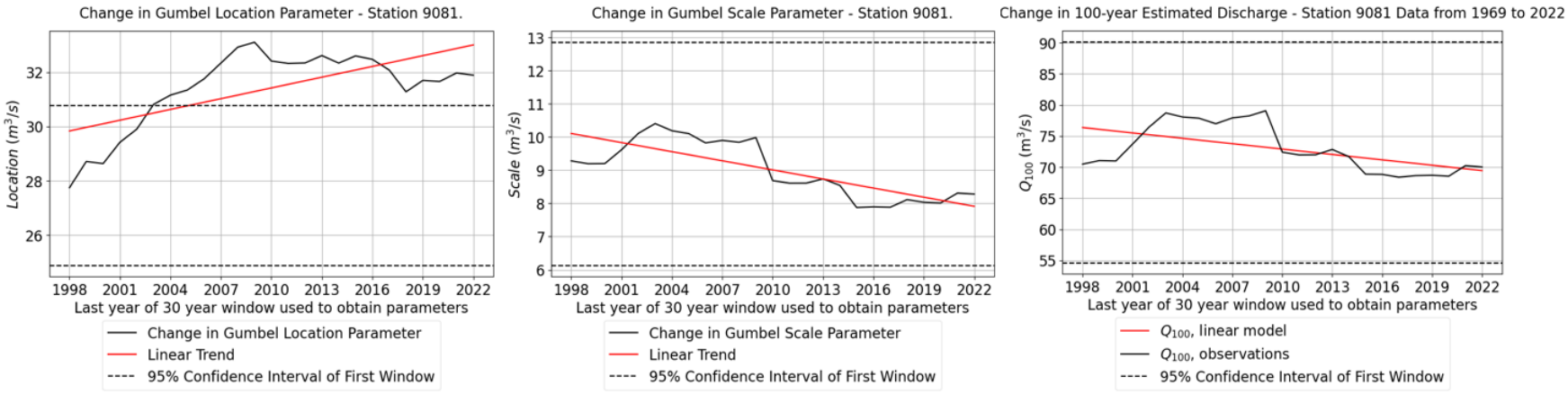


Figure 71: Evolution of Gumbel Parameters and Estimated Return Levels for Station 9081 using 30-year Sliding Window

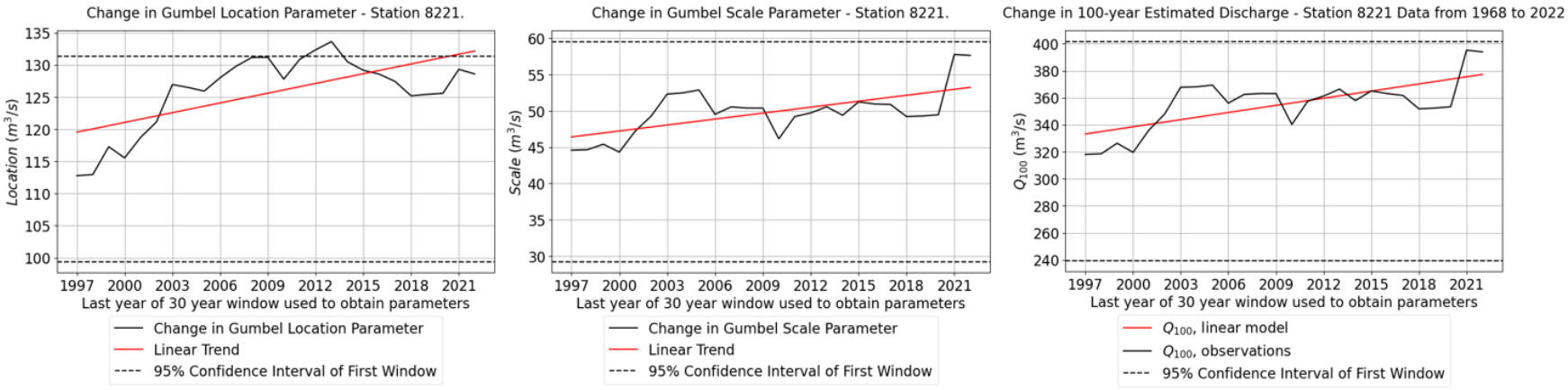


Figure 72: Evolution of Gumbel Parameters and Estimated Return Levels for Station 8211 using 30-year Sliding Window

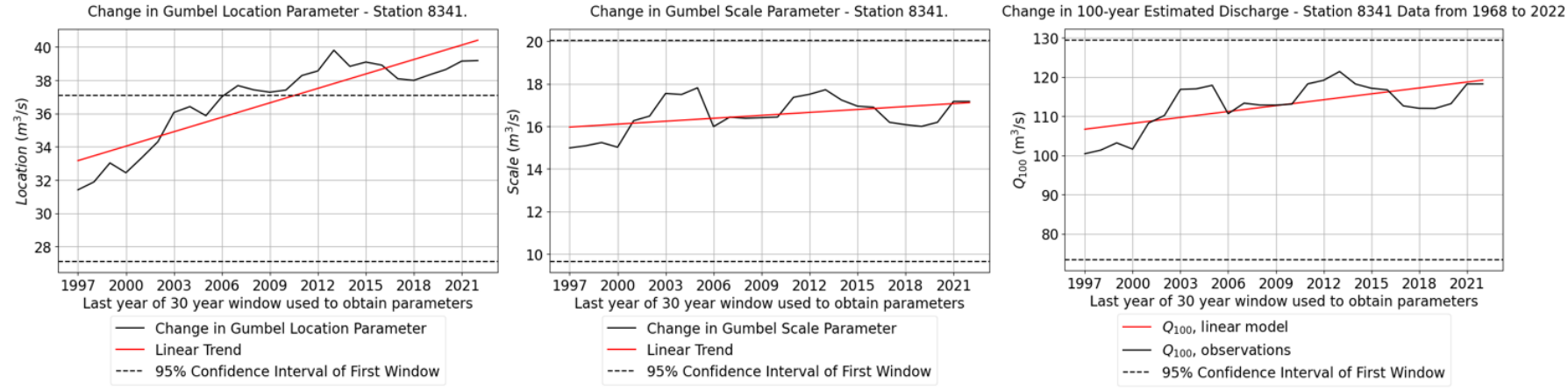


Figure 73: Evolution of Gumbel Parameters and Estimated Return Levels for Station 8341 using 30-year Sliding Window

APPENDIX C: EXTREME VALUE MODELS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

