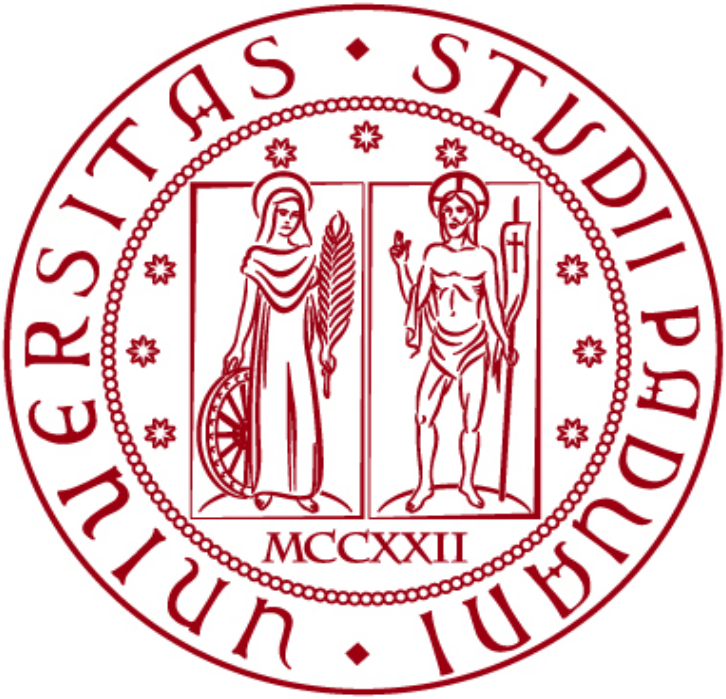
UNIVERSITY OF PADOVA
Civil and environmental department

Master's Degree in
Environmental Engineering



IMPACT OF INITIAL MOISTURE CONDITIONS ON THE GENERATION OF FLOOD EVENTS

Student: Davide Caimotto

Supervisor: Chiar.mo Prof. Gianluca Botter
Co-supervisor: Chiar.mo Prof. Marco Marani

Academic year: 2023-2024

ABSTRACT

Extreme events, such as floods, are increasing in frequency and intensity, making it essential to understand the factors that influence them. The objective of this thesis is to analyze the impact of initial soil moisture conditions on the generation of flood events and develop a statistical tool to study the impact of climate change on the magnitude of floods. Specifically, the study is conducted by exploiting rainfall and streamflow records observed in five study sites located in the Veneto region. The analysis starts by defining a bivariate joint distribution of peak flow rate and antecedent cumulative soil moisture. It was possible to determine that the antecedent moisture calculated over the past 5 days is more correlated with the discharge as compared to the antecedent moisture calculated over longer time-windows. Three models were compared to study extreme events: the Generalized extreme value theory (GEV), the Simplified Metastatistical Extreme Value Distribution (SMEVD) based on a single class of data and the moisture accounting SMEVD. In particular, the SMEV distribution is derived from the more generalized Metastistical Extreme Value (MEV) approach, which allows the use of a much larger sample of data than GEV, enabling a better description of extreme events, even in the context of climate change. The moisture accounting SMEVD is a model that allows the fit of flow events by dividing them into classes. In this case, three classes were used, created according to moisture level: dry soil, average wet soil and wet soil. Different methodologies such as Goodness of fit and Cross Validation were used to measure the performance of the models. Goodness of fit showed better performance using the GEV, the moisture accounting SMEVD has lower performance only for the extreme event related to storm Vaia. Cross-Validation, on the other hand, showed better performance using the moisture accounting SMEVD model, highlighting the advantages of using this model over GEV. The moisture accounting SMEV model, can also be used to make assumptions about climate change, in fact in the last phase of the research it was used to hypothesize three climate change scenarios. The first scenario was developed, assuming a 50% reduction in the frequency of events in the soil class with the highest moisture content. The second scenario instead assumes a 15% increase in the scale parameter of the Gamma distribution used to fit the events to the model. The third scenario (the most pessimistic) combines the changes assumed in the first two. The first and last climate change scenarios identify a decrease in magnitude for events associated with low return times, while an increase in magnitude is observed for events with high return times. In contrast, the second scenario shows an increase in magnitude for events associated with all return times.

INDEX

IMAGES

TABLES

EQUATIONS

1. INTRODUCTION

Extreme events are becoming a growing problem in the world. An extreme event is defined as a rare weather event (unusually high intensity) for a given location and/or time of the year. This definition was given by the IPCC, Intergovernmental Panel on Climate Change, the most important international organization for assessing climate change. The IPCC was founded in 1988 by the World Meteorological Organization (WMO) and the United Nations Environment Program (UNEP) to provide the world's population a clear view of the current and future climate conditions. The IPCC also prepares various types of reports, one of which is for politicians or policy makers to implement laws to try to limit extreme events. There are many types of extreme events, such as storms, heavy rain, floods, droughts, heat waves and wildfires. River floods occur when the water level exceeds the level of the banks or causes them to collapse due to excessive load, spilling water into the surrounding areas. This phenomenon is usually caused by intense weather events, such as intense and prolonged rainfall or the rapid melting of snow due to high temperatures. Climate change acts by causing higher temperatures, which favors more intense rainfall and accelerated snowmelt in some regions, thus increasing the risk of flooding. The consequences of floods are significant and can cause damage to both infrastructure and the local ecosystem. Climate change is mainly caused by the increase in greenhouse gases due to the use of fossil fuels. This phenomenon can be contained in built-up areas with the construction of reservoirs, but this is not always possible for economic or space reasons. The aim of this thesis is to analyze the impact of initial moisture conditions on the generation of flood events. To do this, three different models will be used to study extreme flood events. The three models are based on the use of two different distributions. The first approach uses the GEV (Generalized Extreme Value) distribution, a statistical distribution commonly used in the analysis of extreme events. The GEV is based on extreme value theory and is typically applied to the annual maxima of the flow time series. This distribution allows the probability of extreme flow events (such as floods) to be modelled based on the observed maxima in each year. The GEV distribution can easily fit the data, as it exploits three types of distributions Gumbel, Fréchet and Weibull. Nevertheless, the GEV distribution has some limitations. One of its main limitations is that it only considers annual maxima for the analysis of extreme

events, neglecting potential useful information that could be found in the other peaks. To overcome these limitations, an alternative approach was considered, the MEV (Metastatistical Extreme Value). This method allows the entire distribution of extreme data to be exploited and is not limited to considering a single value for each year. Through this assumption, it is possible to achieve greater precision in the description of the tail of the distribution, in other words a better fit between the distribution and the extreme events. The use of MEV also makes it possible to explore how the distribution of extremes may vary as a function of pre-existing climatic conditions. The other objective of this thesis is to develop three hypothetical climate change scenarios in order to observe how future climate conditions change. These future climate scenarios will be compared with the current climate condition and will make it possible to observe how the events associated with the same return time vary in intensity. By comparing these models and analyzing hypothesized scenarios, new knowledge will be acquired on the dynamics governing flood events and useful indications for flood risk management in the context of climate change. The expected results will be able to contribute to improving the ability to forecast extreme events and to refine adaptation strategies to new climatic conditions. In the following chapters, the methodologies used will be presented and explained in detail. In addition, the main analysis steps are explained.

2. METHODS

In the following chapter, the methodologies used to perform the analysis of extreme events are presented. In particular, the types of models used to describe the course of extreme events, the procedure for calculating the parameters of the distributions and the procedures used to process and analyze the results will be analyzed. Each methodology will be discussed in relation to the research objectives and starting hypotheses, highlighting the advantages and possible limitations of each approach. The reasons behind the choice of these methods and their suitability for the specific study context will also be explained.

2.1 Extreme Value Theory

Extreme Value Theory (EVT) was born out of the need to describe the probability of extreme events, that is, events characterized statistically by a low probability of occurrence. These events are often associated with particularly violent and intense weather phenomena, capable of causing serious damage both to the environment where they occur and to the population. Extreme Value Theory, therefore, can be used as a study tool for processes such as the hydrologic cycle at locate or global scales, floods, wind speeds, earthquake magnitudes, ecological processes, marine storm-surge levels, pollutant dispersion dynamics, and many other applications, including for financial and programming use.

The frequency of extreme events is associated with the so-called return period (Tr), which can be defined as the mean interarrival time between the occurrence of two subsequent events with intensity/magnitude equal or greater than an assigned intensity. Return period, however, presents some problems such as:

- availability of only certain percentiles;
- interpolation may induce errors;
- Non-robustness of nonparametric estimates;
- Underestimation of tails associated with theoretical pdfs;

Extreme value theory succeeds in making up for these shortcomings and providing more robust estimation of high return periods.

Early studies regarding these topics were carried out by: Maurice René Fréchet, in 1927, presenting the Fréchet distribution; in 1928, Ronald Fisher and Leonard Henry Caleb Tippett show the generalized distribution of extreme values; an important contribution was made by mathematician Emil Julius Gumbel, who defined the concept of distribution.

2.1.1 Traditional Extreme Value Theory (EVT)

There are two possible equivalent approaches, to study the probability distribution of extremes, associated with a random variable:

- Block Maxima Method (BM)
- Peak Over Threshold Method (POT)

Block Maxima theory (BM) allows a data set to be divided into non-overlapping blocks of equal length (usually one year). Within each block, the maximum observed value is considered. These maxima represent extreme events for each block. As in the case analyzed in this thesis, extreme flow events.

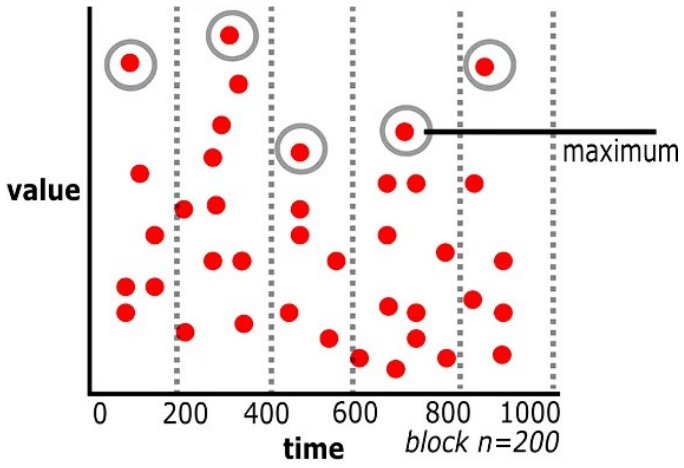


Figure 1 – Example Block Maxima Method

The maxima extracted from each block can be described using the GEV distribution also called the generalized extreme value distribution. The GEV can be defined as a family of distributions (Gumbel, Fréchet, Weibull) used to describe the extreme values extracted from a data sample. They can be extreme values understood as the maximum values or as the minimum values observed in a period. It is usually used in the fields of hydrology

or meteorology. Its equation can be given in two forms, through the cumulative distribution function (CDF) or through the density function (pdf).
The CDF of the GEV is given by:

$$G(x) = \begin{cases} \exp\left(-1 + \xi\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}, & if\ \xi \neq 0 \\ \exp\left(-exp\left(-\frac{(x-\mu)}{\sigma}\right)\right), & if\ \xi = 0 \end{cases} \quad \textit{(Eq. 1)}$$

Deriving the CDF yields the density function (pdf):

$$g(x) = \begin{cases} \frac{1}{\sigma}\left(1 + \xi\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}-1} \exp\left(-\left(1 + \xi\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{\xi}}\right), & if\ \xi \neq 0 \\ \frac{1}{\sigma}exp\left(-\frac{(x-\mu)}{\sigma}\right)\exp\left(-exp\left(-\frac{(x-\mu)}{\sigma}\right)\right), & if\ \xi = 0 \end{cases} \quad \textit{(Eq. 2)}$$

Where $\mu \epsilon$ R corresponds to the position parameter, $\sigma > 0$ corresponds to the scale parameter and $\xi \epsilon$ R corresponds to the shape parameter. Depending on the value assumed by the shape parameter, the GEV distribution uses a different distribution, controlling the tails accordingly:

I. $\xi = 0$ Gumbel distribution
II. $\xi > 0$ Fréchet distribution
III. $\xi < 0$ Weibull distribution

Tails can be of three types:

- Exponential, described by the Gumbel distribution
- Heavy, defined by the power law and described by the Fréchet distribution
- Finite, defined in a finite range and described by the Weibull distribution

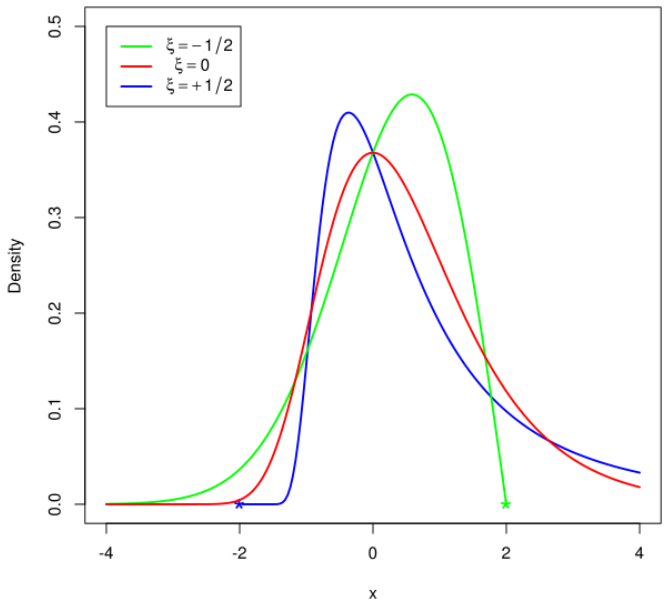


Figure 2 - Example of the different tails decay, depending on the shape parameter (ξ). In green the Weibull distribution, in red the Gumbel distribution and in blue the Fréchet distribution.