C.1. EXTREME VALUE DISTRIBUTIONS

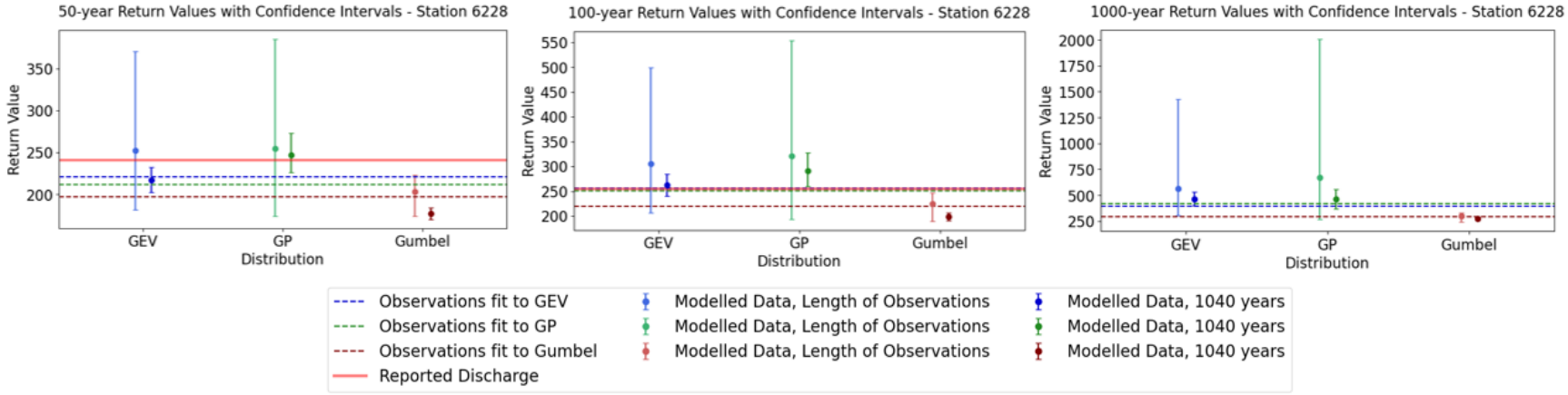


Figure 74: Extreme Value Distributions - Station 6228

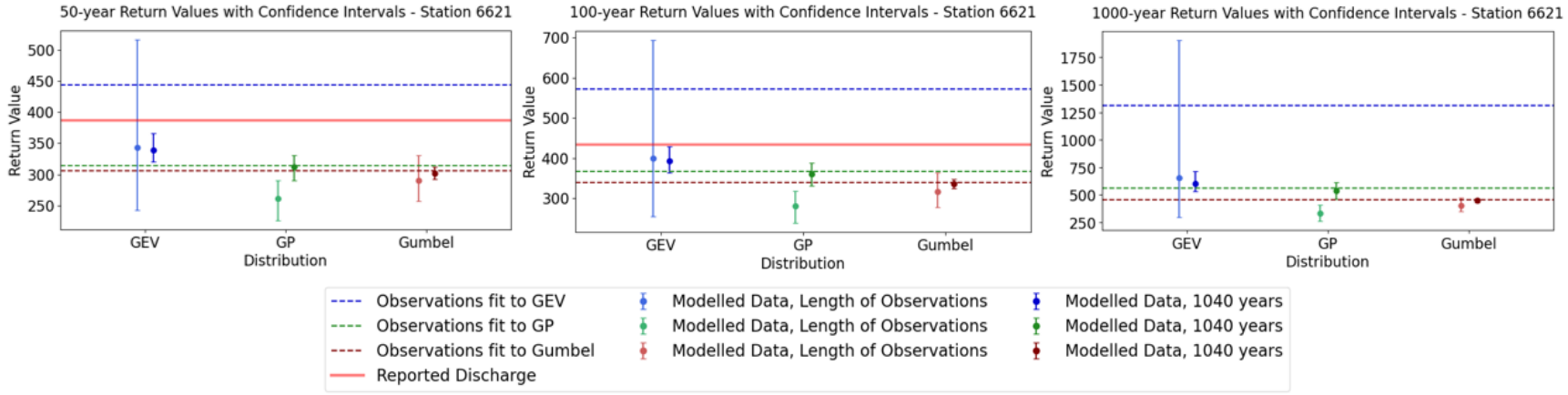


Figure 75: Extreme Value Distributions - Station 6621

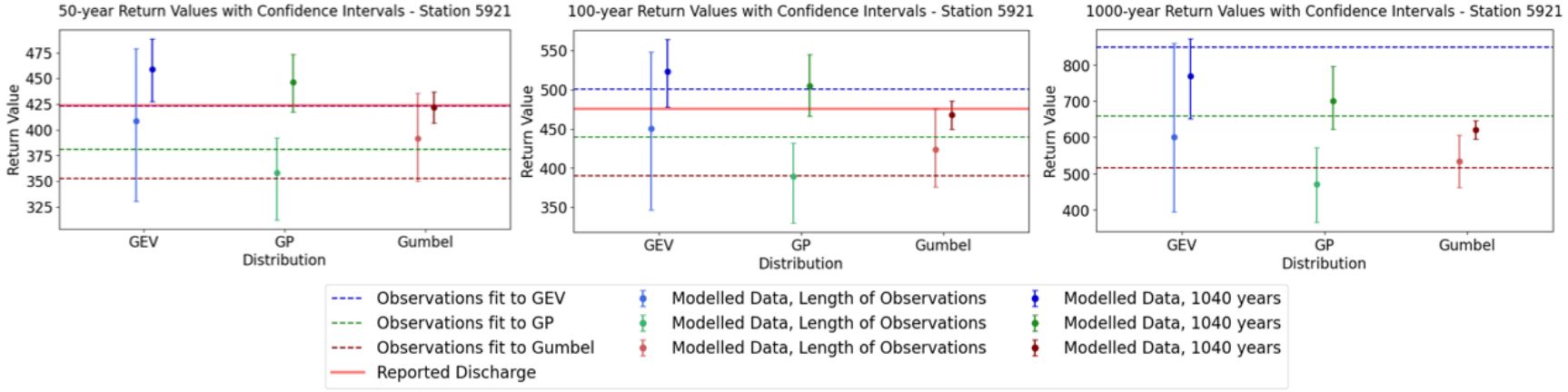


Figure 76: Extreme Value Distributions - Station 5921

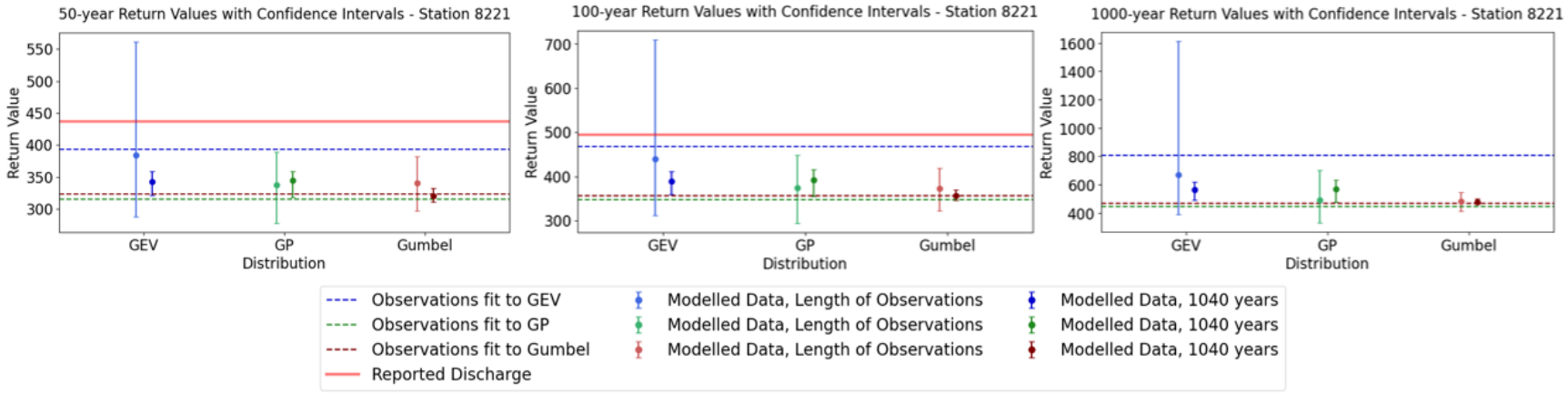


Figure 77: Extreme Value Distributions - Station 8221

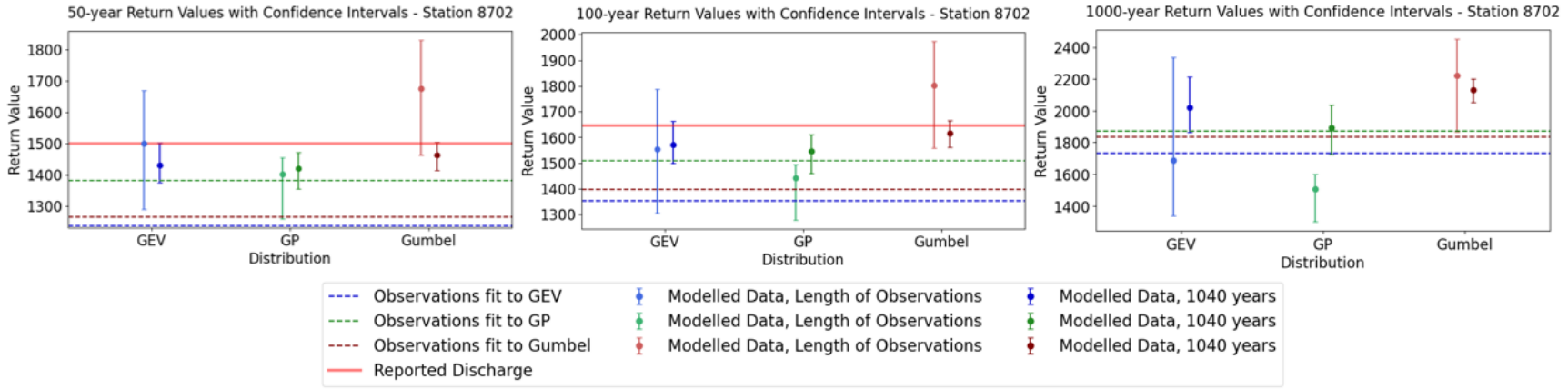


Figure 78: Extreme Value Distributions - Station 8702

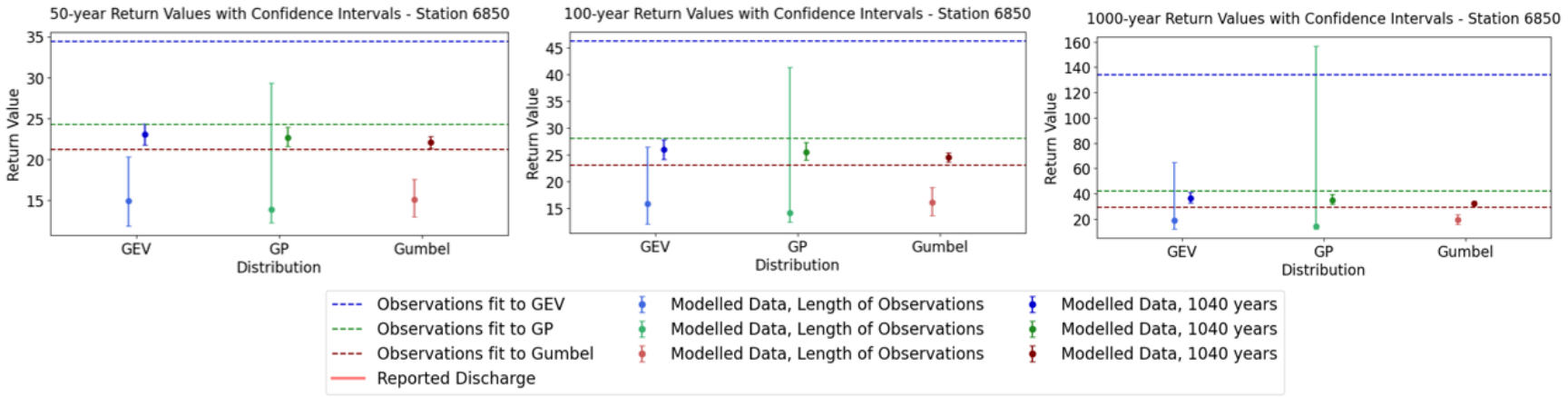


Figure 79: Extreme Value Distributions - Station 6850

The percent differences in 100-year estimated discharges obtained from **observed** and **modelled** data are presented in Table 18. For example, at Station 5921 GEV estimates obtained from modelled data the same length as observations were ten percent lower than GEV estimates obtained from observations. However, as will be discussed later on in this section, the best fit model for observations, modelled data resampled to be the same length as observations, and 1,040 years of modelled data varies which partially contributes to the differences in estimates between the datasets. Therefore, the percent difference of the 100-year discharge obtained from the best fit model is also provided.

Table 18: Percent Difference in 100-year Estimated Discharges Obtained from Observed and Modelled Data

Station	GEV		GP		Gumbel		Best Fit Model[2]	
	Modelled, Length of Obs.[1]	Modelled, 1040-years	Modelled, Length of Obs.	Modelled, 1040-years	Modelled, Length of Obs.	Modelled, 1040-years	Modelled, Length of Obs.	Modelled, 1040-years
8221	-6%	-17%	-15%	-12%	+4%	0%	-6%	-16%
5921	-10%	+4%	-13%	+11%	+8%	+17%	-10%	+4%
6228	+16%	+2%	+22%	+12%	+2%	-10%	+16%	+10%
6621	-41%	-31%	-23%	-1%	-7%	-1%	-41%	-31%
8702	+13%	+14%	-4%	+2%	+22%	+14%	+13%	+12%
6850	-65%	-43%	-50%	-7%	-30%	+4%	-43%	-7%

Notes:
1. Percent differences between observations and modelled data, length of observations is shown to give an idea of how much estimates obtained from modelled data randomly sampled to be the same length as observations can vary from those obtained from observations.
2. The percent difference between estimates is calculated using the 100-year discharge from the best fit model for each dataset (the best fit model between observations, modelled data the length of observations, 1,040-years of modelled data varies).
3. A positive percentage indicates that estimates obtained from modelled data, either length of observations or 1,040-years, are higher than those obtained from observed data.

Percent differences between **GEV** (or **GP**) and **Gumbel** estimates obtained from **1,040 years of modelled data** are shown in Table 19. At the stations of interest GEV and Gumbel estimates of Q_{100} vary between 3 and 24 percent and GP and Gumbel estimates vary between 4 and 30 percent. Additionally, the difference between estimates increases with increasing RP.

Table 19: Percent Difference between GEV (or GP) and Gumbel Estimates obtained from 1,040 years of Modelled Data

Station	Percent Difference from Gumbel Estimate					
	$Q_{GEV,50}$	$Q_{GEV,100}$	$Q_{GEV,1000}$	$Q_{GP,50}$	$Q_{GP,100}$	$Q_{GP,1000}$
8221	+6%	+8%	+15%	+6%	+9%	+15%
5921	+8%	+11%	+19%	+5%	+7%	+11%
6228[1]	+18%	+24%	+42%	+28%	+30%	+35%
6621	+11%	+14%	+26%	+3%	+7%	+16%
8702[2]	-2%	-3%	-5%	-3%	-4%	-11%
6850	+4%	+8%	+8%	+4%	+8%	+6%

Notes:
1. Gumbel was not a good fit for the 1,040 years of modelled data at Station 6228.
2. Station 8702 exhibited light tail behavior.

C.1.1. THRESHOLD ESTIMATION FOR GP DISTRIBUTION

Table 20: Thresholds for Generalized Pareto Distribution

Station ID	Threshold		
	Observations	Modelled Data (Length of Obs.)	Modelled Data (1,040 years)
6228	60	60	60
6621	60	65	65
5921	100	100	120
8221	100	100	94
8702	400	400	470
6850	2.5	6	6