The GEV distribution is obtained as a result of the extreme value theorem. The theorem states that under certain assumptions the distribution of normalized maxima (or minima) referring to a sequence of independent and identically distributed variables converges using one of three distributions (Gumbel, Fréchet, Weibull). Normalized maxima are calculated by considering independent and identically distributed random variables, denoted as $X_1, X_2, \dots, X_n$. *Mn* is defined as the maximum of the random variables $M_n = \max(X_1, \dots, X_n)$. The maximum *Mn* is standardized using linear transformations:

$$Z_n = \frac{M_n - b_n}{a_n} \qquad \textit{(Eq. 3)}$$

Where the values a_n and b_n correspond to normalisation constants dependent on n.

The Gumbel distribution is a continuous probability distribution with two parameters α and u, which is used to describe the extreme values of a continuous stochastic series. The parameter α is used for scale changes, if α >1 there is a compression, otherwise if α <1 there is an extension.

The parameter u controls the translation of the pdf, rightward if u>0, leftward if u<0.

$$P(x) = Pr\{X \leq x\} = exp\{-exp[-\alpha(x - u)]\}$$
$$p(x) = \alpha\, exp\{-exp[-\alpha(x - u)] - \alpha(x - u)\} \qquad \textit{(Eq. 4)}$$

$P(x)$ represents the cumulative probability of not exceeding and $p(x)$ is the density of probability. The Gumbel distribution is characterized with a mean (μ), a variance (σ^2), and by a coefficient of skewness (γ).

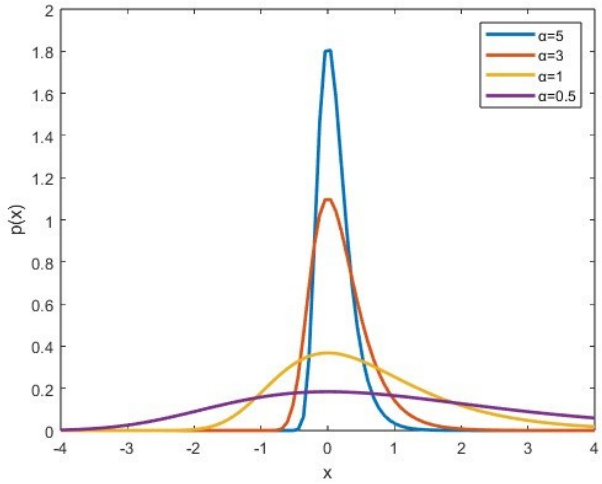


Figure 3 - Gumbel distribution with variation of scale parameter α..

Fréchet Distribution is a continuous probability distribution with three parameters α, u and θ, where the last one corresponds to the shape parameter. The tail decay follows the power law and it is particularly significative and long compared to the Gumbel tail.

$$P(x) = Pr\{X \leq x\} = exp\{-[\alpha(x - u)]^{-\theta}\}$$
$$p(x) = \alpha\theta\, [\alpha(x - u)]^{-\theta-1} - \; exp\{-[\alpha(x - u)]^{-\theta}\} \qquad \textit{(Eq. 5)}$$

Where x > u

Also, the Fréchet distribution is characterized with a mean (μ), a variance (σ^2), and by a coefficient of skewness (γ).

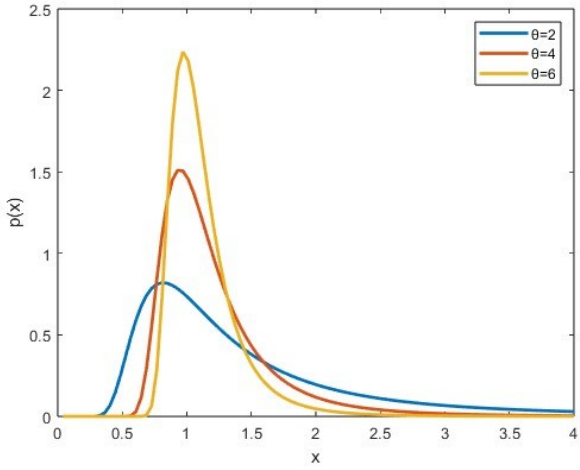


Figure 4 - Fréchet distribution with variation of shape parameter θ.

Weibull Distribution is a continuous probability distribution with three parameters α, u and θ. The distribution changes shape depending on the value of the shape parameter, going from an exponential to a bell-shaped distribution (see Figure 5).

$$P(x) = Pr\{X \leq x\} = 1 - exp\{-[\alpha(x - u)]^{\theta}\}$$

$$p(x) = \theta\alpha[\alpha(x - u)]^{\theta-1} exp\{-[\alpha(x - u)]^{\theta}\} \quad \textit{(Eq. 6)}$$

$$\text{Where } x \geq 0$$

Usually, this type of distribution is used to adjust the pattern of minimum.

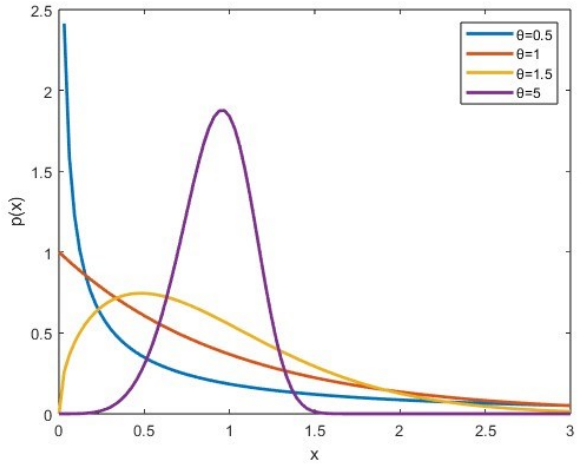


Figure 5 - Weibull distribution with variation of scale parameter θ.

The Peak Over Threshold Theory (POT) can be used as an alternative to Block Maxima and consists of considering a series of events, above a set threshold (u). As can be easily guessed, the lower the threshold, the greater will be the amount of data selected above it, on the contrary if the threshold is placed high.

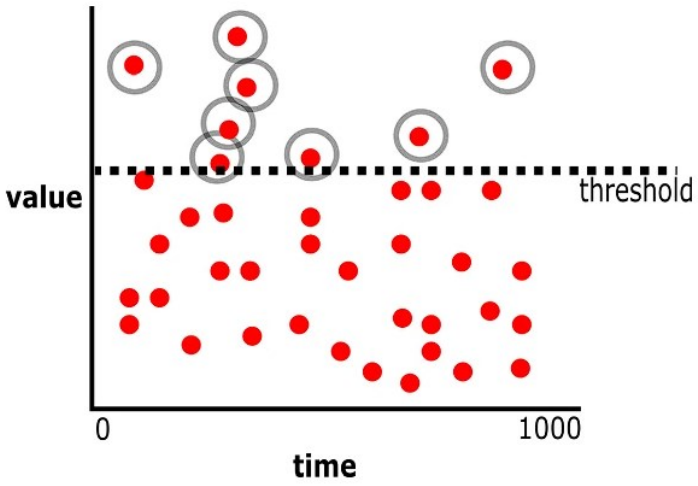


Figure 6 - Example Peak Over Threshold Method

One of the main problems with the POT method is the appropriate selection of the threshold. If the threshold is too high, one runs the risk of having too small a sample, limiting the robustness of the estimates. If, on the other hand, the threshold is too low, too much data is included, no longer focusing on true extreme events and violating the assumptions of the method. The choice of threshold could be problematic, if an increasing trend or a decreasing trend occurs in the sample under analysis. This is because the threshold is fixed and cannot be changed according to the trend in the data. To overcome this problem and others, a hybrid of the two methods (BM and POT) is used, which is presented and discussed in chapter 3.2.

2.1.2 Metastatistical Extreme Value Theory (MEV)

The extreme value method (EVT) analyzed in Chapter 2.1, needs a large amount of data to extrapolate extreme values, either through the block maxima (BM) method or the Peak Over Threshold (POT) method. Moreover, once extracted, there turns out to be little observational data to adequately describe the chosen distribution (Gumbel, Weibull, Frechet, etc). To overcome this problem, a variation of EVT, called Metastatistical Extreme Values (MEV), defined by Professor Marco Marani & Massimiliano Ignaccolo (2015), was introduced. This approach makes it possible to take advantage of all available data, thus offering a reduction in uncertainty in the estimation of high quantile extremes, 50 % greater than the traditional extreme value method (EVT).
The method can be summarized according to these steps:

- Consider the entire temporal series of data available;
- Select all peaks in the time series of data, trying not to consider noise, due to measurement errors or other issues;
- Divide the time period into annual blocks;
- Fit a distribution (Weibull, Gamma, Gumbel, etc.) to the different period in order to characterize the variability of the maxima within each period;
- Construct the MEV by combining all selected periods
- Estimate the probability of extreme events
- Validation and Analysis of Results Obtained

The method allows one to identify the number of events in each block, denoted by n, and the parameter values, represented by $\vec{\theta}$, of the parent distribution $F(x;\vec{\theta})$ as realizations of stochastic variables. The probability distribution of the maxima for each block can be defined, via the total probability theorem as:

$$G(x) = \sum_{n=1}^{\infty} \int_{\Omega_{\vec{\theta}}} F\left(x;\vec{\theta}\right)^n g\left(n;\vec{\theta}\right) d\vec{\theta} \qquad \textit{(Eq. 7)}$$

Where $F\left(x;\vec{\theta}\right)^n$ is the parent distribution for each annual block, $g\left(n;\vec{\theta}\right)$ is the joint probability distribution of the number of events in one year and $\Omega_{\vec{\theta}}$ signifies the entire population of the parameter's values. The MEV approach therefore uses all the observable

information available to characterize the probability distributions of common events in each data set, rather than simply considering only the extreme values of F(x).

The expression of the probability distribution of maxima $G(x)$ can be simplified by substituting the integral in the set $\Omega_{\vec{\theta}}$ in Eq. 7 with the sample average computed over all the blocks in the time series, obtaining:

$$G(x) = \frac{1}{M}\sum_{j=1}^{M} F\left(x;\vec{\theta}_j\right)^{n_j} \qquad \textit{(Eq. 8)}$$

where M are the annual data recording blocks and the distribution chosen is applied to each jth block ($j = 1, \dots, M$). Usually, the blocks considered are one year in size.

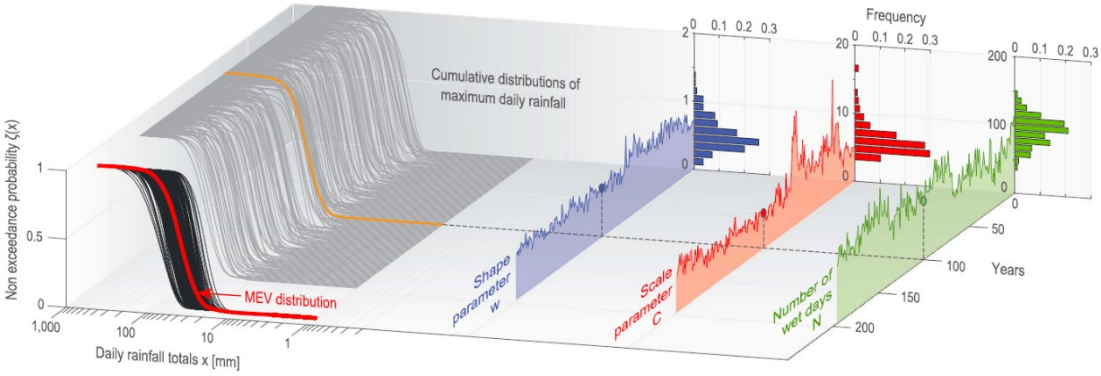


Figure 7 - Graphical representation of the MEV approach. Image taken from the scientific article "On the emergence of rainfall extremes fromordinary events," written and edited by E. Zorzetto, G. Botter, and M. Marani.