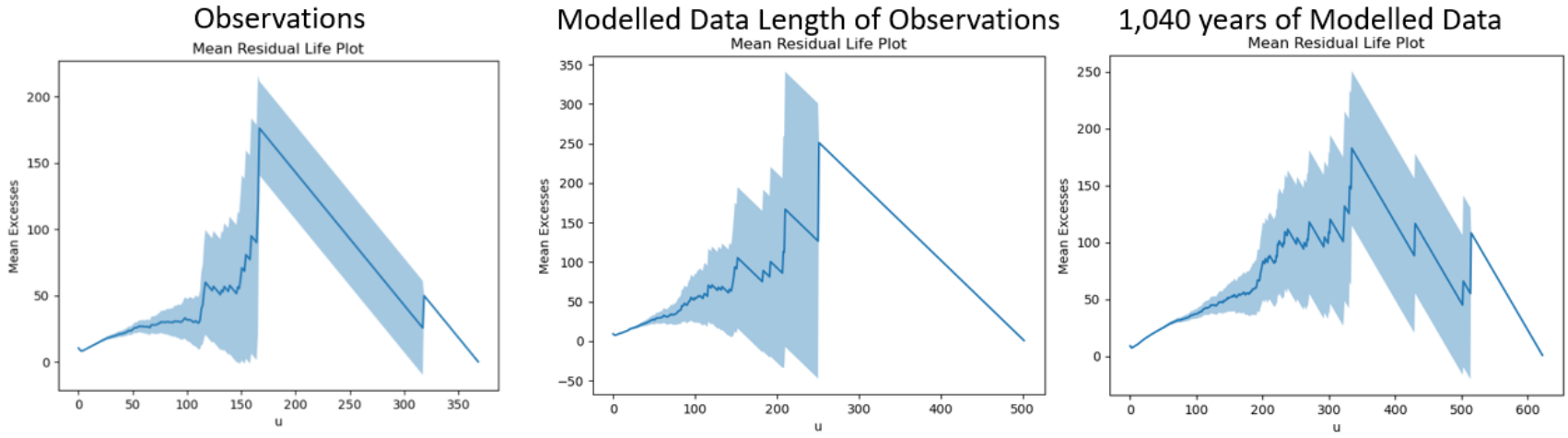


Figure 80: MRLP for Station 6228

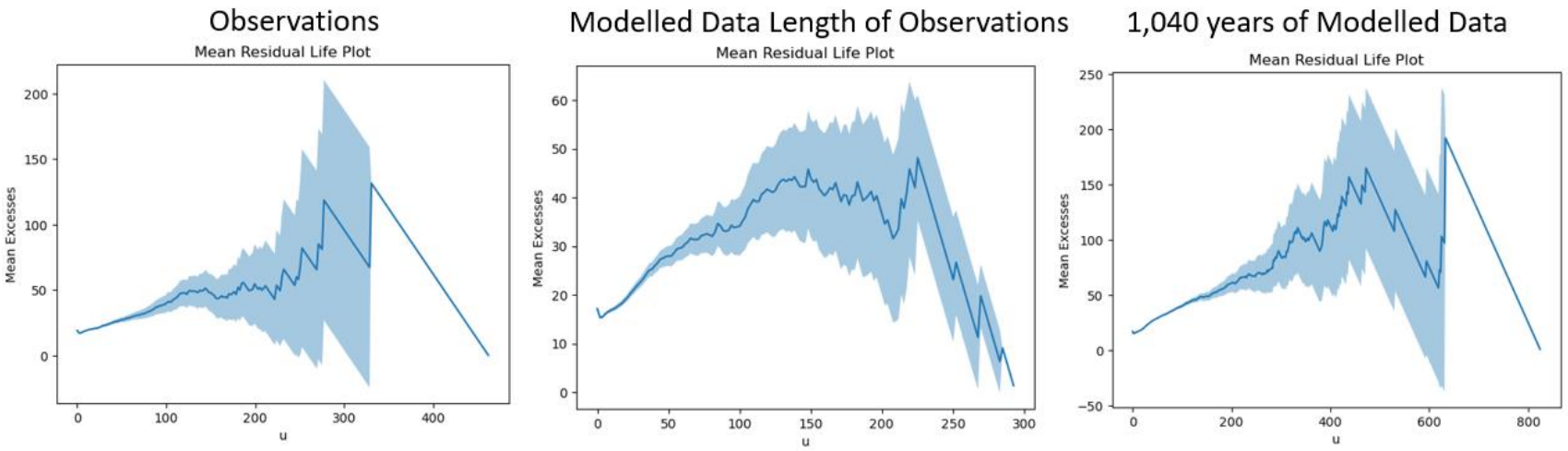


Figure 81: MRLP for Station 6621

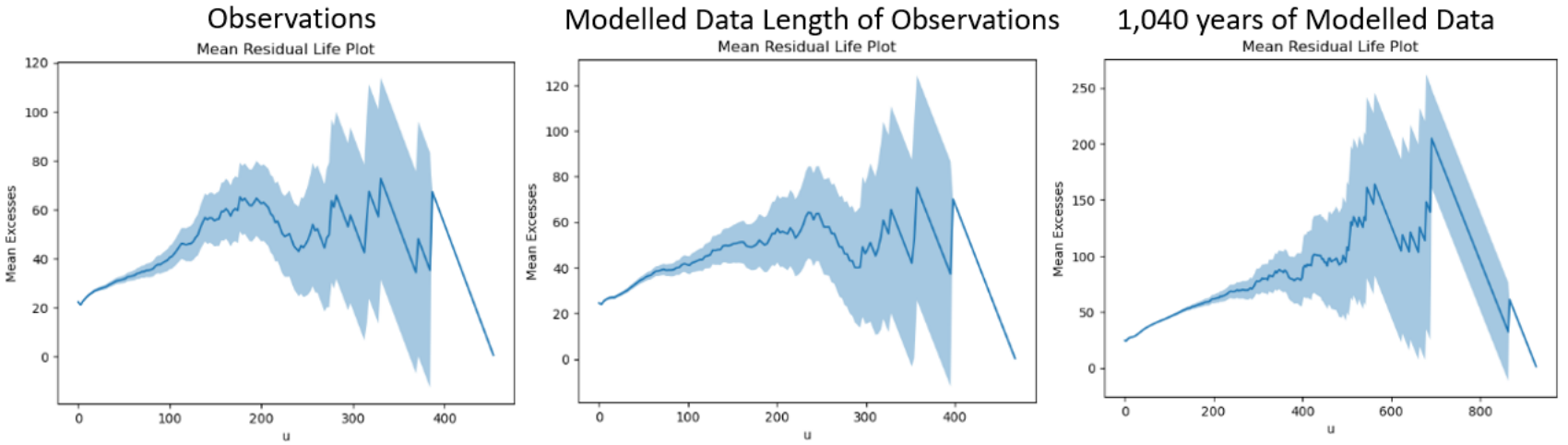


Figure 82: MRLP for Station 5921

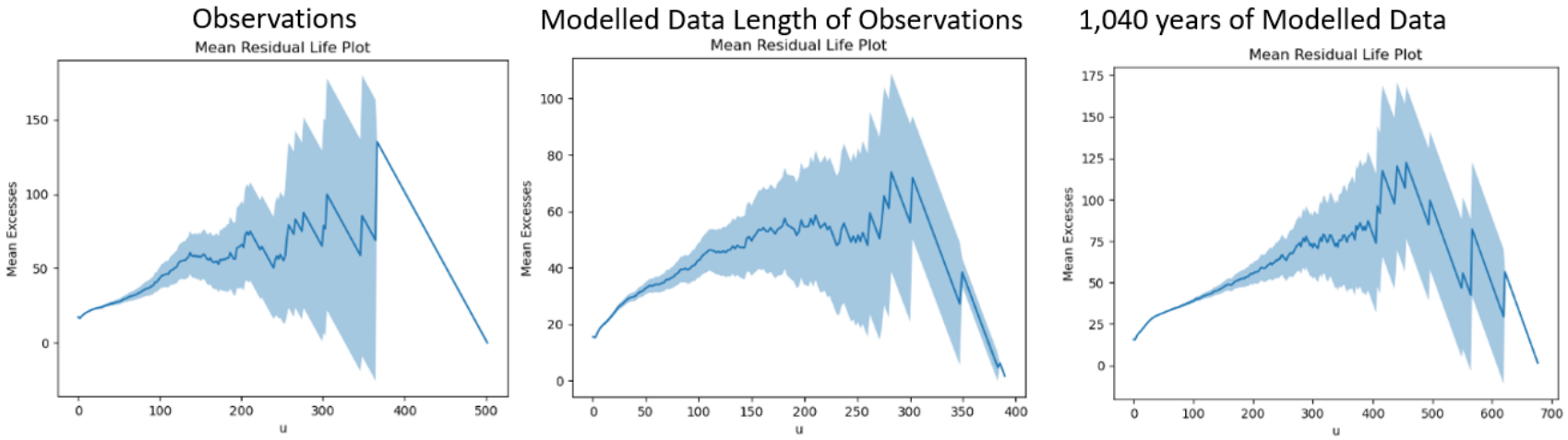


Figure 83: MRLP for Station 8221

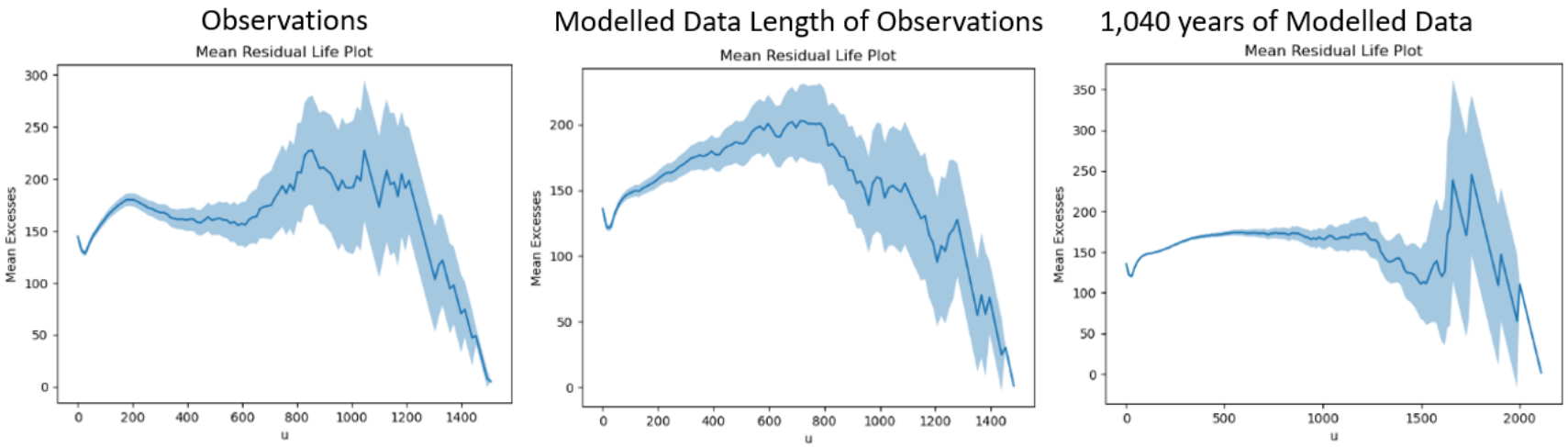


Figure 84: MRLP for Station 8702

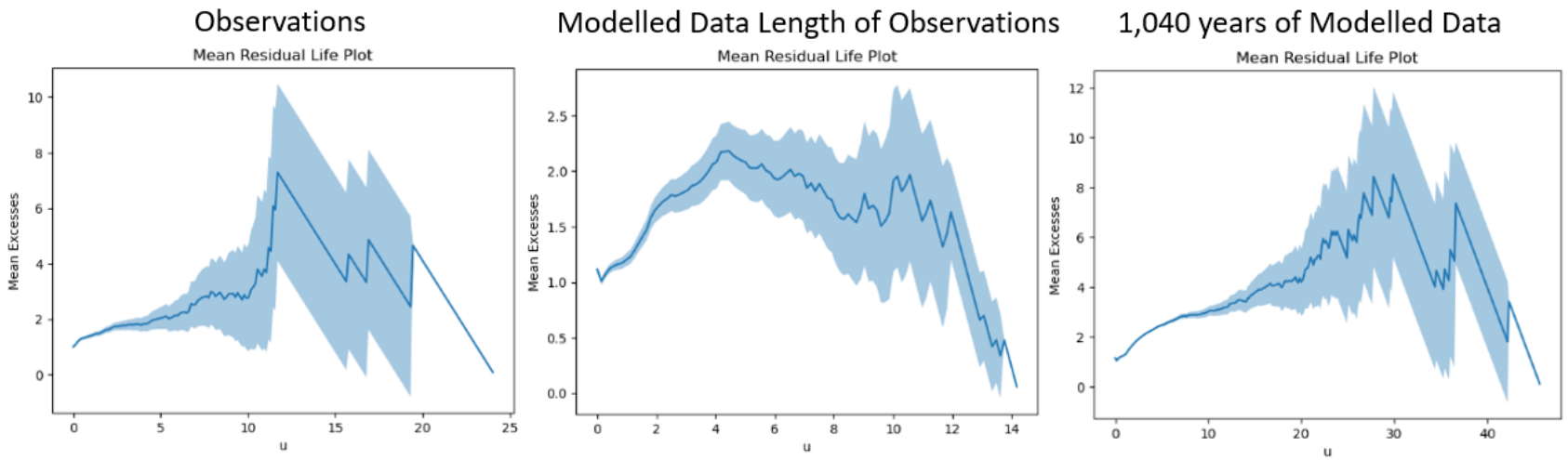


Figure 85: MRLP for Station 6850

C.1.2. GOODNESS OF FIT

At all stations included in this analysis, the Q-Q plot and Kolmogorov Smirnov tests, introduced in Section 2.4, indicated that all three distributions were a good fit for all three datasets except for Station 6228. The Kolmogorov Smirnov test indicated that the Gumbel distribution was not a good fit for the 1,040 years of modelled data at Station 6228. The alpha value of the KS test for the Gumbel distribution was 0.04979, therefore, the null hypothesis that the data fits the Gumbel distribution is rejected. The Q-Q plot is shown in Figure 86.

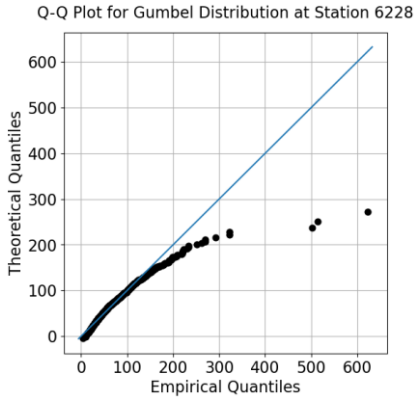
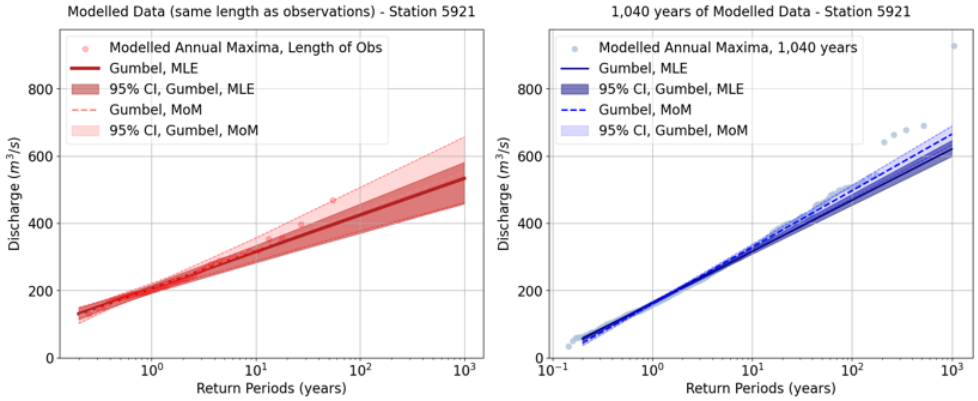
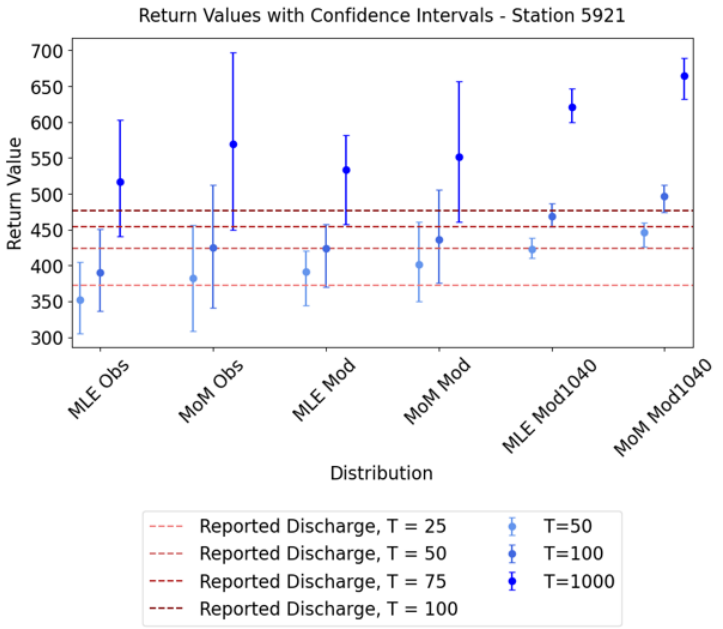


Figure 86: Gumbel Q-Q plot for Station 6228

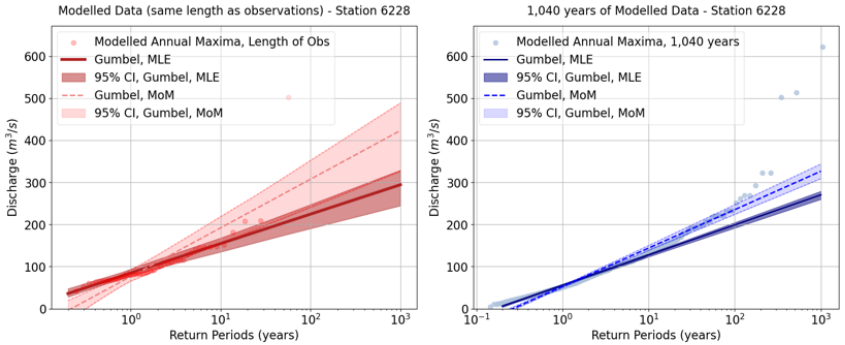
C.2. PARAMETER ESTIMATION METHODS



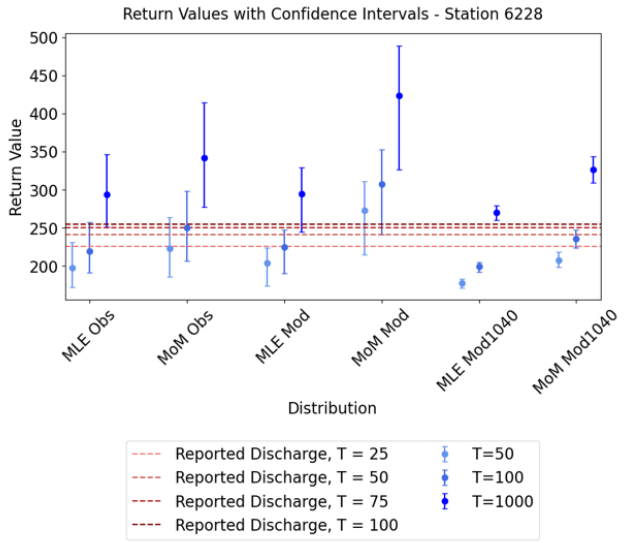
Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	54 years	Observed	613
Gumbel - MoM	54 years	Observed	614
Gumbel - MLE	54 years	Modelled	593
Gumbel - MoM	54 years	Modelled	594
Gumbel - MLE	1,040 years	Modelled	12071
Gumbel - MoM	1,040 years	Modelled	12090

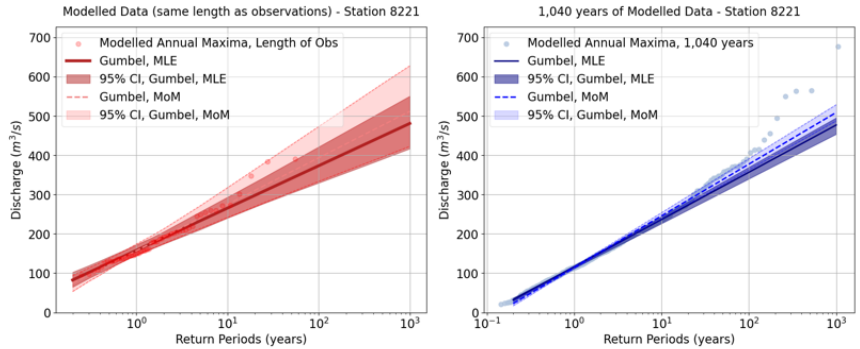


C.2 PARAMETER ESTIMATION METHODS

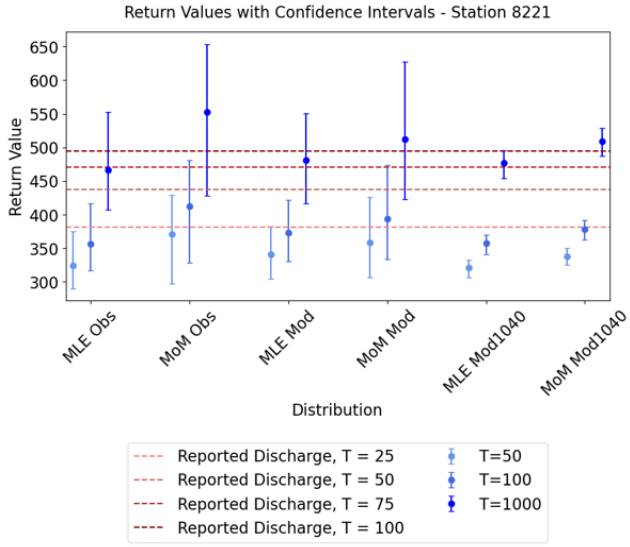


Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	56 years	Observed	573
Gumbel - MoM	56 years	Observed	577
Gumbel - MLE	56 years	Modelled	575
Gumbel - MoM	56 years	Modelled	594
Gumbel - MLE	1,040 years	Modelled	10554
Gumbel - MoM	1,040 years	Modelled	10659

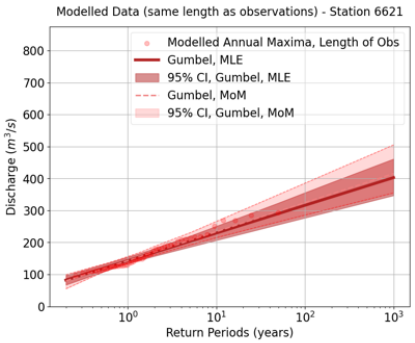


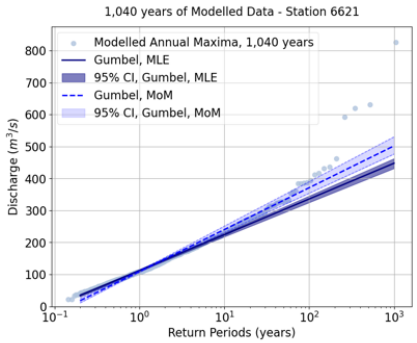


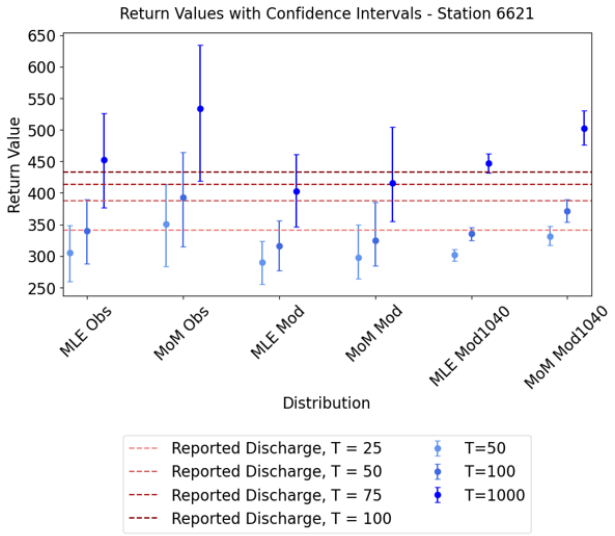
Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	55 years	Observed	611
Gumbel - MoM	55 years	Observed	615
Gumbel - MLE	55 years	Modelled	605
Gumbel - MoM	55 years	Modelled	606
Gumbel - MLE	1,040 years	Modelled	11562
Gumbel - MoM	1,040 years	Modelled	11579



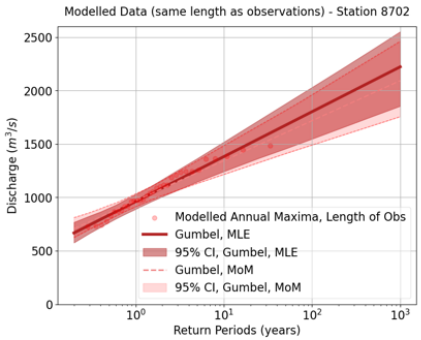
APPENDIX C: EXTREME VALUE MODELS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

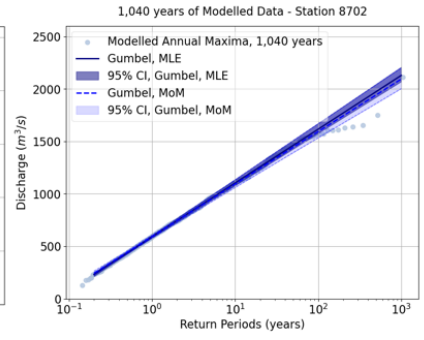


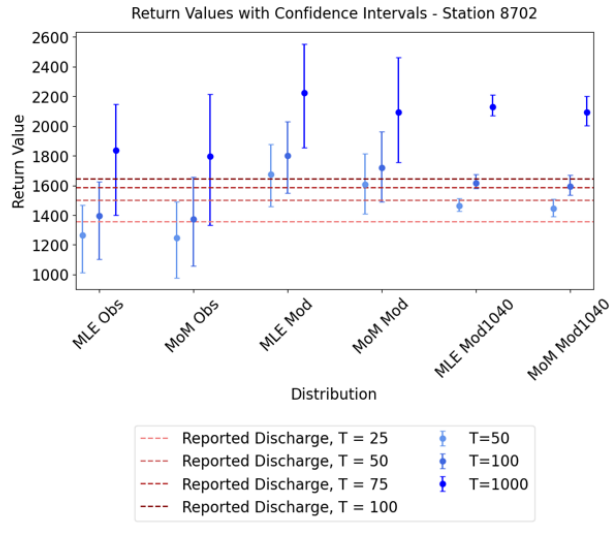




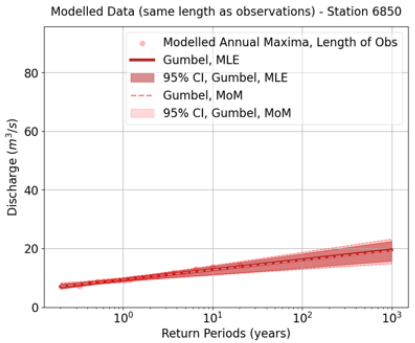
Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	49 years	Observed	548
Gumbel - MoM	49 years	Observed	552
Gumbel - MLE	49 years	Modelled	517.7
Gumbel - MoM	49 years	Modelled	517.9
Gumbel - MLE	1,040 years	Modelled	11444
Gumbel - MoM	1,040 years	Modelled	11492

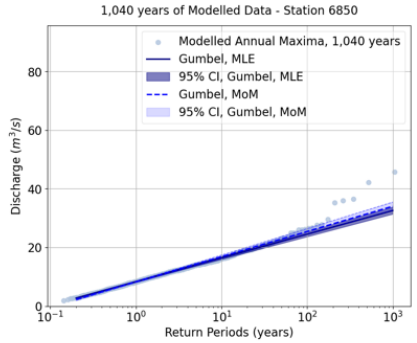


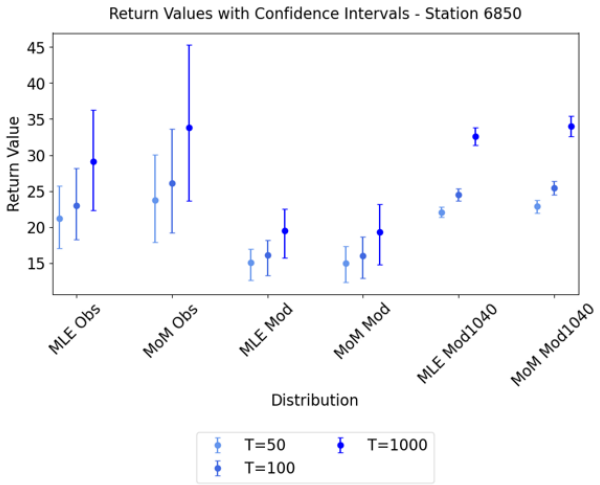




Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	33 years	Observed	454.7
Gumbel - MoM	33 years	Observed	454.8
Gumbel - MLE	33 years	Modelled	450
Gumbel - MoM	33 years	Modelled	451
Gumbel - MLE	1,040 years	Modelled	14534
Gumbel - MoM	1,040 years	Modelled	14536







Distribution - Fit Method	Length of Time Series	Observed or Modelled	AIC
Gumbel - MLE	20 years	Observed	110
Gumbel - MoM	20 years	Observed	111
Gumbel - MLE	20 years	Modelled	83
Gumbel - MoM	20 years	Modelled	83
Gumbel - MLE	1,040 years	Modelled	5921
Gumbel - MoM	1,040 years	Modelled	5930

Table 21: Percent Difference in Estimates of the 100-year Discharge using MLE and MoM

Station	Percent Difference between Estimates Obtained from MLE and MoM		
	Observations	Modelled Data, Length of Obs.	1,040-years of Modelled Data
8221	14%	5%	6%
5921	8%	3%	6%
6228 [1]	12%	27%	49%
6621	13%	2%	9%
8702	2%	5%	1%
6850	12%	0%	4%

Note:

1. Gumbel was not a good fit for the 1,040 years of modelled data at Station 6228 when using MLE or MoM.

APPENDIX D: EVENT SETS INFLUENCE ON DISCHARGE ESTIMATES AND THEIR UNCERTAINTY

D.1. RECORD LENGTH

Table 22: Percent Decrease in Width of Confidence Interval from a Record Length of 65 years

Station	Return Level	Percent Decrease in Width of Confidence Interval from a Record Length of 65 years							
		Gumbel				GEV			
		130 years	195 years	260 years	1040 years	130 years	195 years	260 years	1040 years
8221	50-year	26%	53%	63%	78%	31%	46%	65%	83%
	100-year	27%	52%	64%	77%	32%	48%	66%	84%
	1000-year	29%	53%	63%	78%	37%	53%	71%	87%
5921	50-year	46%	64%	71%	83%	58%	69%	69%	85%
	100-year	45%	65%	72%	84%	61%	72%	72%	87%
	1000-year	47%	66%	72%	84%	71%	80%	81%	90%
6228	50-year	36%	57%	69%	83%	42%	71%	78%	88%
	100-year	36%	57%	70%	83%	46%	74%	81%	89%
	1000-year	37%	57%	69%	83%	58%	82%	88%	93%
6621	50-year	37%	51%	66%	82%	39%	49%	67%	83%
	100-year	38%	51%	66%	82%	42%	53%	70%	85%
	1000-year	40%	52%	67%	82%	52%	64%	78%	89%
8702	50-year	21%	41%	51%	75%	17%	26%	52%	71%
	100-year	20%	41%	51%	75%	15%	26%	52%	70%
	1000-year	19%	40%	51%	75%	5%	24%	53%	68%
6850	50-year	29%	43%	57%	71%	38%	50%	50%	88%
	100-year	25%	38%	50%	75%	39%	52%	52%	87%
	1000-year	27%	36%	55%	73%	38%	55%	55%	87%