Figure 7 is an example of a graphical representation of the MEV approach, used for an analysis of rainfall extremes. As can be seen of the diagram on the left, the MEV distribution (represented by a curve in red), is obtained from the average of the cumulative distributions of annual maximum rainfall. The parameters of the Weibull distribution are estimated using the Probability Weighted Moments (PWMs) method, which succeeds in attributing greater importance to the tail's distributions and is more reliable for small sample size. Similarly, the L-moments method that was used in this thesis (see pages 16-17). On the right, the pdfs of the shape and scale parameters associated with the Weibull distribution $F(x; C, w)$ are highlighted. Finally, the pdf of the numbers of wet days, based on a number of years, is depicted.

In conclusion, the MEV method provides the following advantages:

- It eliminates the asymptotic assumption and it does not require a large number of observations for each year;
- Elimination of the assumption of poison distribution of the data;
- Allows the annual maxima to be extrapolated not only from the tail of F(x), but from the entire data set;
- In the case of the MEV-Weibull approach, allows the distribution of extreme events to be described by either a thin tail (associated with an exponential trend) or a fat tail (trend according to a power law. This can be deduced by varying the shape parameter.

2.1.3 Simplified Metastatistical Extreme Value Theory (SMEVD)

This simplified version of the MEVD, developed by Marra et al. (2019), allow one to neglect of interannual variability, captured by the standard MEV in equation 8 by j. SMEVD is usually adopted to improve parameter estimation in the presence of a small annual number of ordinary events. By doing so, the simplified equation can be defined:

$$G(x) \cong F\left(x;\vec{\theta}\right)^{n} \qquad \textit{(Eq. 9)}$$

$F\left(x;\vec{\theta}\right)^{n}$ is the cumulative distribution of ordinary events and *n* is the average number of events per year. This method is commonly used when one does not have enough ordinary values for each year to be able to apply classical MEVD. So, the main difference is that classical MEVD considers the cumulative distribution function $F\left(x;\vec{\theta}\right)^{n}$ for each year, while SMEVD considers the cumulative distribution function, applied over all ordinary events (even of different years).

2.1.4 Moisture-accounting SMEVD

This method is based on the theory of simplified metastatic extreme values (SMEVD) but has been modified to apply it on a sample of ordinary events, divided into wetness classes. In this case, the sample will be divided into three soil moisture classes (dry class 1, medium class 2, wet class 3). This assumption makes it possible to make assumptions about climate change (see chapter 4.6 for a detailed description). The resulting equation is very similar to the SMEVD equation, the only difference being that the producer symbol is added in front of the cumulative distributions. The resulting formula is as follows:

(Eq. 10)

$$G(q) = \prod_{1}^{k} F_k\left(q; S_{0_k}\right)^{n_k}$$

On the above equation, k corresponds to the number of wetness classes (in this case equal to 3), n_k is the mean number of events per year, q are the flow peaks, S_{0_k} is the cumulative soil moisture subdivided in classes and $F_k\left(q; S_{0_k}\right)$ is the cumulative distribution of ordinary values for each class.

The distributions that are commonly used with MEVD or SMEVD to parameterize cumulative function distributions (CDFs) are as follows: Weibull distribution, Gamma distribution or generalized Pareto distribution (GPD). Below is a brief description of the Gamma distribution, chosen for the parameterization of the data in this thesis.

Gamma distribution is described by two parameters α and β. α corresponds to the shape parameter and β corresponds to the scale parameter. The Gamma distribution is a probability distribution defined over the positive real numbers (R^+).

$$P(x) = Pr\{X \leq x\} = \frac{\gamma\left(\alpha, \frac{x}{\beta}\right)}{\Gamma(\alpha)} \qquad \textit{(Eq. 11)}$$

$$p(x) = \frac{1}{\theta^k \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

Where $x \geq 0$

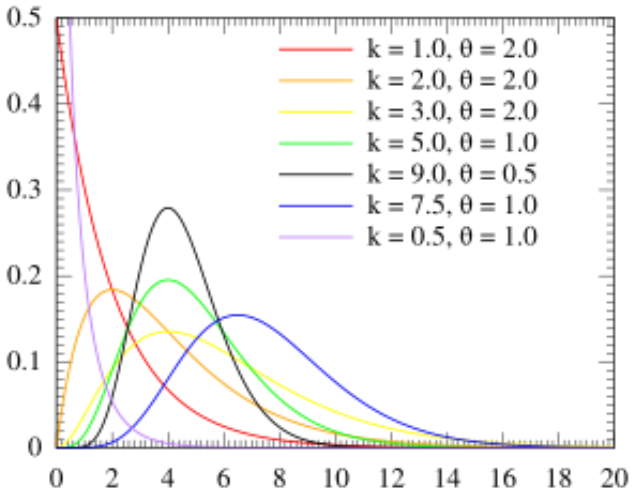


Figure 8 - Gamma distribution with variation of parameters α and β. The graph uses a different notation, but k = α e ϑ = ß.

2.2 Optimal choice of model parameters

In this chapter, it is shown how to identify and determine the parameters of the models seen in Chapter 2.1 on the basis of observed data sets. A fundamental prerequisite for proceeding in parameter identification is to consider a set of random, independent, identically distributed observations x = $(x_1, \dots, x_n)$. It is also considered a probability density function $p(x|\theta)$, defined in the domain $\Gamma = R^+$, where θ is a vector of parameters. The goal of this phase is to determine the vector θ, through the method of maximum likelihood or the Method of L-moments.

Method of Maximum Likelihood (ML)

This method was compared with the L-moments method in order to choose the one that best estimates the parameters of the chosen distribution. The method of Maximum likelihood is based on the assumption of the vector θ, consisting of the parameters $a_1, \dots, a_m$ and a sequence of random variables $x_1, \dots, x_n$ (IID), distributed according to the distribution $p(x|\theta)$. Under these assumptions it is possible to define the joint probability distribution function or also calls Likelihood function of X and θ:

$$\begin{aligned} L(\theta) &= p_{joint}(x|\theta) = p(x_1, \dots, x_n|a_1, \dots, a_m) \\ &= p(x_1|a_1, \dots, a_m) \cdot \dots \cdot p(x_n|a_1, \dots, a_m) \\ &= \prod_{i=1}^{N} p(x_i|\theta) \end{aligned} \qquad \textit{(Eq. 12)}$$

It is important to note that it is a probability distribution function (pdf), if x is variable and θ is fixed, conversely if x is constant and θ is variable then it can be taken as a likelihood function ($L(\theta)$). Sometimes, it can be convenient to work with the Likelihood function in logarithmic format:

$$l(\theta) = \ln[L(a_1, \dots, a_m)] = \sum_{i=1}^{N} \ln[L(a_1, \dots, a_m)]$$

In other words, the method of maximum likelihood aims to maximize the function $L(\theta)$ or $l(\theta)$. To choose the optimal values of the pdf's parameters, this can be possible by applying the following conditions:

$$\begin{cases} \dfrac{\partial L(a_1, \dots, a_m)}{\partial a_1} = 0 \\ \dots \\ \dfrac{\partial L(a_1, \dots, a_m)}{\partial a_m} = 0 \end{cases} \qquad \begin{cases} \dfrac{\partial l(a_1, \dots, a_m)}{\partial a_1} = 0 \\ \dots \\ \dfrac{\partial l(a_1, \dots, a_m)}{\partial a_m} = 0 \end{cases} \qquad \textit{(Eq. 13)}$$

or similarly:

$$\left.\frac{\partial L(\theta)}{\partial \theta_i}\right|_{\theta_i^*} = 0 \quad \Leftrightarrow \quad \left.\frac{\partial l(\theta)}{\partial \theta_i}\right|_{\theta_i^*} = 0$$

The system of equations to find the parameters of the distribution, must be solved numerically, since the calculations can be very time-consuming. The model also presents biases, on the basis of the considerations made.

Despite this, the ML method has some advantages such as: it can also be applied to distributions with divergent moments; it is very general and therefore has many applications; it allows all data to be studied by giving equal importance to each one; and it can be used even in the absence of some data, knowing only their position with respect to a certain threshold. However, the MLE may suffer from the presence of extreme events in cases where the sample is limited.

Method of L-Moments

The method of L-moments is a variant developed to overcome certain limitations of the classical method of moments. It is based on linear combinations, in fact the L stands for linear combination. This technique can be used to estimate the parameters of a distribution and is particularly useful when distributions have a heavy or asymmetric tail (distributions very common in hydrology and finance). The method of L-moments compared to the method of maximum likelihood has several advantages such as greater robustness and sensitivity in the case of outliers. This method is based on linear combinations of statistical orders. By statistical orders, one means a sample of data of size n and ordered in ascending order from X_1 the minimum and X_n the maximum. In addition, L-moments are created from statistical orders, but weighing them linearly, instead of raising them to powers, as in the classical method. This reduces the influence of extreme values. The

calculation of the L-moments on Probability Weighted Moments (PWMs) polynomials and the generic formula (of order r) can be expressed in this way:

$$b_r = \frac{1}{n}\sum_{i=1}^{n} X_i \cdot \frac{(i-1)(i-2)\cdots(i-r)}{(n-1)(n-2)\cdots(n-r)} \qquad \textit{(Eq. 14)}$$

where b_0 corresponds to the sample mean and b_1, b_2, b_3 are used to calculate L-moments. From equation 14, L-moments can be calculated, the first four in particular are L_1, L_2, L_3 and L_4. The first L-moment (L_1) represents the sample mean and corresponds to the first moment of the traditional method, defined by the following formula:

$$L_1 = b_0 = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad \textit{(Eq. 15)}$$

The second L-moment (L_2) corresponds to the variance and is used to measure dispersion. It is calculated with the following equation and is based on the first two PWMs:

$$L_2 = b_0 - 2b_1 \qquad \textit{(Eq. 16)}$$

The third L-moment (L_3) allows to measure the skewness, using b_0, b_1 and b_2 defined with PWMs. In other words, L_3 allows the asymmetry of a distribution to be measured.

$$L_3 = b_0 - 6b_1 + 6b_2 \qquad \textit{(Eq. 17)}$$

In the case analyzed in this thesis, the L-moments method was preferred to the maximum likelihood method for calculating the parameters of the gamma distribution calibrated on peak events. The parameters (α, β) of the gamma distribution can therefore be calculated using the following equations:

$$t = \frac{L_2}{L_1}$$

If $0 < t < 0.5$ then $z = \pi t^2$ and $\alpha = \frac{1-0.3080 \cdot z}{z-0.05812 \cdot z^2+0.01765 \cdot z^3}$

If $0.5 \leq t < 1$ then $z = (1-t)$ and $\alpha = \frac{0.7213 \cdot z-0.5947 \cdot z^2}{1-2.1817 \cdot z+1.1213 \cdot z^2}$

$$\beta = \frac{L_1}{\alpha} \qquad \textit{(Eq. 18)}$$

2.3 Model Performance Measures

Various methods such as goodness of fit, cross-validation, Kling-Gupta efficiency (KGE), Nash-Sutcliffe efficiency (NSE), Root Mean Square Error (RMSE) and Standard Deviation (σ) were used to measure the performance of the models. Specifically, this chapter explains the various methods in detail and highlights their main advantages.

2.3.1 Q-Q Plot

The goodness of fit of a sample can be observed using a graph, called the Q-Q Plot, which stands for Quantile-Quantile plot. It can be used to compare the distribution of two data sets, or it is often used to assess whether a data set follows a certain theoretical distribution. The method consists of considering a series of observations in ascending order, then assigning an index *i* that stands for the position of the data, and finally calculating the theoretical cumulative density function (CDF) of each observation (as shown in Table 1). The graph (see figure 9) is constructed by relating the theoretical quantiles on the y-axis and the sampled quantiles on the x-axis. The sampled quantiles correspond to the annual maxima extracted from ordinary events (e.g. if there are 30 years of data, 30 annual maxima are extracted).

X_i ordered	i	P*(X_i)
X_1	1	$\frac{1}{N+1}$
X_2	2	$\frac{2}{N+1}$
...	...	...
X_n	N	$\frac{N}{N+1}$

Table 1 - ordering procedur

The theoretical quantiles are estimated by the theoretical distribution taken into consideration, via the following equation:

$$P(x_i) \cong G(\hat{x}) \qquad \textit{(Eq. 19)}$$

$P(x_i) = \text{cumulative frequency of non} - \text{exceedance}$

$G(\hat{x}) = \text{theoretical distribution}$

$\hat{x} = \text{Theoretical quantiles}$

$x_i = \text{Sample quantiles}$

At this point, the graph can be constructed, and it is interesting to observe the closer points are to a 45° sloping line, the greater the fit with the distribution considered. The coefficient of determination, also known as R^2, was used to estimate the goodness of fit. This index measures the link between the variability of the data and the correctness of the statistical model used. An $R^2 = 1$ indicates that the model perfectly predicts the observed data, meaning it explains all the variability. It is calculated by the formula:

$$R^2 = 1 - \frac{SSR}{SST} \qquad \textit{(Eq. 20)}$$

$SSR = \sum_{i=1}^{n} (y_i - \hat{y}_i)\text{^}2 = \text{Sum of Squared Residuals}$

$SST = \sum_{i=1}^{n} (y_i - \bar{y})\text{^}2 = \text{Total Sum of Squares}$

$y_i = \text{observed values}$

$\hat{y}_i = \text{predicted values}$

$\bar{y} = \text{mean of values}$

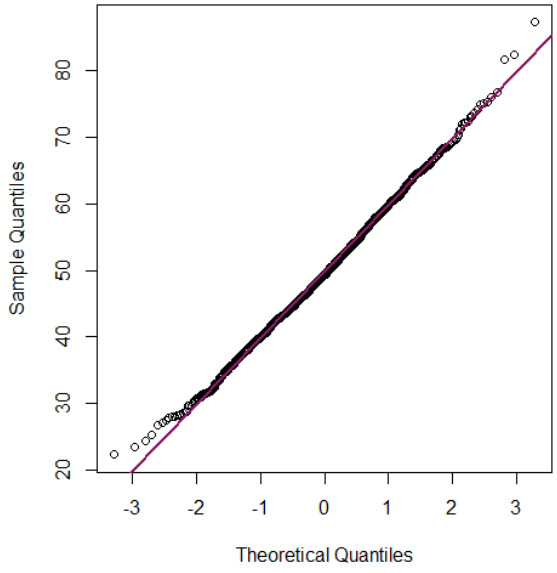


Figure 9 – Q-Q Plot

2.3.2 Cross Validation

Cross validation is a method used to evaluate the performance of the model and allows the prediction of the probability of unobserved events. The method consists of dividing the ordinary events of the flow peaks into two parts. The first part of data, also called the training set, is the data that will be used exclusively in the training phase of the model. During this phase, the model will learn the relationship between the input variables and the output variables. The second part of data, on the other hand, is called the validation set and is used to test the performance of the model. This procedure is repeated a number of times, on each occasion the two data sets (calibration and validation set) are randomly extracted from the sample of flow peaks of ordinary events (detailed explanation chapter 4.5).

2.3.3 KGE, NSE, RMSE and Standard deviation

Kling-Gupta Efficiency (KGE)

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \qquad \textit{(Eq. 21)}$$

- r: is the correlation coefficient between observed and simulated events;
- α: is the ratio of the standard deviation of the simulated and observed events;
- β: is the ratio of the averages of the simulated and observed events.

Takes into account correlation, bias, and variance to evaluate model performance. Ideal value = 1.

Nash-Sutcliffe Efficiency (NSE)

$$NSE = 1 - \frac{\sum_1^n (q_{est} - q_{obs})^2}{\sum_1^n (q_{obs} - \overline{q_{obs}})^2} \qquad \textit{(Eq. 22)}$$

- q_{obs}: Are the observed events;
- q_{est}: are the events predicted by the model.

Measures how well the model reproduces the observed data compared to using the average of the data. Ideal value = 1.

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_1^n (q_{obs} - q_{est})^2}{n}} \quad \textit{(Eq. 23)}$$

- q_{obs}: Are the observed values;
- q_{est}: are the events predicted by the model;
- n: is the number of observed values.

Measures the size of the mean square error between observed and predicted values.
Ideal value = 0.

Standard Deviation (σ)

The standard deviation is a statistical measure that quantifies the dispersion of data with respect to their mean. It indicates how far the data deviates on average from the central value (the arithmetic mean). A low standard deviation indicates that the data are concentrated around the mean, while a high standard deviation indicates that the data are more scattered and distributed away from the mean. The standard deviation is calculated using the following formula:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2} \quad \textit{(Eq. 24)}$$

Where n corresponds to the amount of data in the sample and $\bar{\varepsilon}$ average of the relative errors. The standard deviation was applied to the relative errors calculated between the observed and estimated flow rates, in order to quantify their dispersion from their mean. The equation used to calculate relative errors is as follows:

$$\varepsilon = \frac{q_{obs} - q_{est}}{q_{obs}} \quad \textit{(Eq. 25)}$$

3. STAGES OF ANALYSIS

In the following sub-chapters, the procedure followed to analyze the impact of initial soil moisture on the generation of flood events will be explained in detail. Specifically, three different methodologies for risk estimation and management of extreme events will be used. The methods are as follows: GEV distribution, Simplified MEV distribution and Modified MEV distribution (theoretical bases explained in chapter 2.1).

3.1 Description of sites selected

The choice of sites was made on the basis of several factors:

- Availability of hourly rainfall and discharge data;
- Small size of the basin;
- Absence or low presence of snow, which could compromise accounts for flood wave generation;
- Availability of stations equipped with instruments for measuring solar radiation, wind speed, humidity and temperature;
- Sufficiently high correlation coefficient between peak flows and cumulative soil moisture;
- Similar equation of the straight line below which cumulative soil moisture fails to generate a flow event;
- Basins tending to an erratic regime (see chapter 3.1.2).

Several basins, located in different parts of the Veneto region, were considered for this analysis. The analyzed basins are located in different locations: Posina at Stancari, Boite at Podestagno, Cordevole at Saviner, Fiorentina at Sottorovei and Padola at Santo Stefano.

Posina at Stancari

The first basin considered is located between mountain ranges called Piccole Dolomiti, in province of Vicenza. The highest peak is Cima Palon, which measures a height of 2232 m.a.s.l. The basin drains water into the Astico stream and covers an area of 117 km^2. The closure section considered, corresponds with the station for flow measurement, located at Posina in Stancari (near the village of Arsiero). The rain gauge stations considered are as follows: Molini (Lakes), Contrà Doppio (Posina), Castana, Passo Xomo (Posina), Brustole' (Velo d'Astico) and Astico in Pedescala.

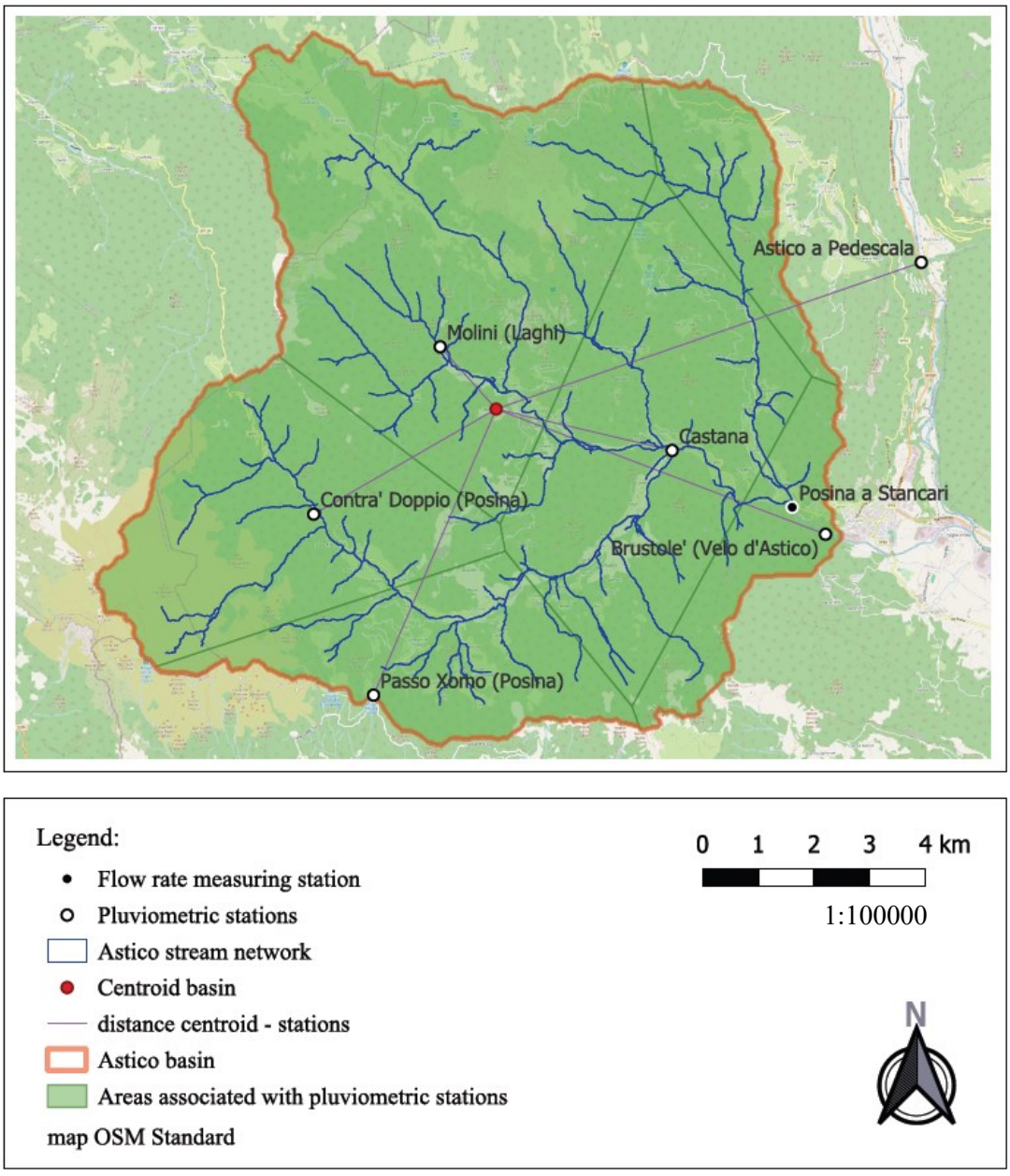


Figure 10 - Posina Basin

Boite at Podestagno

The second basin considered is located between mountain groups of the Dolomites, in province of Belluno. The surrounding peaks exceed 2500 m in altitude. The basin drains water into the Boite stream and covers an area of 30 km^2. The closure section considered, corresponds with the station for flow measurement, located at Boite a Podestagno. The rain gauge station considered is located near to station for flow measurement.