D.1.1. GUMBEL DISTRIBUTION

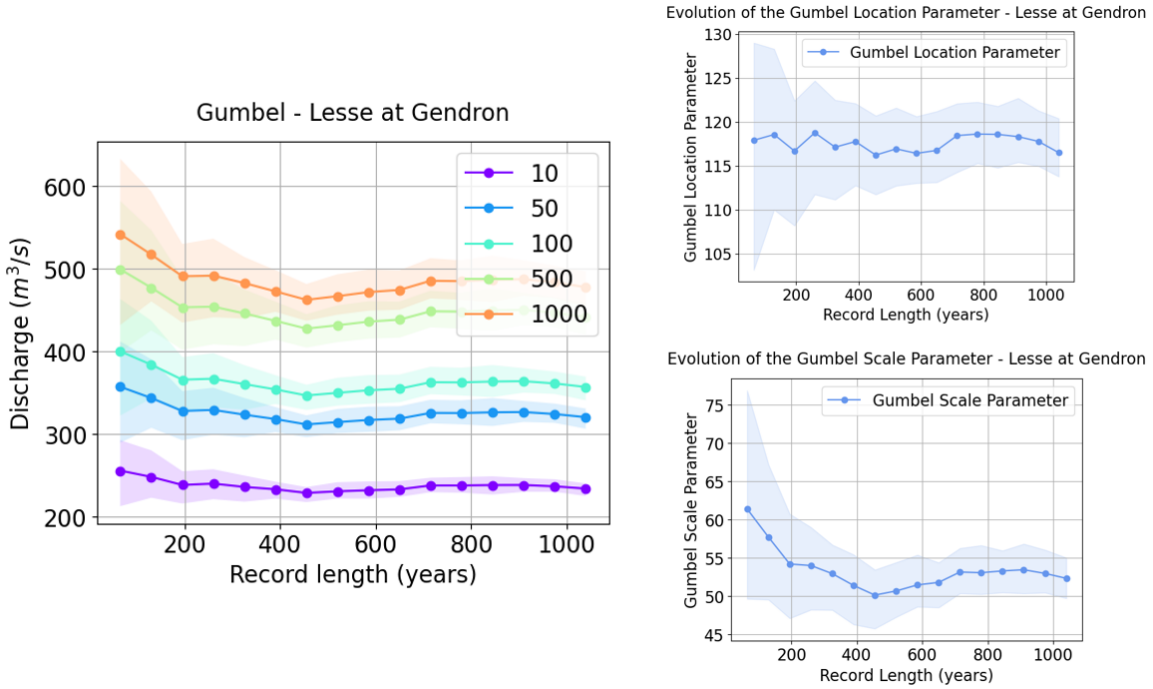


Figure 87: Estimated Discharges for varying Record Lengths - Station 8221

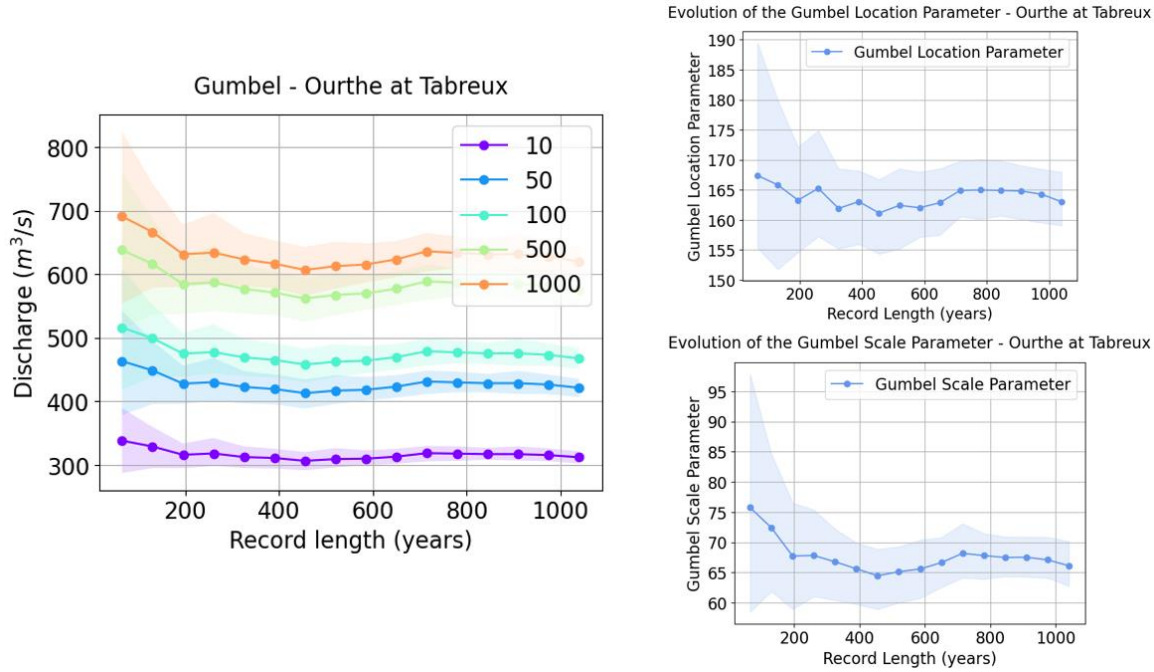


Figure 88: Estimated Discharges for varying Record Lengths - Station 5921

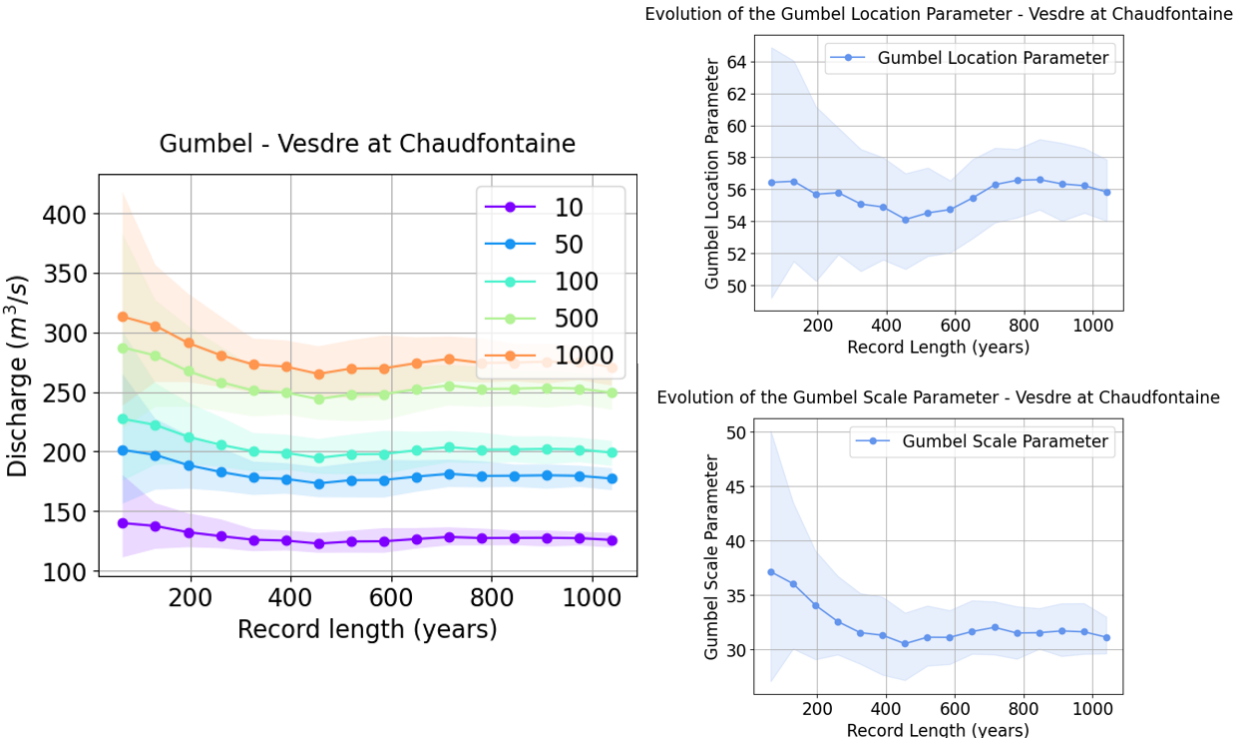


Figure 89: Estimated Discharges for varying Record Lengths - Station 6228

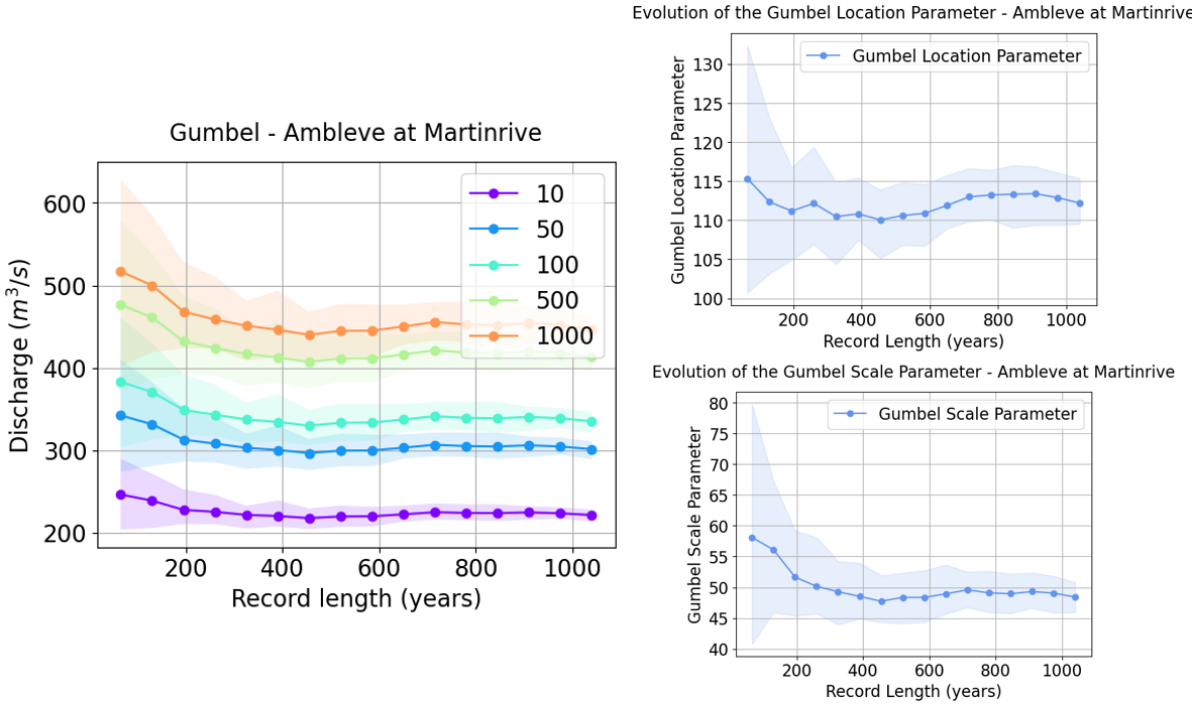


Figure 90: Estimated Discharges for varying Record Lengths - Station 6621

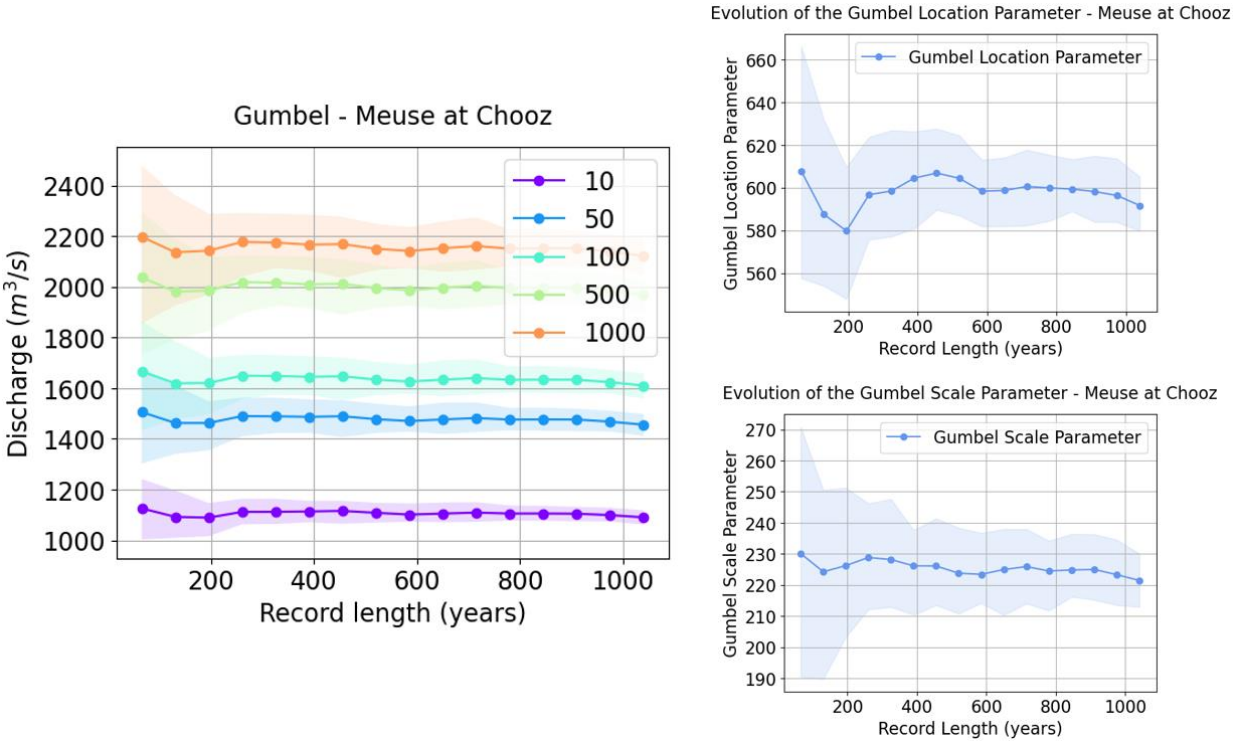


Figure 91: Estimated Discharges for varying Record Lengths - Station 8702

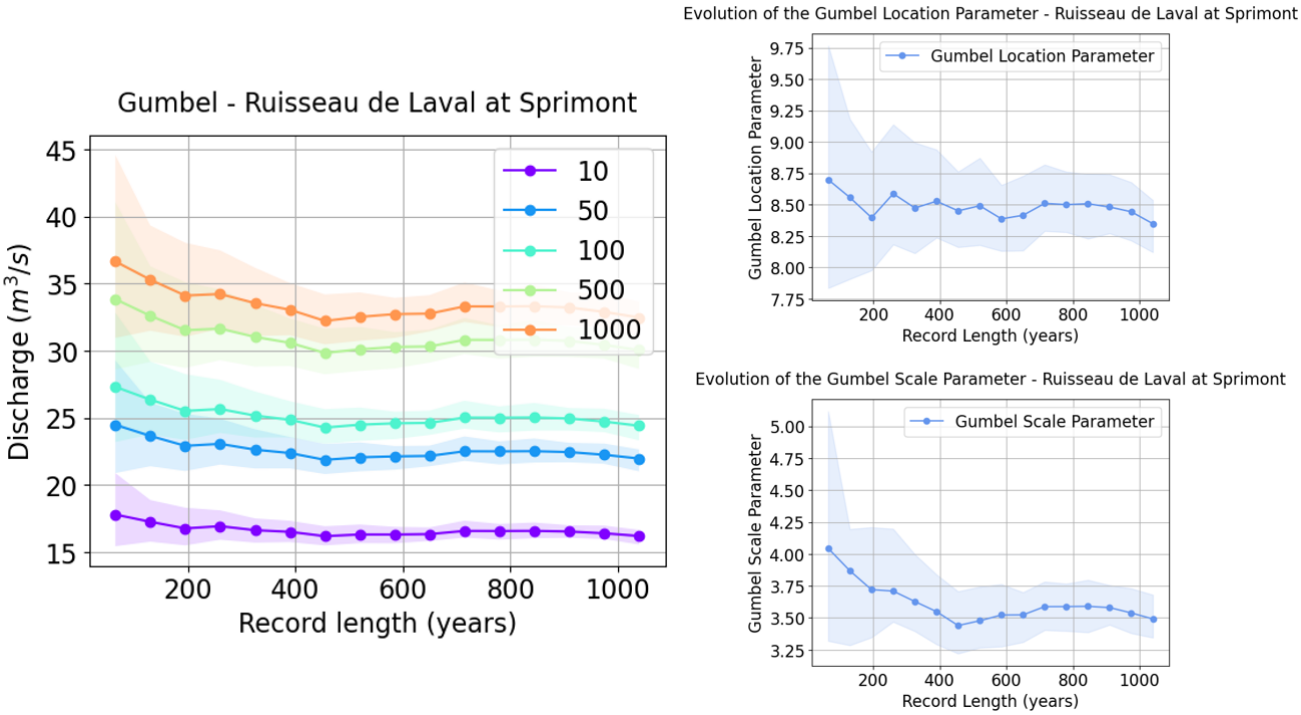


Figure 92: Estimated Discharges for varying Record Lengths - Station 6850

D.1.2. GEV DISTRIBUTION

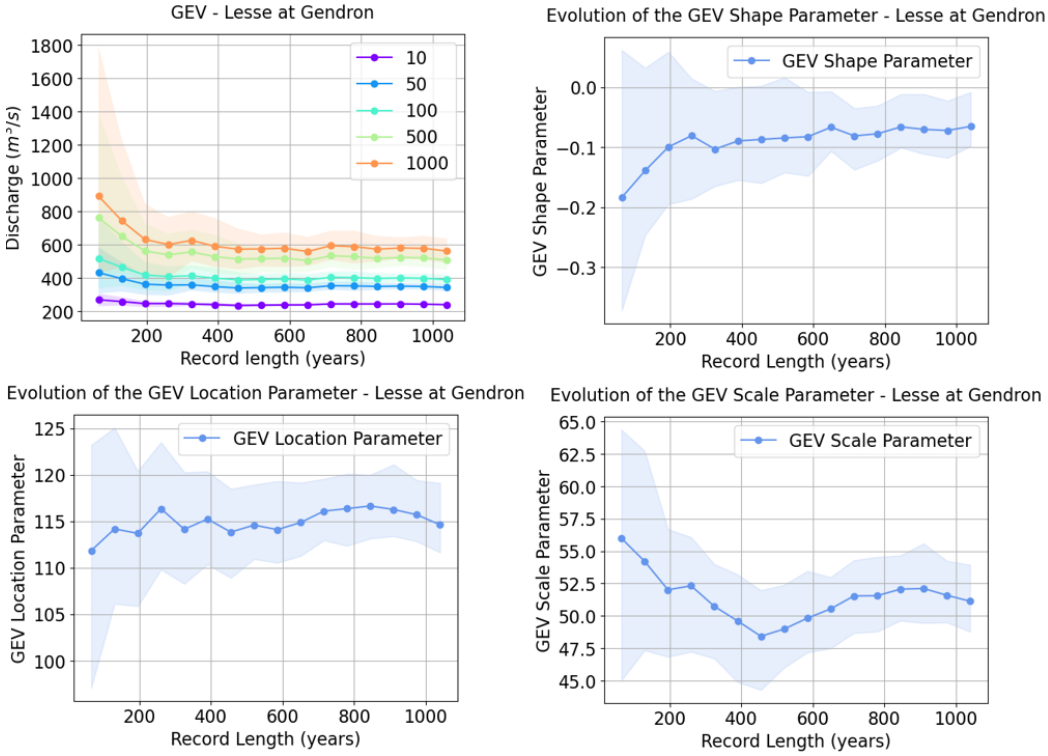