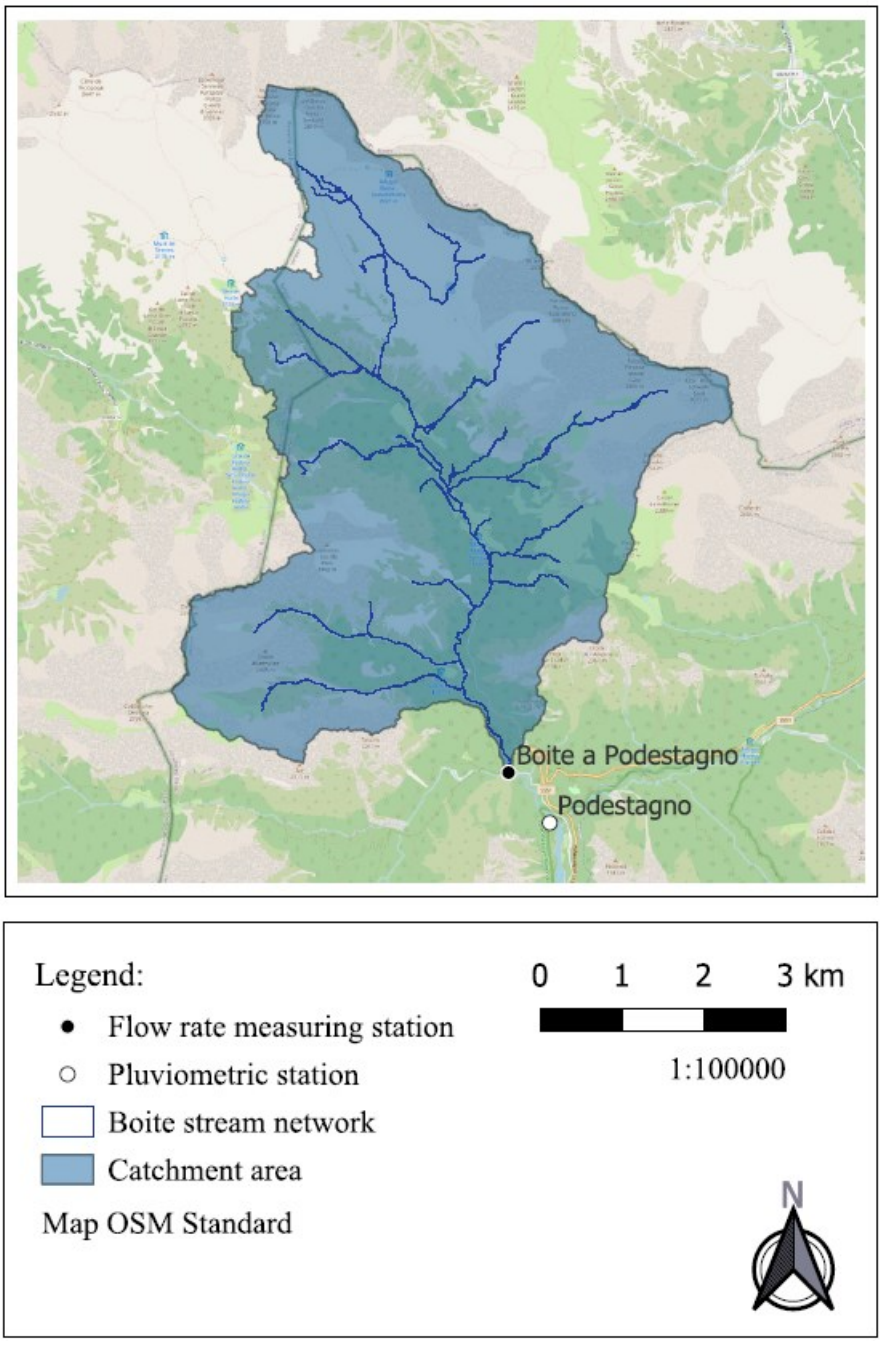


Figure 11 - Boite Basin

Cordevole at Saviner

The third basin considered is located is located between mountain groups of the Dolomites, in province of Belluno. The source is located at an elevation of 1919 m.a.s.l. The basin drains water into the Cordevole stream and covers an area of 105 km^2. The closure section considered, corresponds with the station for flow measurement, located at Cordevole a Saviner. The rain gauge station considered is located near to station for flow measurement.

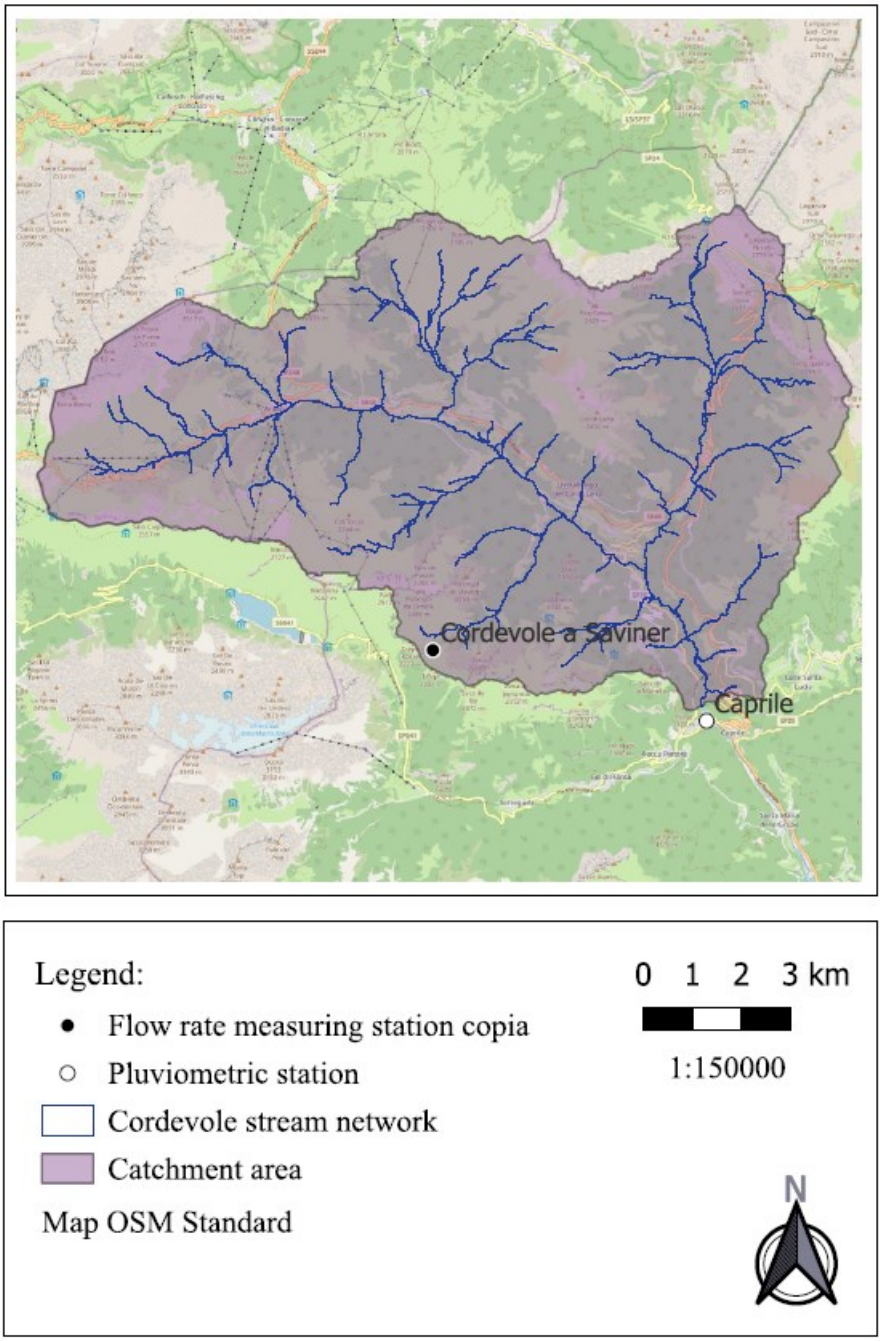


Figure 12 - Cordevole basin

Fiorentina at Sottorovei

The fourth basin is the Fiorentina basin and is located near Alleghe, in the Agordino territory. The highest peak is Cima Cernera, which measures a height of 2664 m.a.s.l. The basin drains water into the Fiorentina stream and covers an area of 57 km^2. The closure section considered, corresponds with the station for flow measurement, located at Fiorentina a Sottorovei. The rain gauge station considered is located near to station for flow measurement.

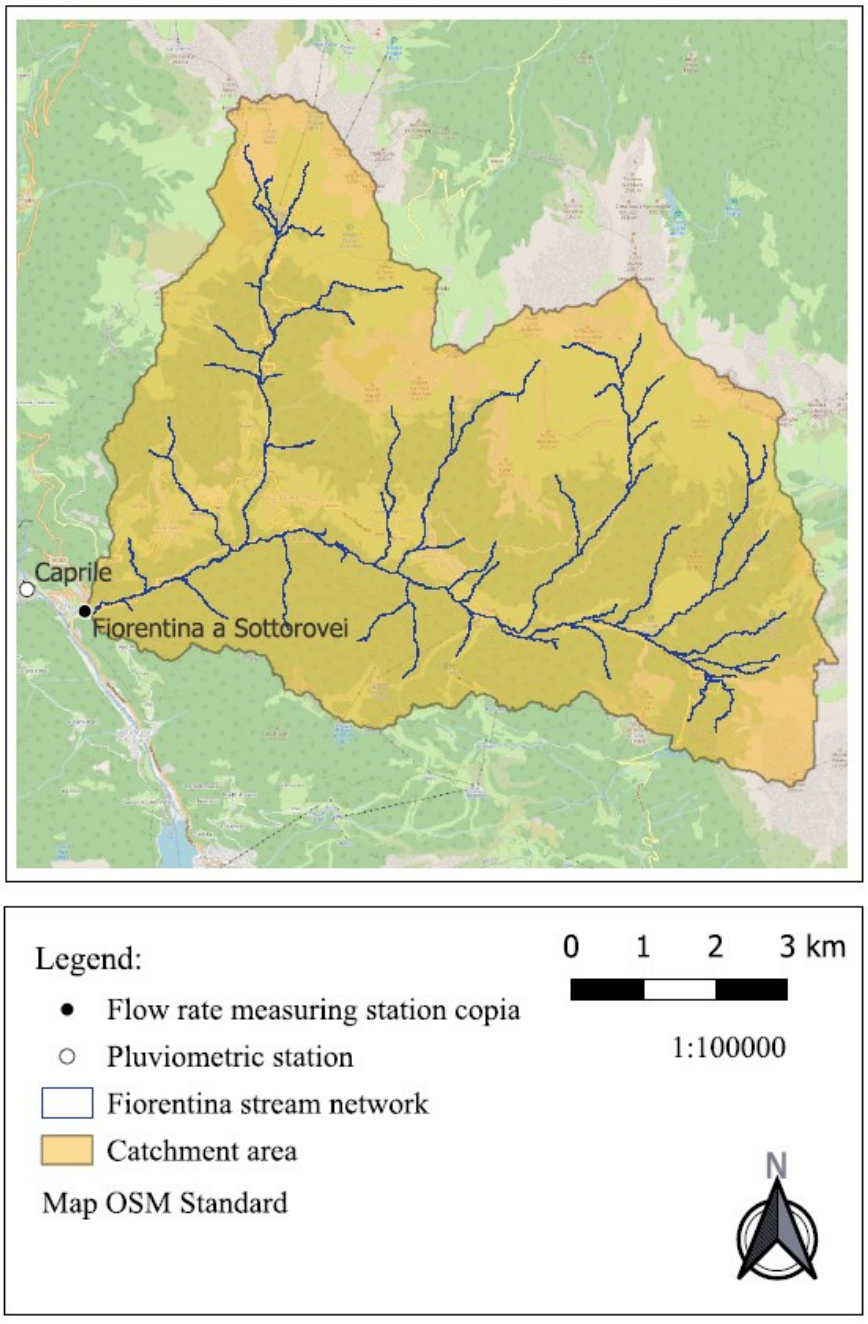
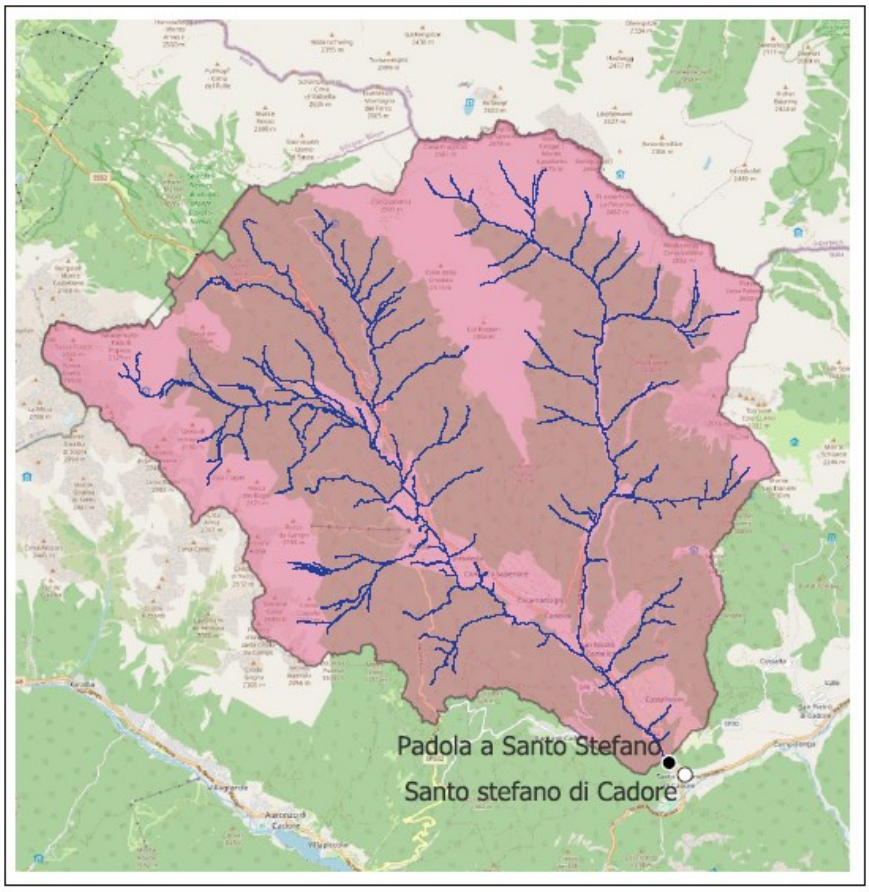


Figure 13 - Fiorentina basin

Padola at Santo Stefano

The last one is the Padola basin and it originates northeast of the Monte Croce di Comelico pass in the Altoatesino territory. The surrounding peaks reach altitudes over 3000 m.a.s.l. The basin drains water into the Padola stream and covers an area of 121 km^2. The closure section considered, corresponds with the station for flow measurement, located at Padola a Santo Stefano. The rain gauge station considered is located near to station for flow measurement.



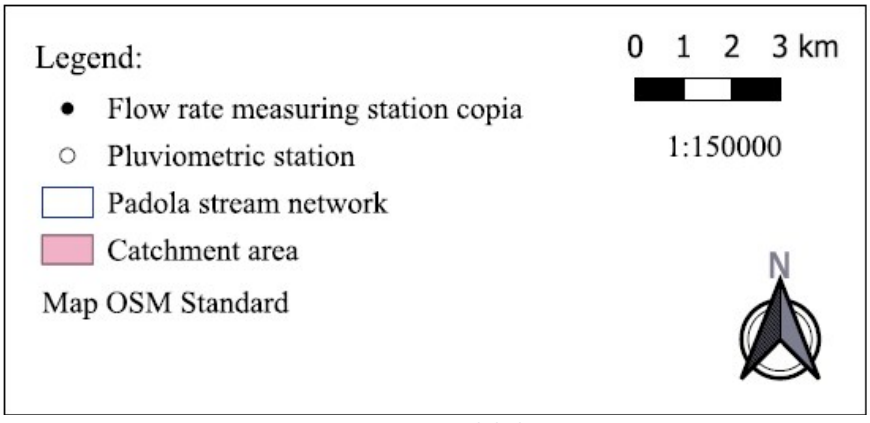


Figure 14 - Padola basin

3.1.1 Tools used for basin delineation

Catchment areas are defined as topographical areas delimited by a topographical watershed collecting water flowing over the ground surface into a specific water body. These areas are selected on the basis of land elevation, which is contained within a digital model called DTM (Digital Terrain Model). It is possible to see an example in image 16. The DTM files were downloaded from the Veneto Region Geoportal. They were acquired using remote sensing instruments called Lidar, which allow the distance of an object or surface to be determined using laser pulses. This instrument can be mounted in various vehicles, such as an airplane, helicopter or even a drone. Going into details, there are two types of surfaces that can be acquired by Lidar. The Digital Surface Model (DSM) and the previously mentioned Digital Terrain Model (DTM). The DSM considers as a surface, not only the ground, but also anthropogenic elements and vegetations that are on it. The DTM, on the other hand, represents the land surface free of obstacles. This surface is obtained by software interpolation from the DSM.

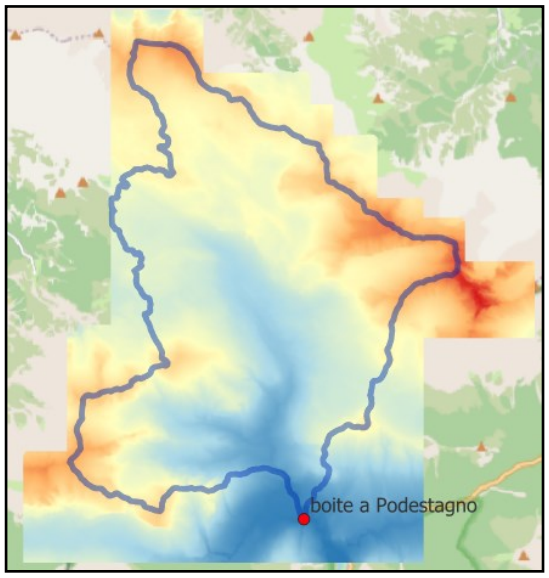


Figure 16 - Digital Terrein Model (DTM) of the Boite a Podestagno basin

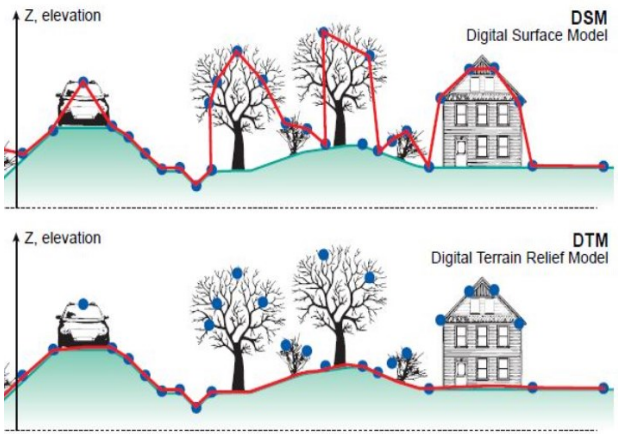


Figure 15 – DSM vs DTM

3.1.2 River flow regimes

It is interesting to note that streams can have different flow regimes. The flow regime refers to how water moves within the watercourse. The flow of water is influenced by various factors, such as the flow rate, the slope of terrain, the section of the channel and the friction between water and terrain. To study this behavior, you can observe the pdf of flow rates, in order to analyze the type of curve. Flow regimes can be classified into two distinct categories: erratic regime and persistent regime.

The erratic regime is characterized by: a small basin size; an unstable flow, therefore with many fluctuations in flow rate; high probability of drought in a river; high probability of the river flooding; high sensitivity to climate changes.
If the pdf is observed in relation to the flow rates, the erratic regime typically takes an asymmetrical, long-tailed form (see Figure 18).

The persistent regime is characterized by: a large basin size; a constant and predictable flow; low probability of drought in a river; low probability of river flooding; low sensitivity to climate change. If the pdf is observed in relation to the flow rates, the erratic regime typically takes on a symmetrical bell shape (see Figure 17).

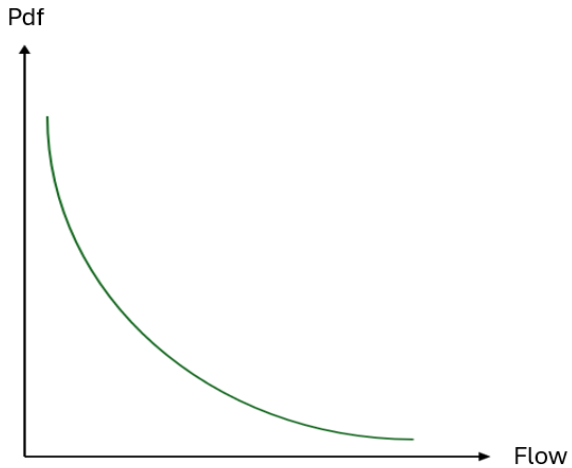


Figure 18 - Erratic regime

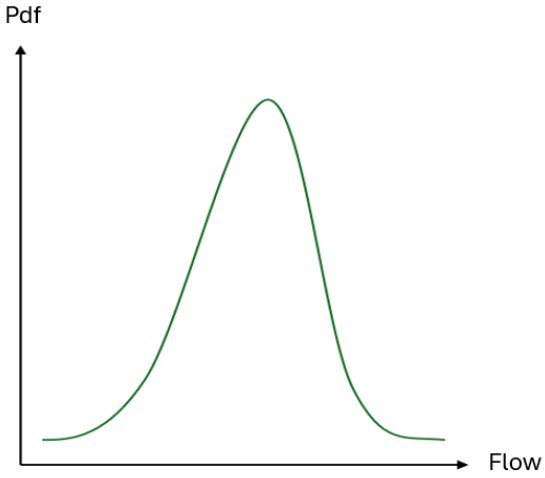


Figure 17 - Persistent regime

In the present case, all the streams in the five basins under analysis are characterized by an erratic regime. The Padola and Posina streams have a higher pdf for low specific flow values. This means that on average, the river regime is characterized by a lower specific flow (see figure 19).

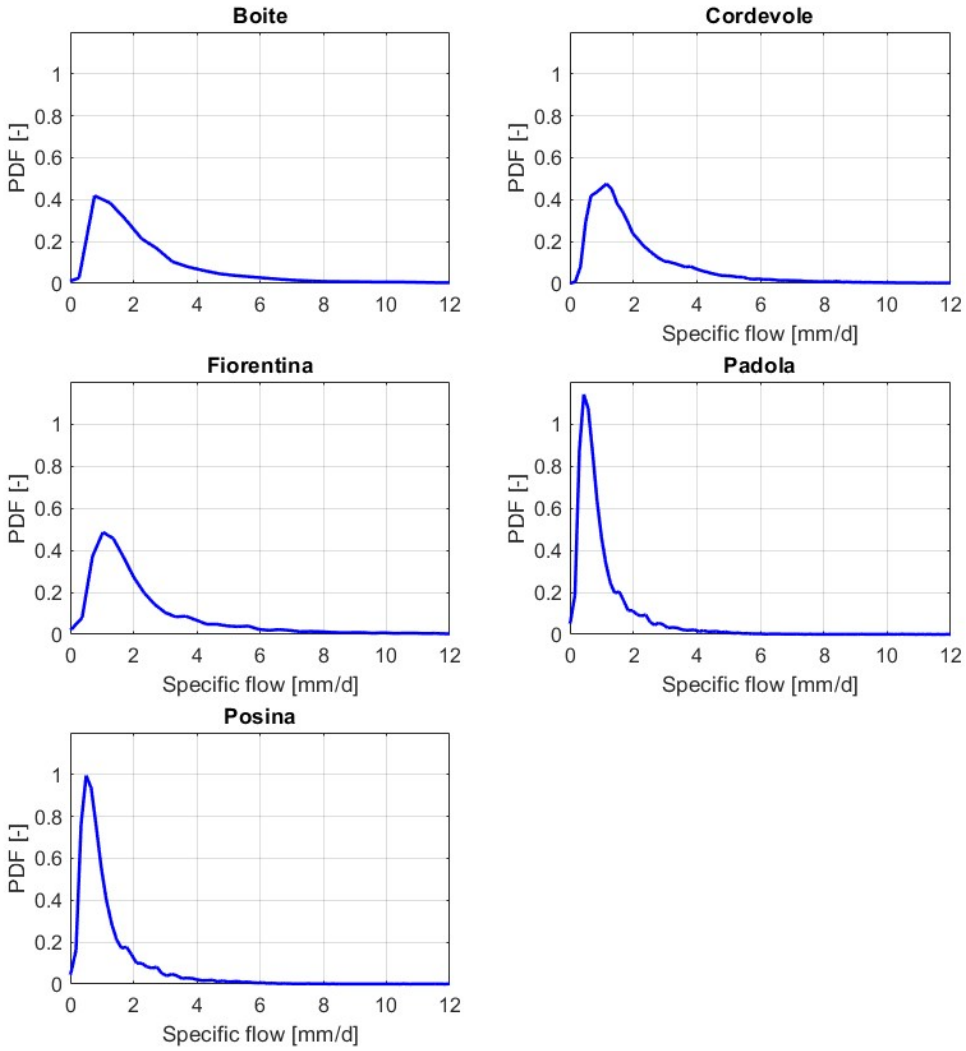


Figure 19 - CDF of the flow rates of the various streams.

3.2 Hydrological datasets

In this chapter, more information on the source data will be given and the procedure followed for calculating the antecedent soil moisture data will be shown. Soil moisture is defined as the result of the difference between cumulative precipitation and evapotranspiration (ET).

The rainfall, flow rate data and the various environmental characteristics of the different basins were requested on the website of ARPAV, a regional agency in the Veneto region that deals with environmental protection and prevention. Specifically, rainfall and flow rate data are defined on an hourly scale and have different time histories according to the basin considered. ARPAV's dedicated portal is called Ambiente Veneto and it is possible to consult and download the time series of data for the last 10 years. It can be reached from following link: https://www.ambienteveneto.it/stazioni/datiorari/ .

For the Boite torrent basin, the Podestagno (Cortina d'Ampezzo) rainfall station number 48 and the Boite flow station in Podestagno, identified by number 214, were considered. Both stations have a 31-year history of hourly data from 01/01/1993 to 31/12/2023.

For the Cordevole stream basin, the Caprile rainfall station number 9 and the Cordevole flow station in Saviner, identified by number 6, were considered. Both stations have a 31-year history of hourly data from 01/01/1993 to 31/12/2023.