Figure 93: Estimated Discharges for varying Record Lengths - Station 8221

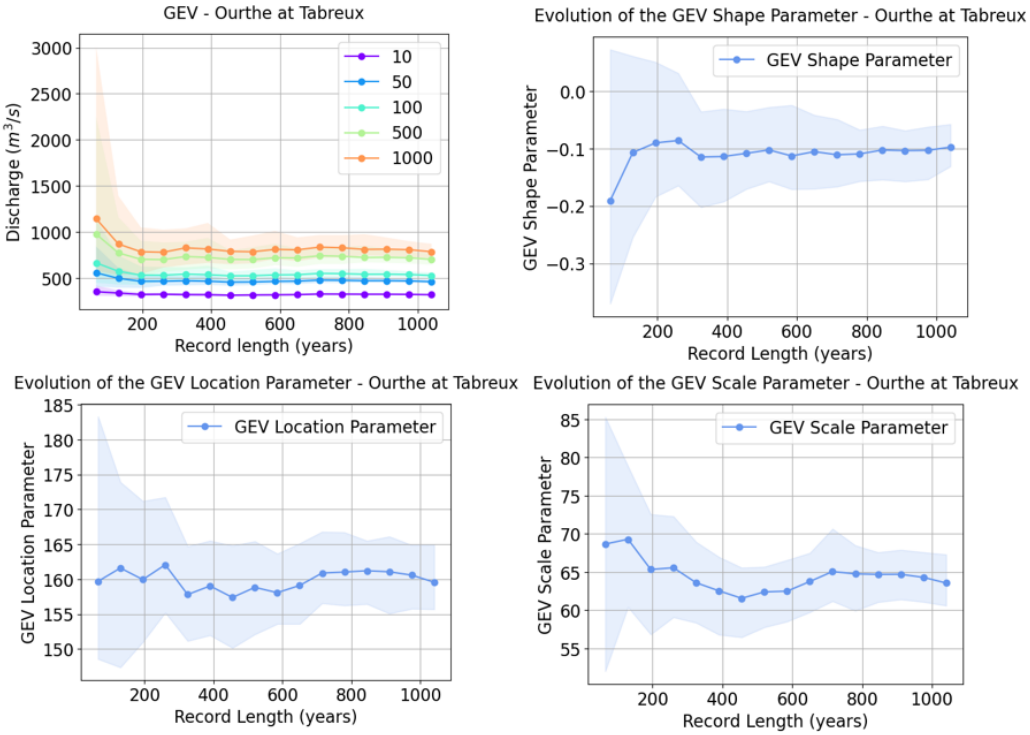


Figure 94: Estimated Discharges for varying Record Lengths - Station 5921

Figure 95: Estimated Discharges for varying Record Lengths - Station 6228

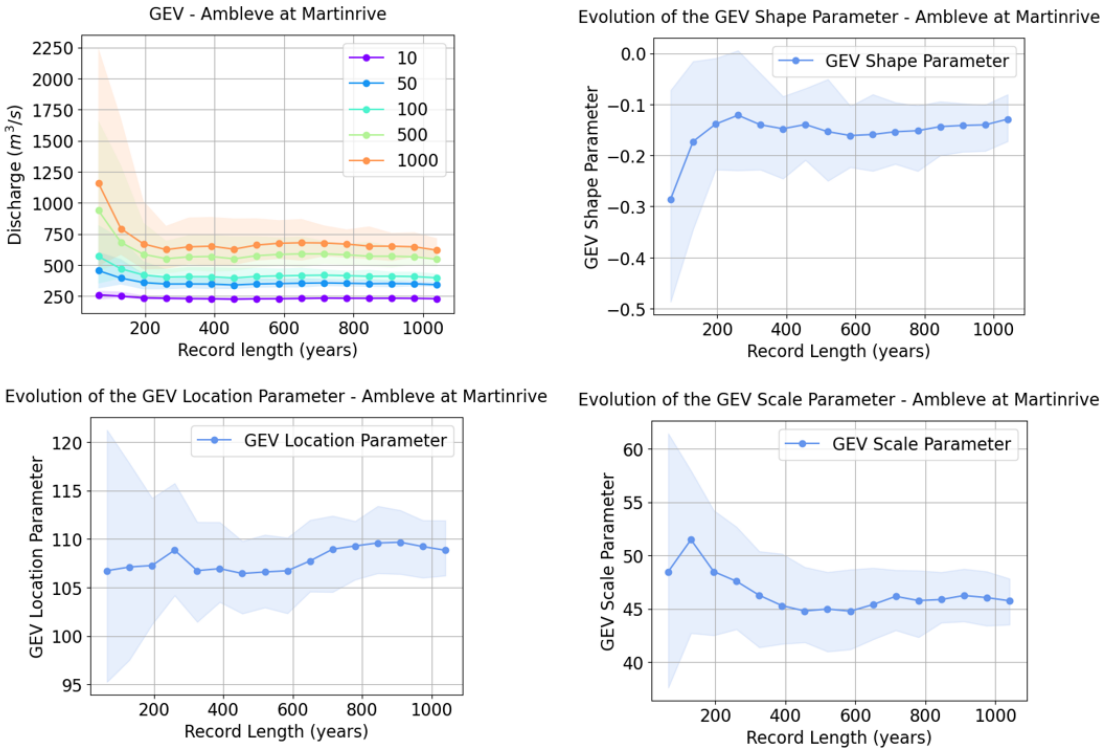


Figure 96: Estimated Discharges for varying Record Lengths - Station 6621

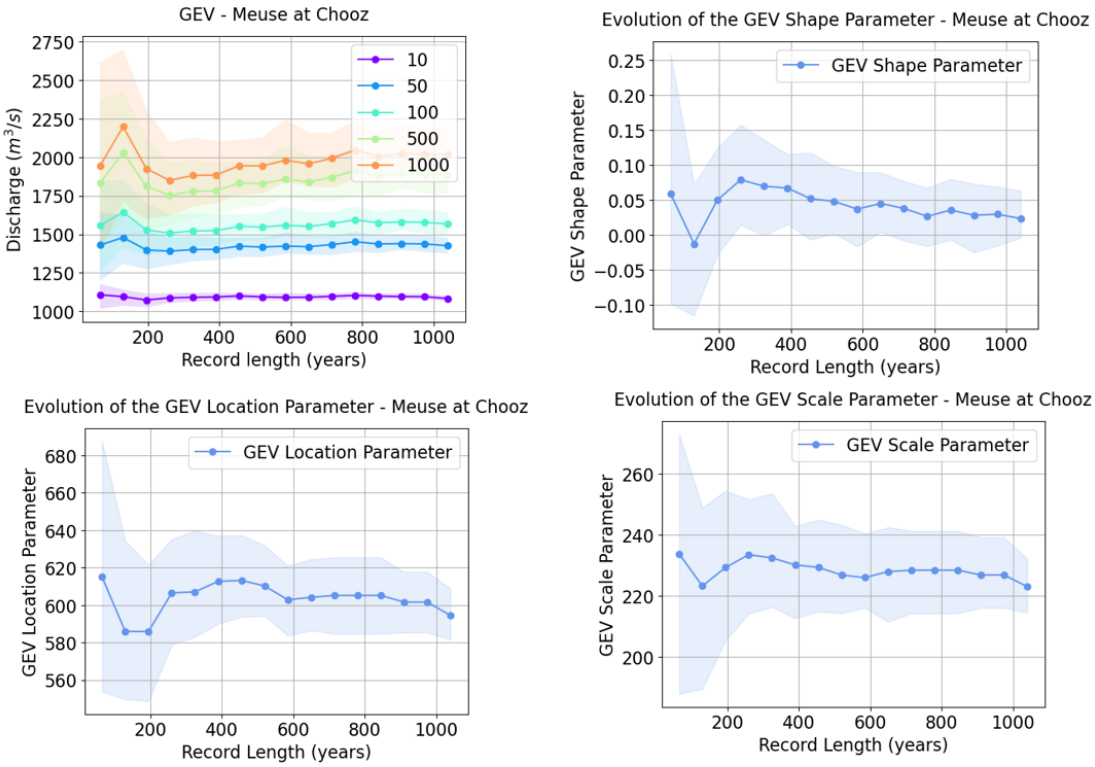


Figure 97: Estimated Discharges for varying Record Lengths - Station 8702

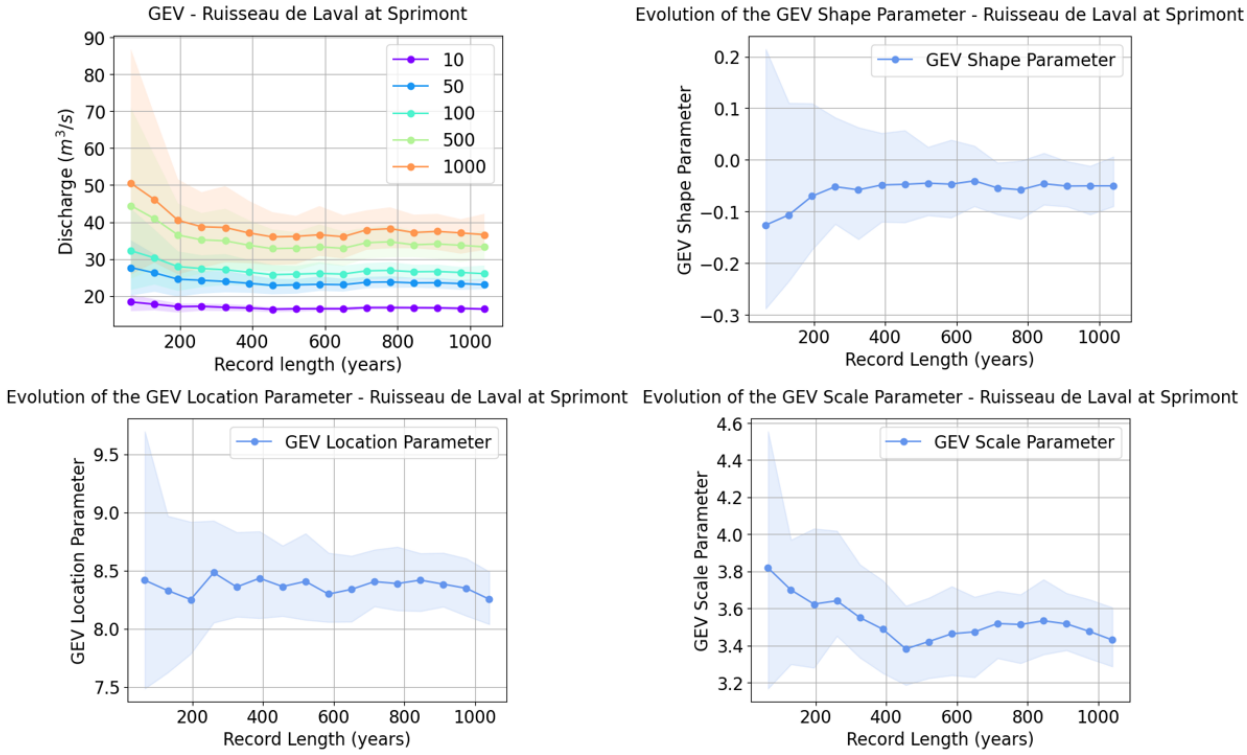


Figure 98: Estimated Discharges for varying Record Lengths - Station 6850

D.2. SEASONALITY

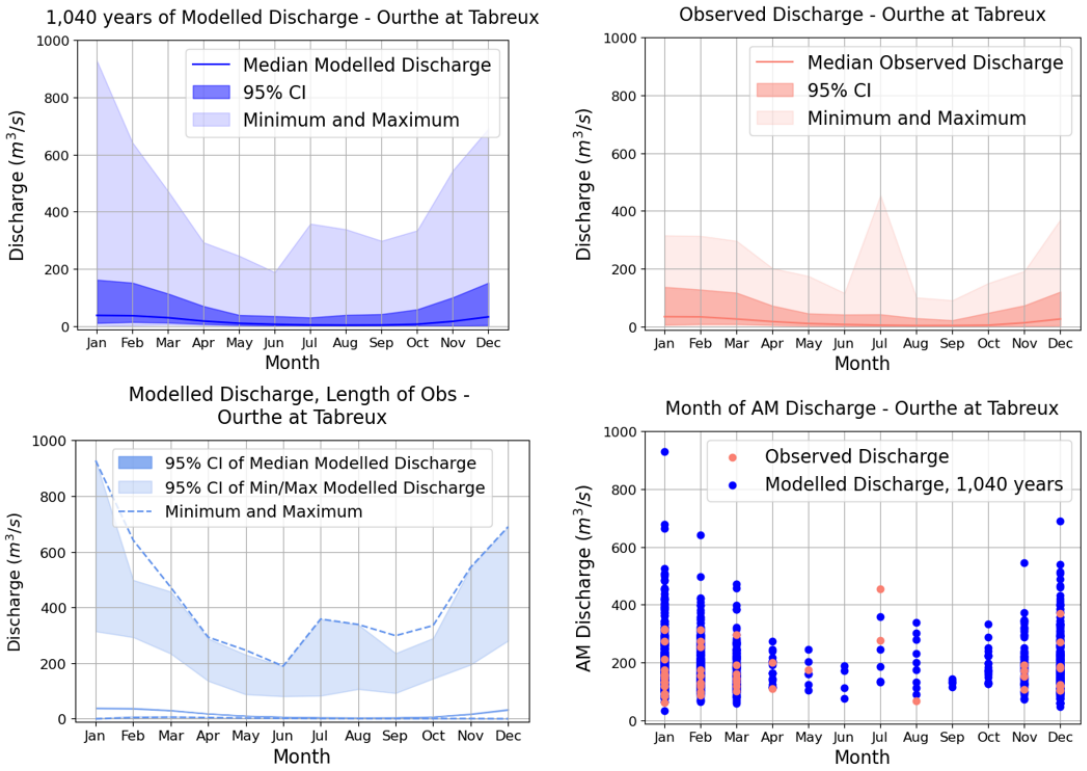


Figure 99: Seasonality of Annual Maximum Discharge - Station 5921

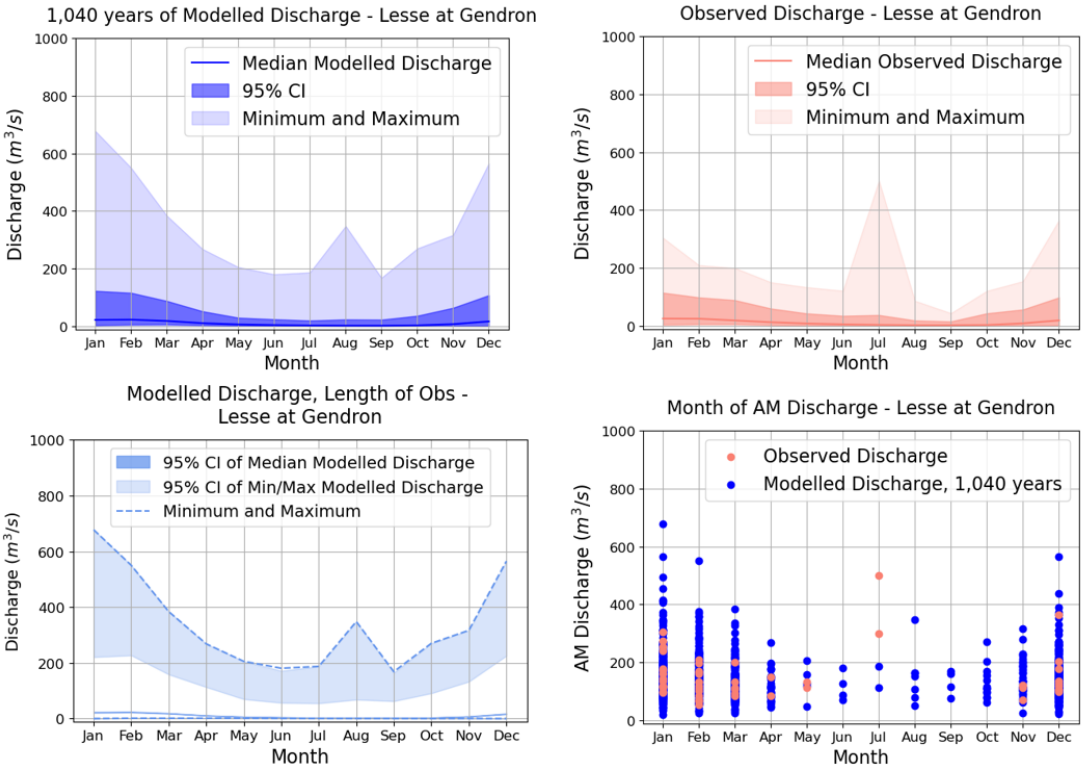


Figure 100: Seasonality of Annual Maximum Discharge - Station 8221