For the Fiorentina stream basin, the same rainfall station was considered as for the Cordevole stream, Caprile number 9, and the Fiorentina flow station at Sottorovei, identified by number 6. Both stations have a 31-year history of hourly data from 01/01/1993 to 31/12/2023.

For the Padola torrent basin, the rainfall station Santo Stefano di Cadore number 58 and the discharge station Santo Stefano di Cadore, identified by number 62, were considered. Both stations have a 34-year history of hourly data from 01/01/1990 to 31/12/2023.

For the Astico torrent basin, 5 rainfall stations were considered Castana number 68, Astico a Pedescala (Posina) number 72, Contrà Doppio (Posina) number 73, Molini (Laghi) number 191, Passo Xomo (Posina) number 192 and the Posina flow rate station in Stancari, identified by number 407. Both stations have a 14-year history of hourly data from 01/01/2010 to 31/12/2023. Data on other environmental characteristics, such as temperature, humidity, wind, solar radiation, etc., are not available at all stations; only a

few have the appropriate instrumentation for measurement. Furthermore, the data have a daily scale and not an hourly one. Time gaps are present in all-time series of data; these could be due to station maintenance periods or instrument malfunctions. Graphs 20-23 show the rainfall time series of the various basins:

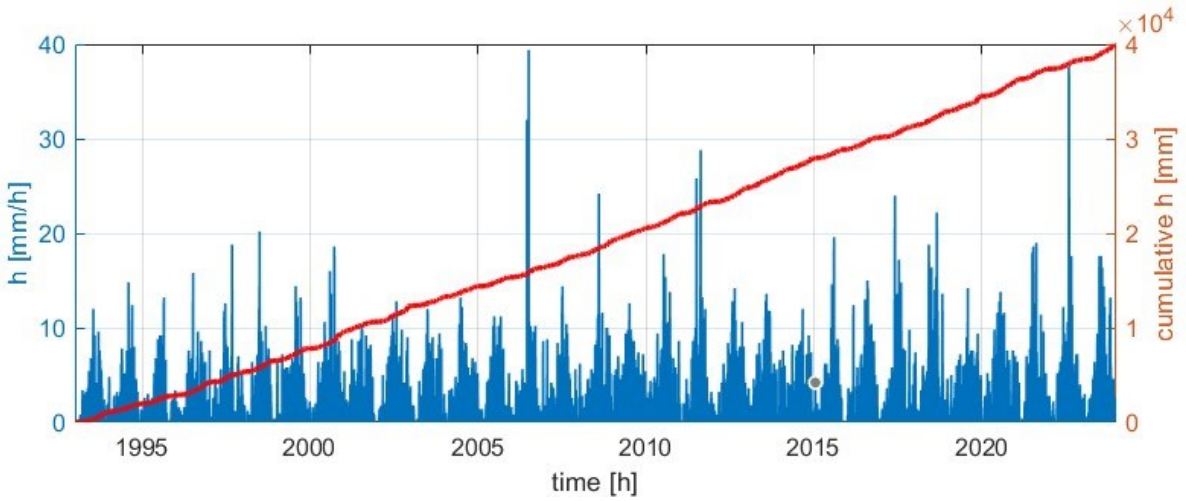


Figure 20 - Precipitation and cumulative precipitation in the Boite basin.

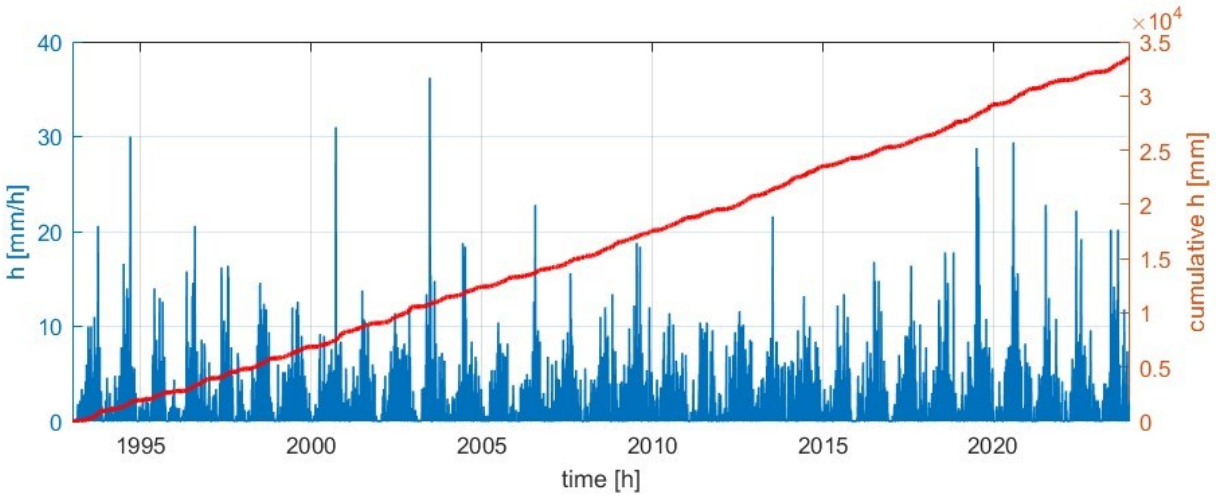


Figure 21 - Precipitation and cumulative precipitation in the Cordevole and Fiorentina basin (same rain gauge station).

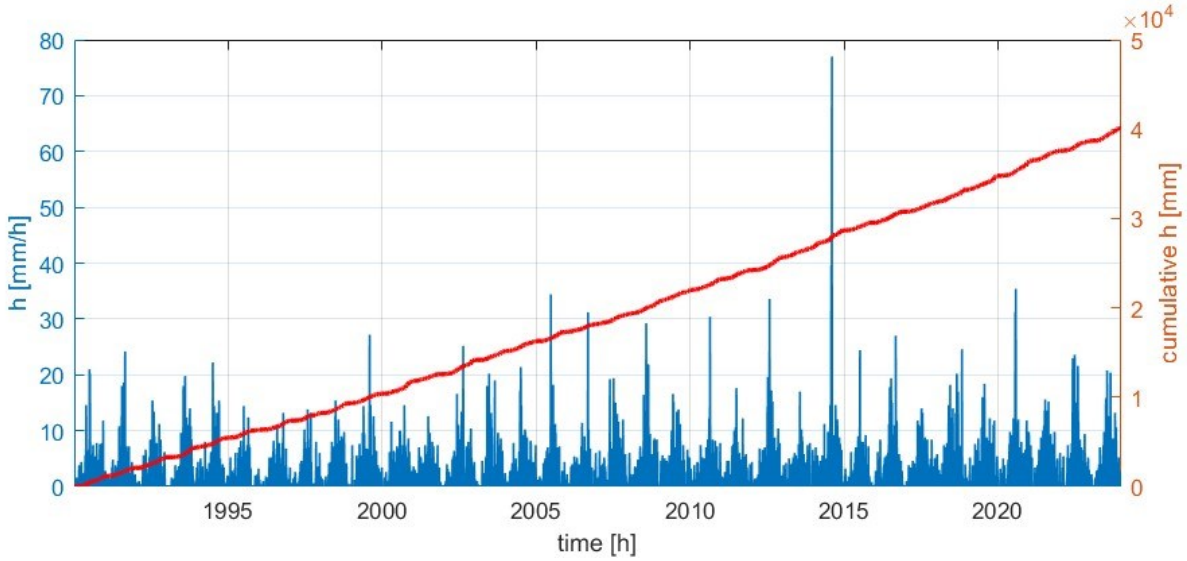


Figure 22 - Precipitation and cumulative precipitation in the Padola basin.

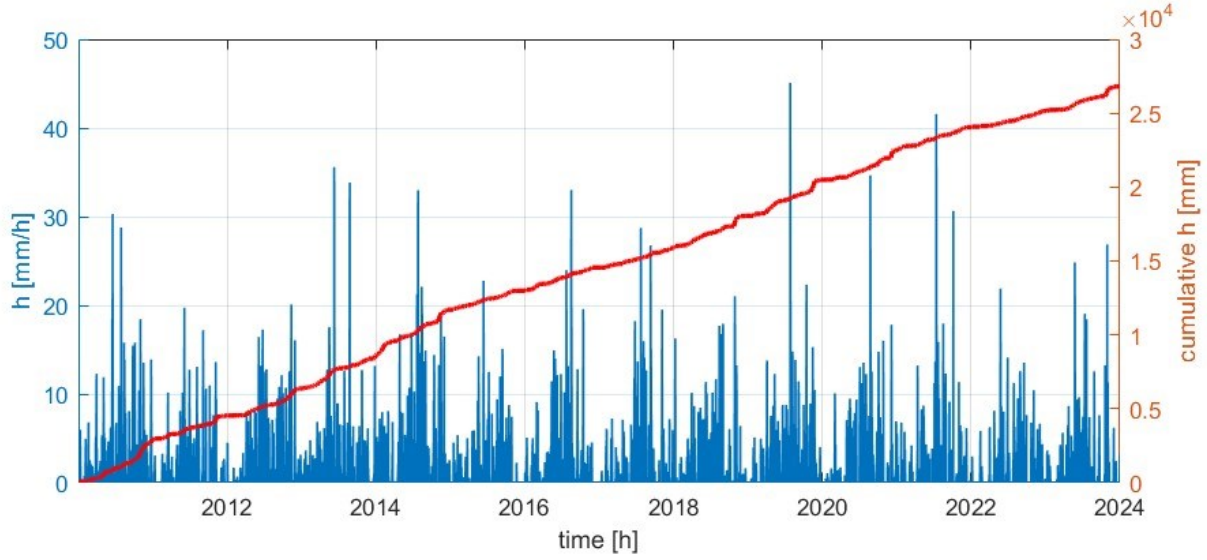


Figure 23 - Precipitation and cumulative precipitation in the Posina basin.

In some periods, some gaps can be seen where flow data are interrupted, these interruptions can be due to various reasons, such as: station maintenance, station installation before or after the considered time frame, etc. Graphs 24-28 show the time series of the flow rates of the various basins. It is also interesting to note, that during the period from 27 Oct 2018 to 3 Nov 2018, very intense flow events are measured in almost all basins (indicated in the graphs with a red circle). This period saw the occurrence of storm Vaia, a strong Mediterranean storm, characterized by hurricane-level wind gusts and heavy rainfall, which caused considerable damage mainly in the regions of Veneto, Trentino Alto Adige and Friuli Venezia Giulia.

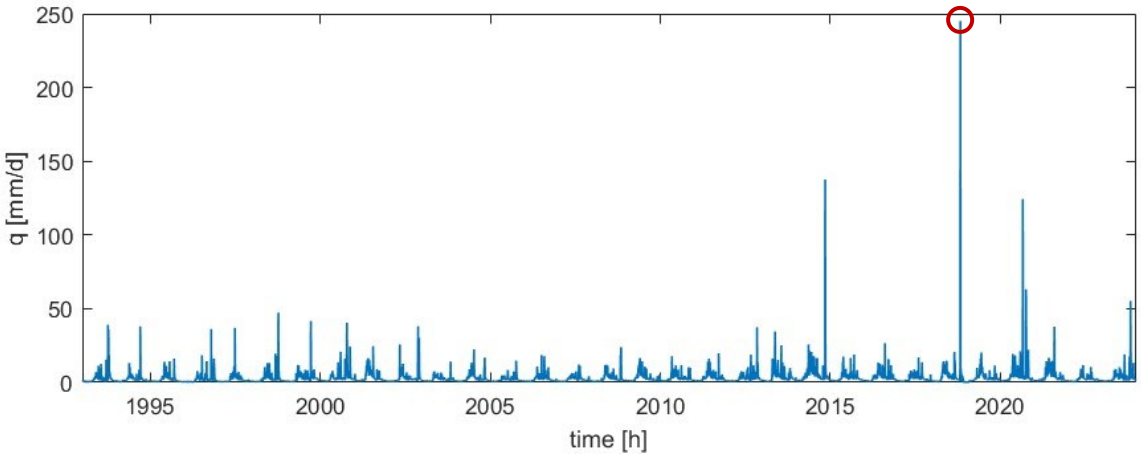


Figure 24 - Boite basin flows.