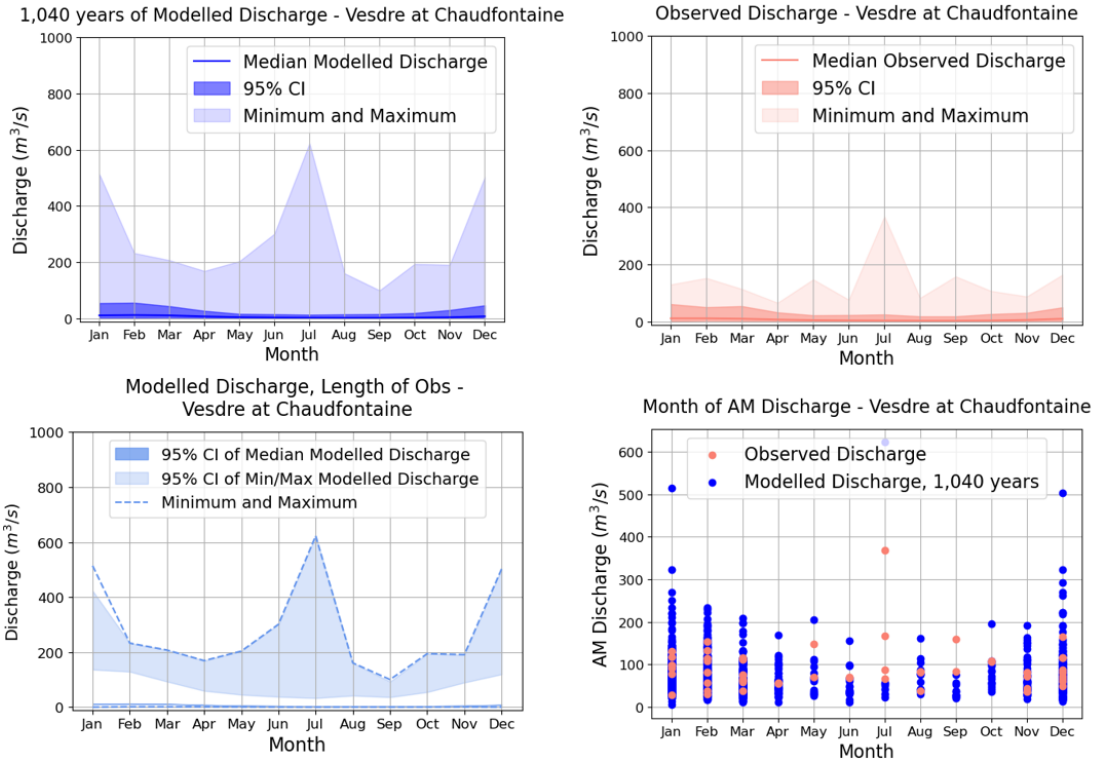


Figure 101: Seasonality of Annual Maximum Discharge - Station 6228

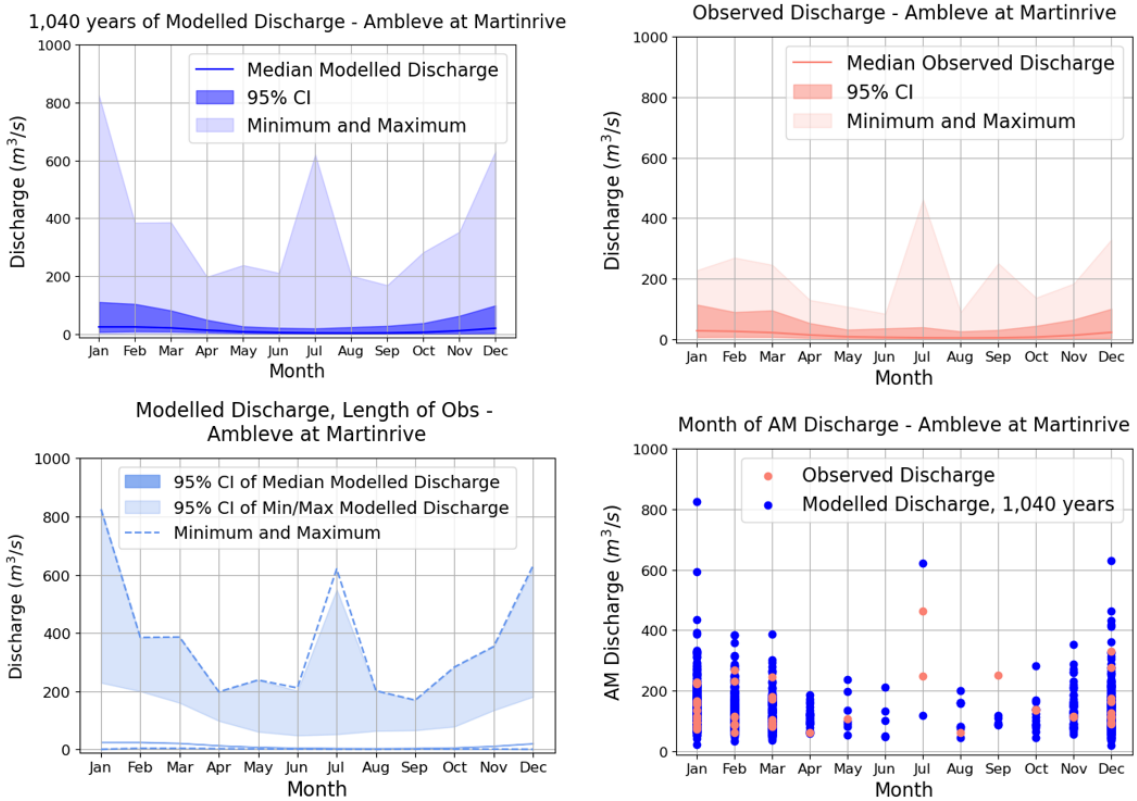


Figure 102: Seasonality of Annual Maximum Discharge - Station 6621

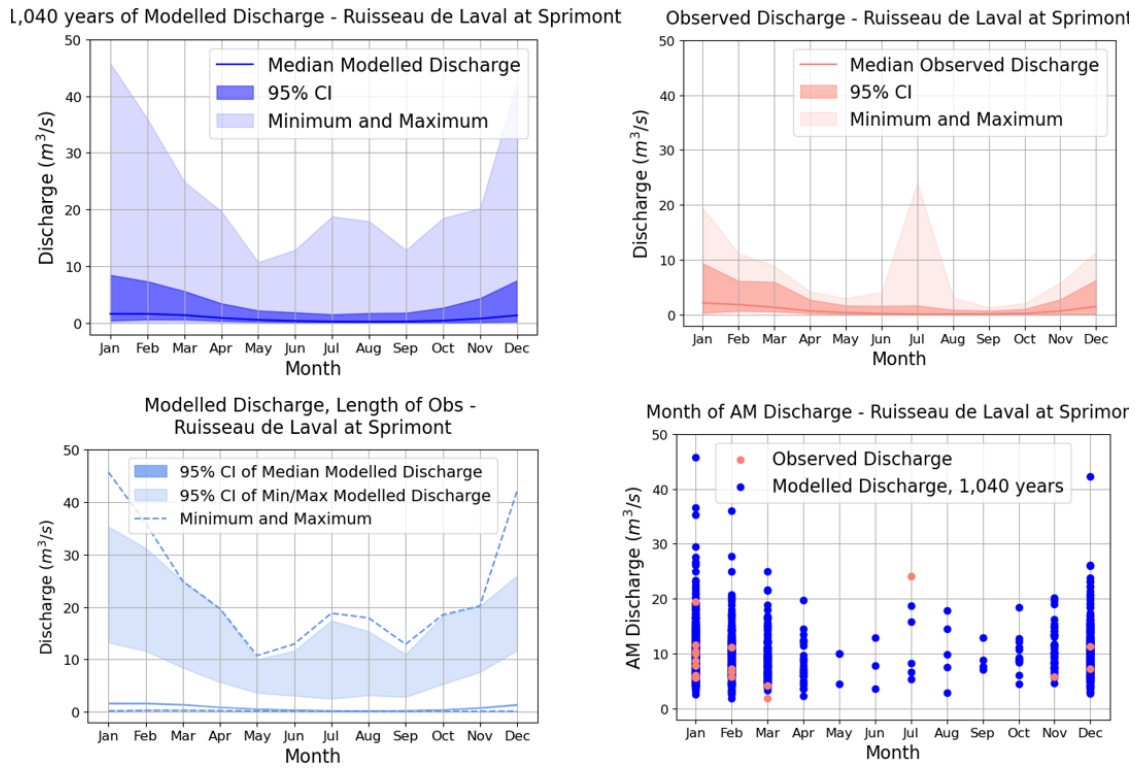


Figure 103: Seasonality of Annual Maximum Discharge - Station 6850

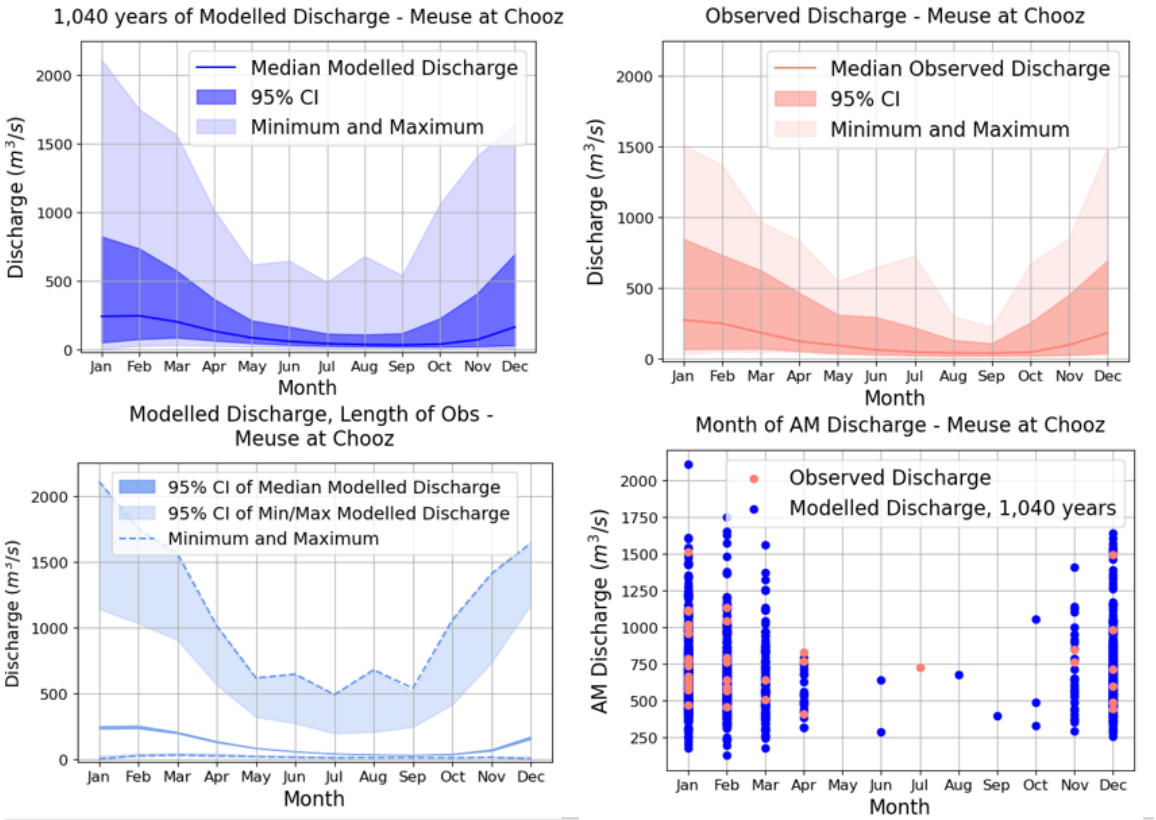


Figure 104: Seasonality of Annual Maximum Discharge - Station 8702

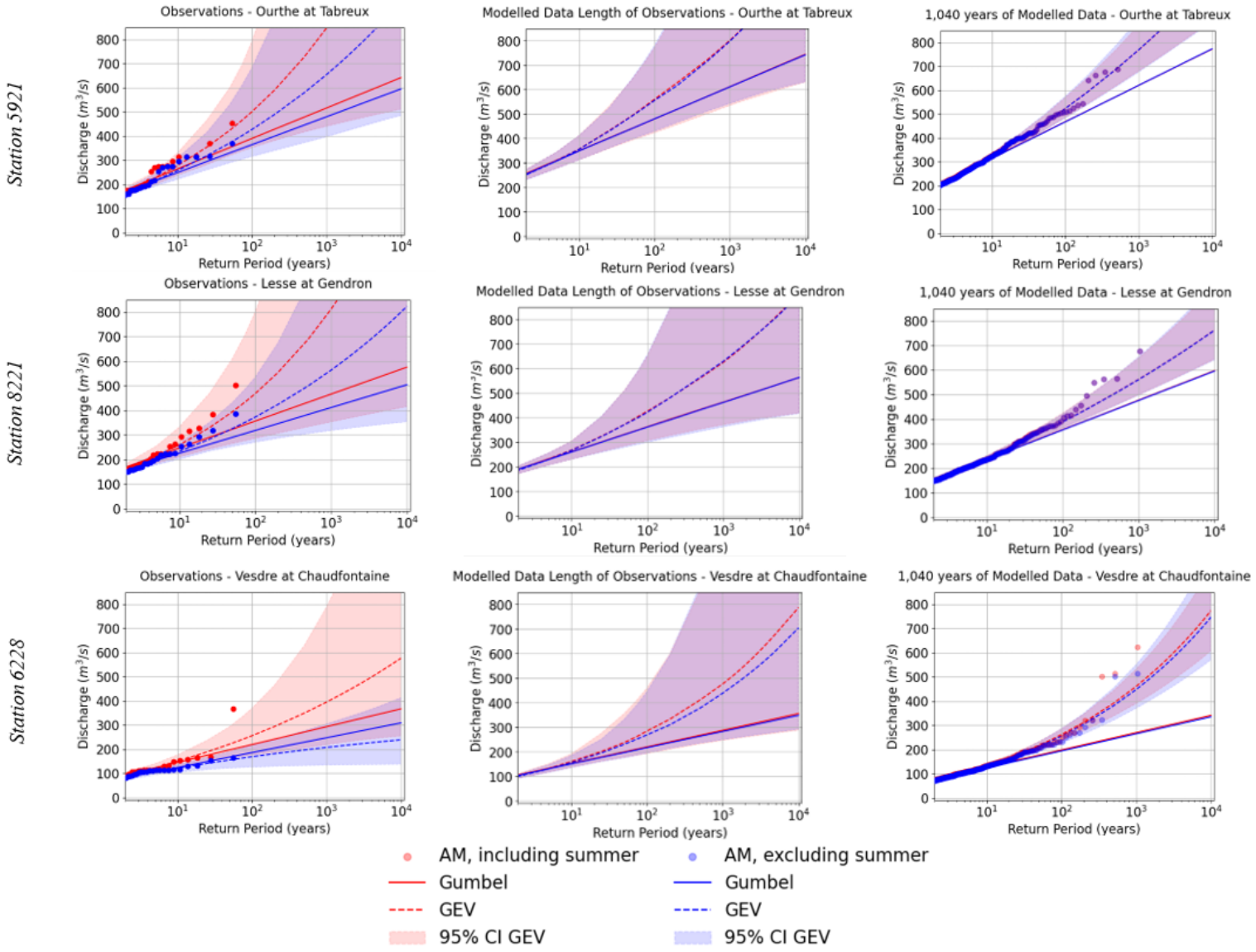


Figure 105: Results of Extreme Value Analysis Including and Excluding Summer Events (April - September) – Stations 5921, 8221, and 6228

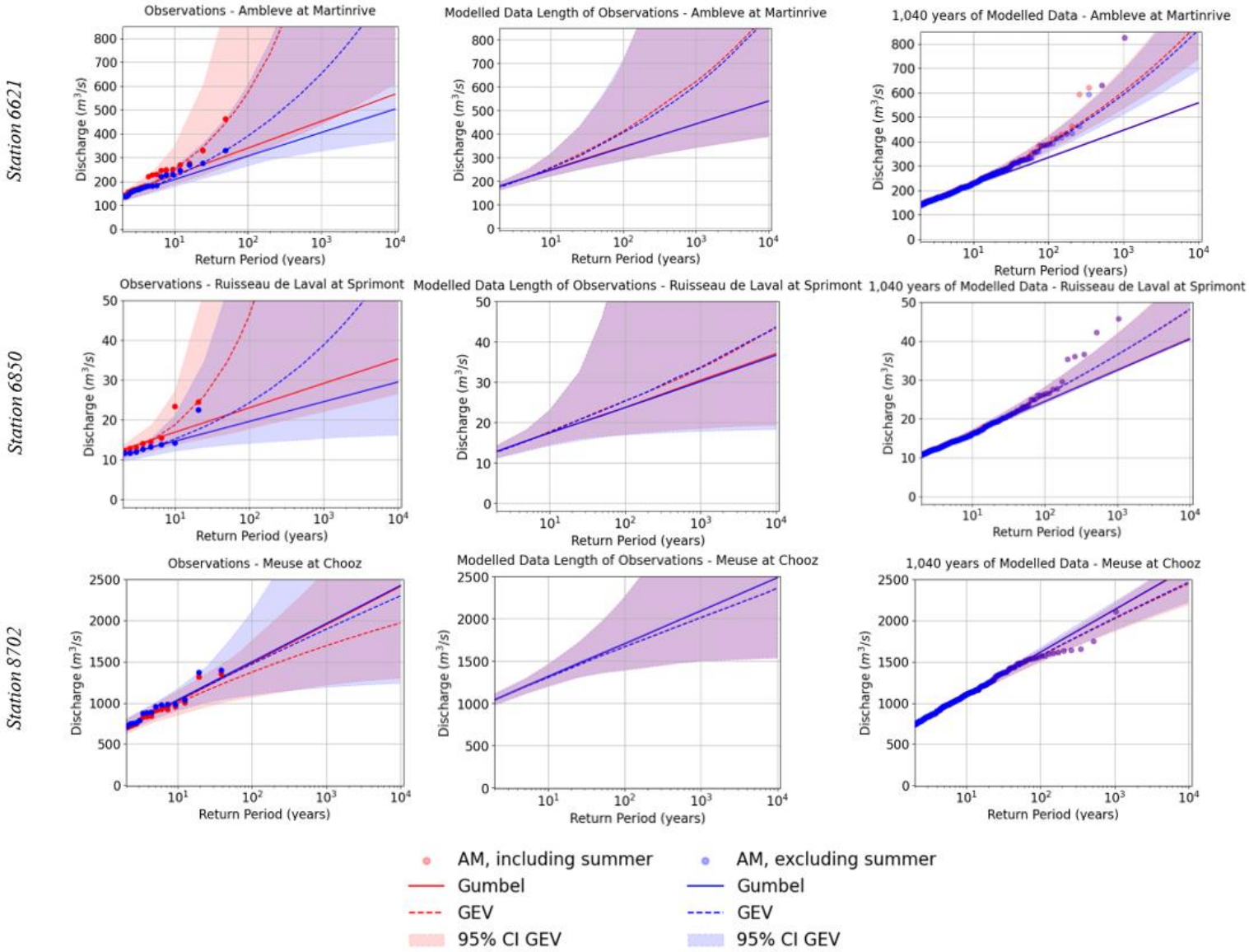


Figure 106: Results of Extreme Value Analysis Including and Excluding Summer Events (April - September) – Stations 6621, 6850, and 8702

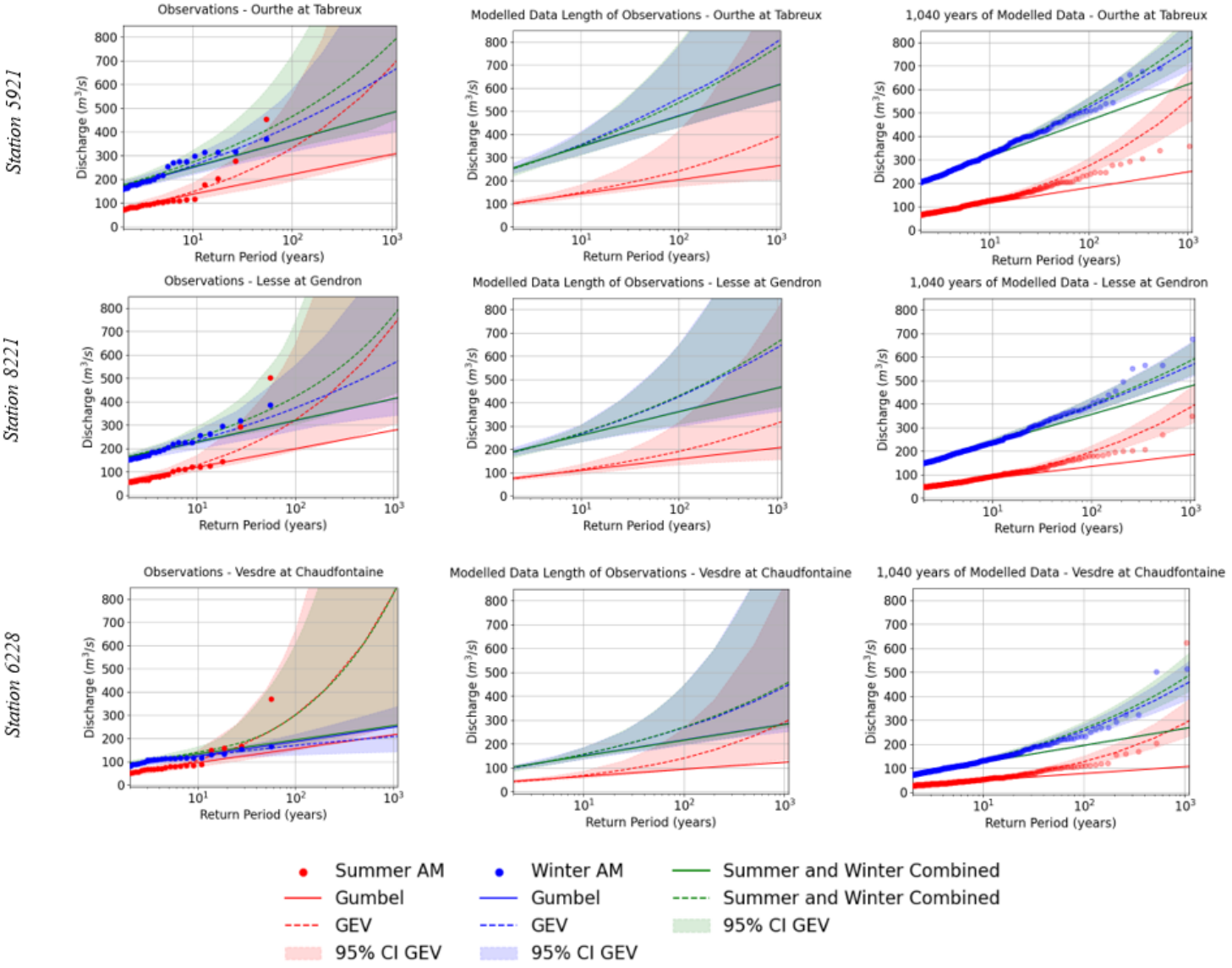


Figure 107: Results of EVA for Summer and Winter (Summer: April - September, Winter: October - March) – Stations 5921, 8221, and 6228

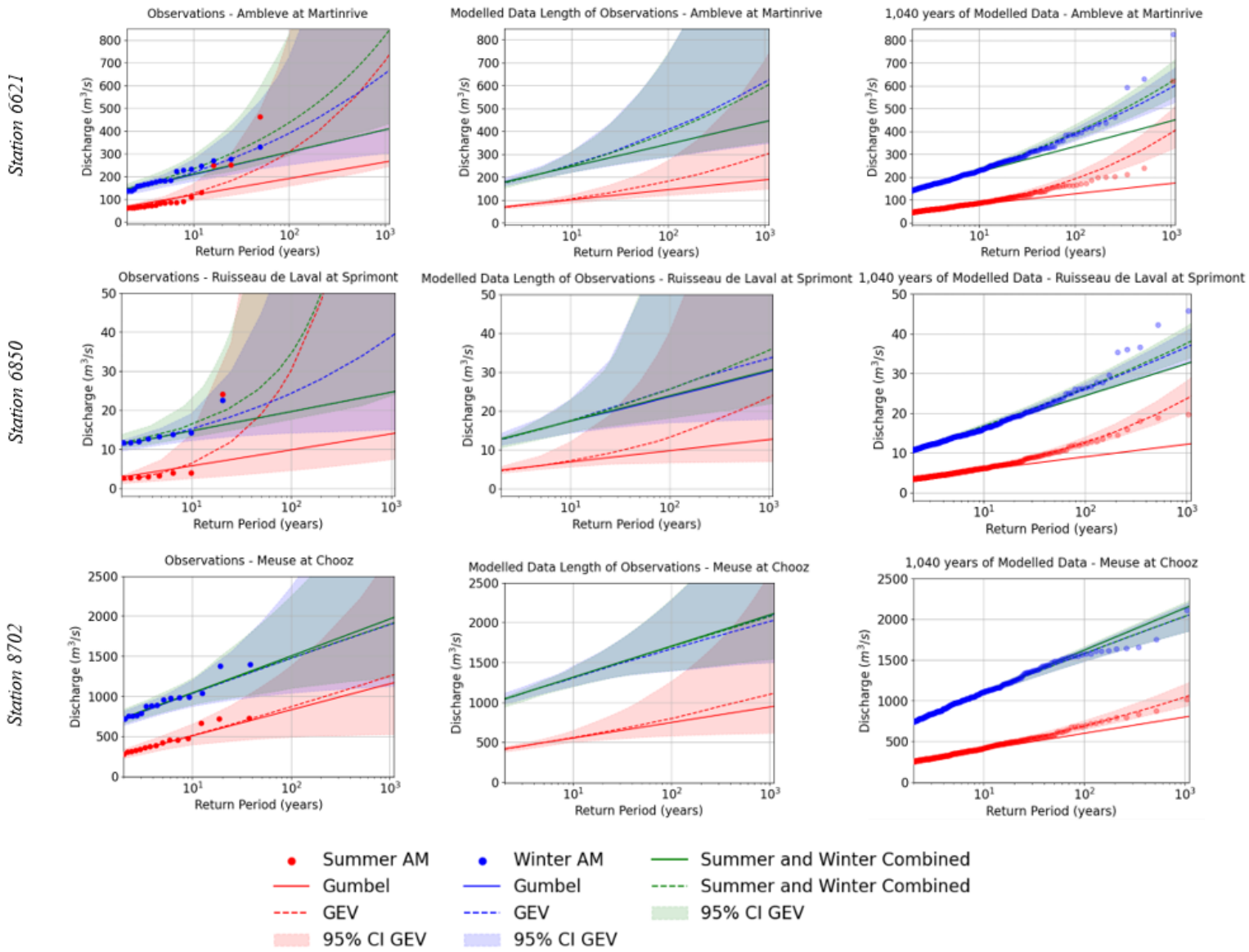


Figure 108: Results of EVA for Summer and Winter (Summer: April - September, Winter: October - March) – Stations 6221, 6850, and 8702

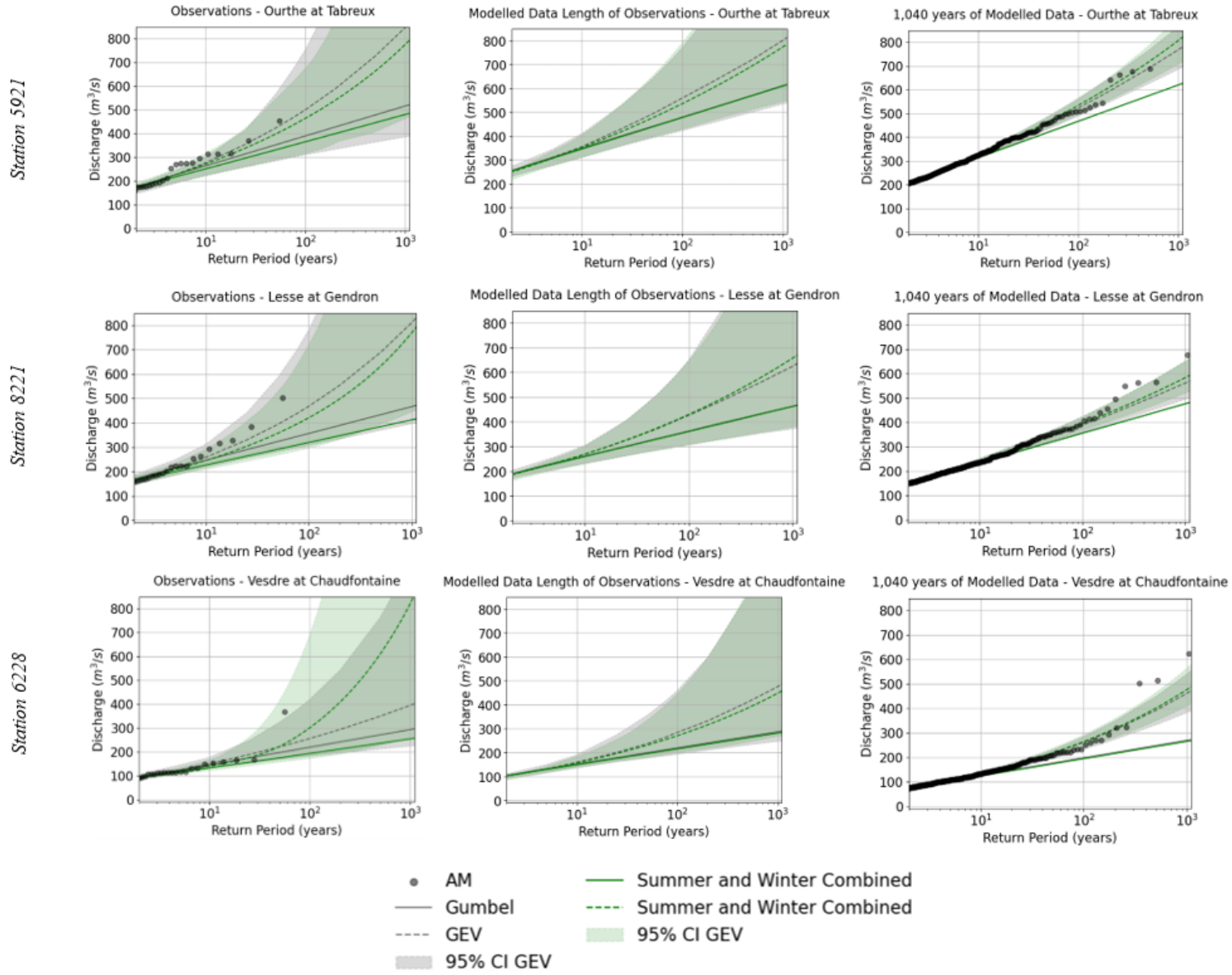


Figure 109: Comparison of Combined EVA (green) and Neglecting Seasonality (grey) (Summer: April - September, Winter: October - March) – Stations 5921, 8221, and 6228

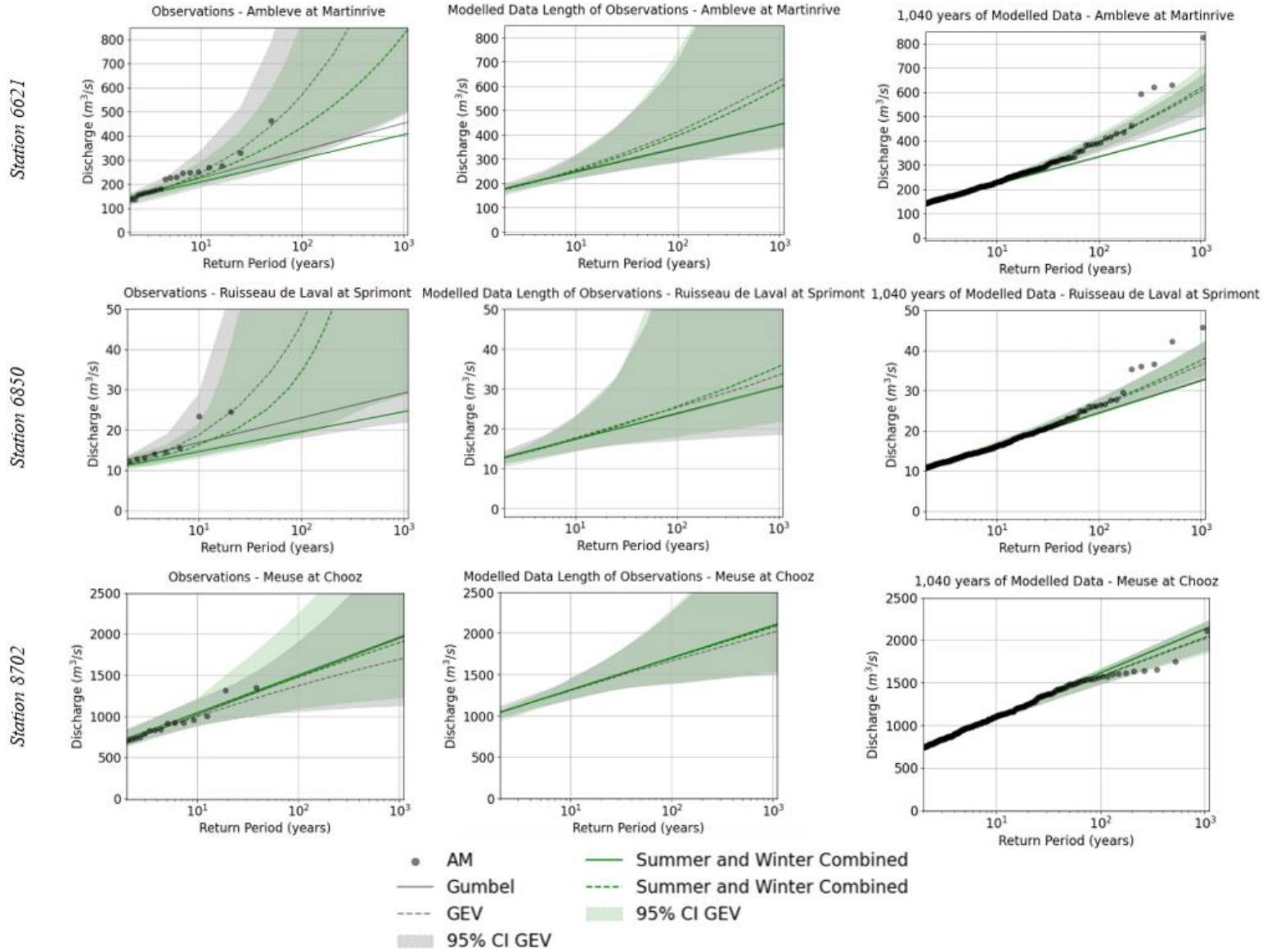


Figure 110: Comparison of Combined EVA (green) and Neglecting Seasonality (grey) (Summer: April - September, Winter: October - March) – Stations 6621, 6850, and 8702