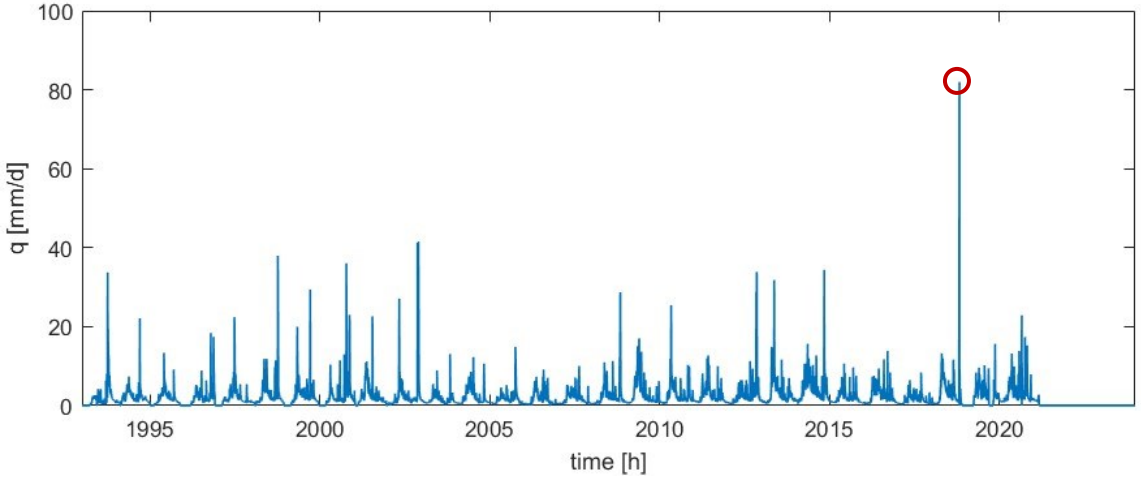


Figure 25 - Cordevole basin flows.

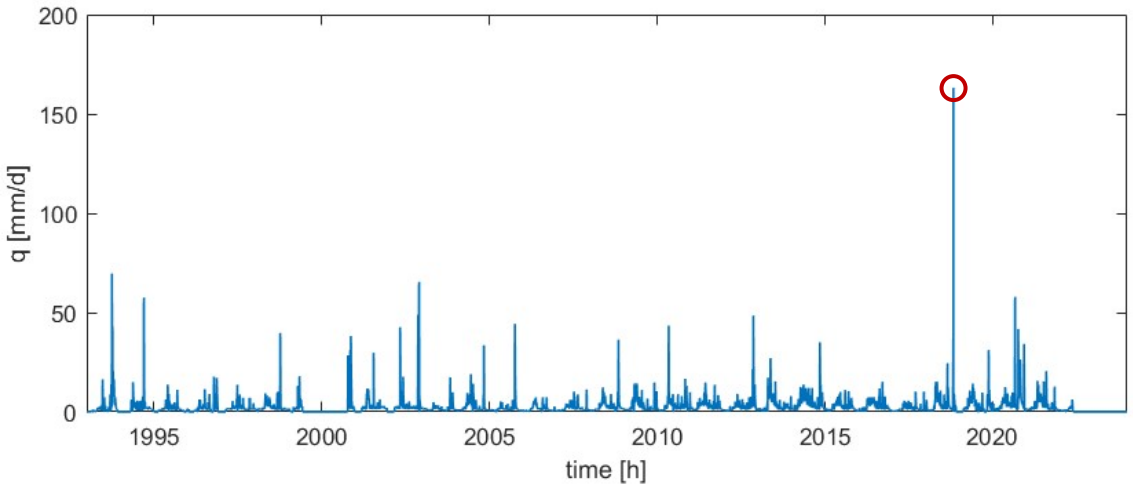


Figure 26 - Fiorentina basin flows.

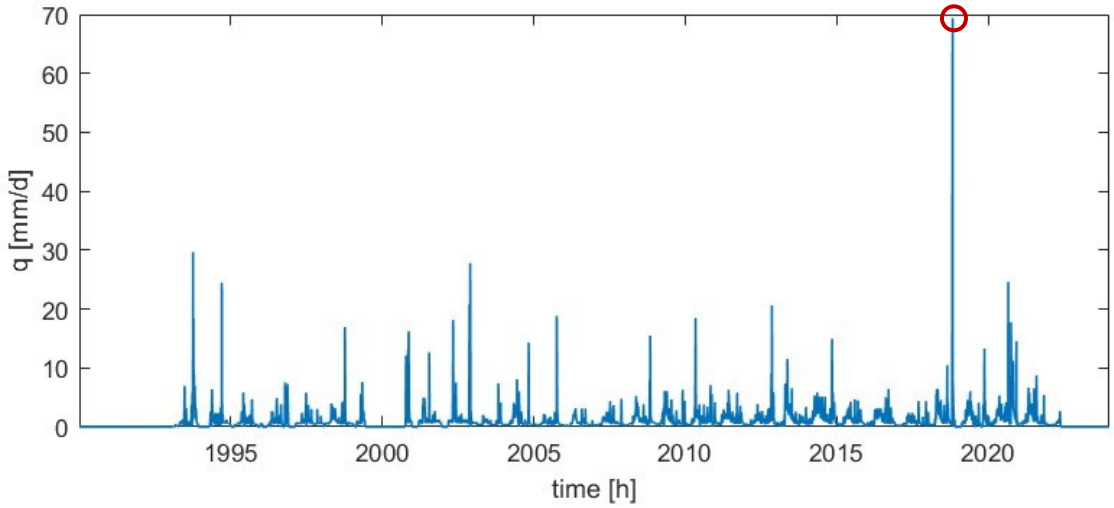


Figure 27 - Padola basin flows.

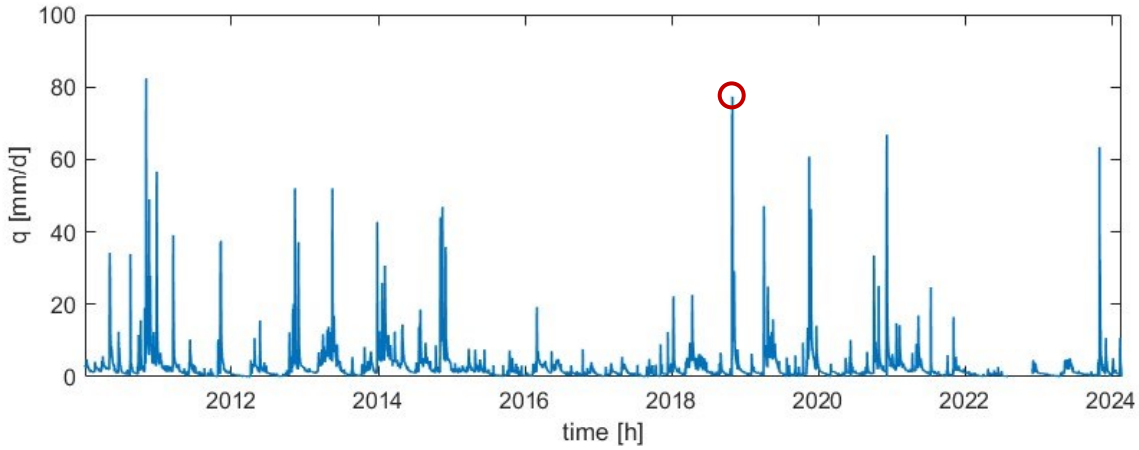


Figure 28 - Posina basin flows.

As explained in chapter 2.1 the advantage of MEV in general is that it can exploit a much larger amount of data to be able to study the behavior of extreme events more precisely. To consider extreme flow events, a hybrid method was used which is based on the two defined approaches of extreme events theory, which are as follows: Block Maxima Theory and Peak Over Threshold Theory (see chapters 2.1.1 for an exhaustive description). This method consists of assuming an increase in flow rate between a local minimum and a peak event; the threshold is different for each basin. The thresholds for different sites were chosen so that they are statistically comparable. This was achieved by observing the entire distribution of flow rates and calculating the associated quantiles. Another assumption made to be able to eliminate the noise in the data (by noise is meant the anomalous fluctuations, due to possible measurement errors of the rain stations) is to consider the data set average, to obtain a smoothed signal. This was done by considering a moving window with an overlap of a 6 hours amplitude, the following method allows the position of the peaks to be maintained over time. The width of the window was not chosen a priori, but trying to have the right number of events, which is neither too high or too low (e.g., considering a 24, 12 or 3 hours window). The resulting dummy curve is exploited to extract the flow rate maxima. This curve was used to consider actual peaks and not peaks due to instrumentation noise.

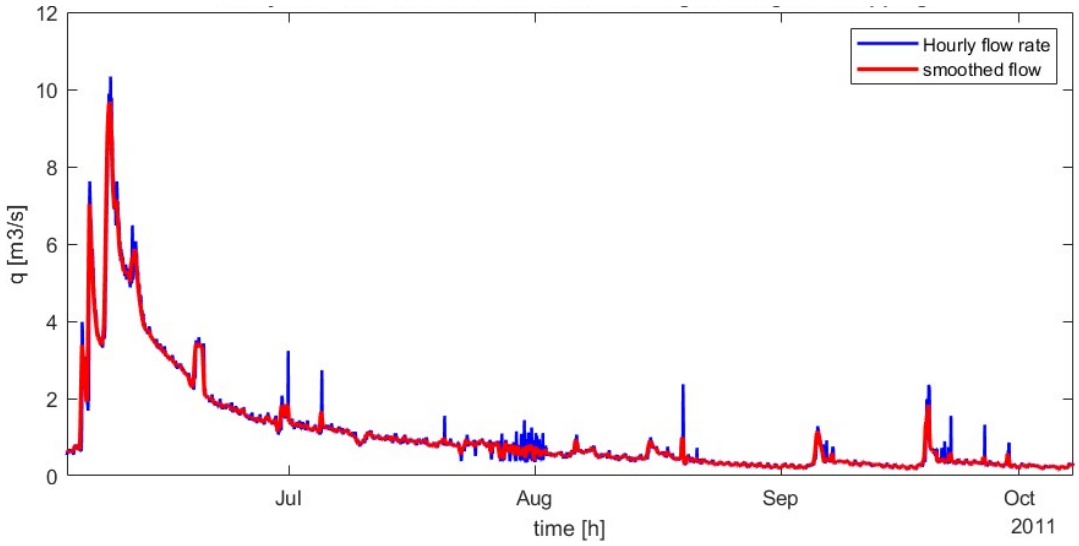


Figure 29 - Zoom in: Hourly flow rate vs. flow rate obtained through average overlapping.

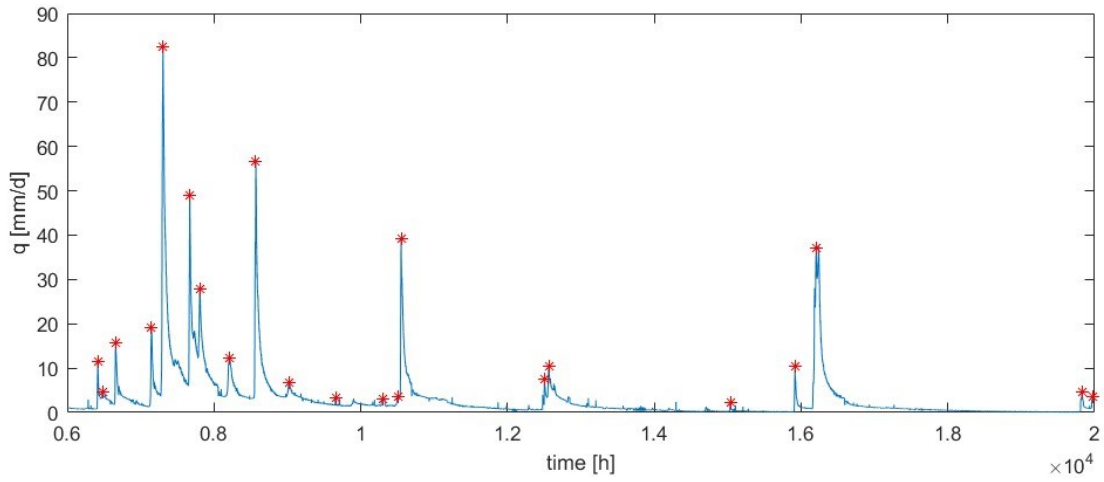


Figure 30 - Zoom in: Graph with peak flow rates highlighted.

The number of maximum flow events extracted from the source dataset changes for each basin. They are summarized in table 2.

basins	number of events
Boite	387
Cordevole	397
Fiorentina	344
Padola	484
Posina	195

Table 2 – Number of events for each basin

3.3.1 Cumulative soil moisture content

The other important factor to consider is precisely the cumulative soil moisture content, which is given by the subtraction of cumulative rainfall and cumulative evapotranspiration. It participates in the formation of the flood wave, so it is important to choose the number of cumulative days carefully. To do this, various ranges of days were considered for comparison: 30, 25, 20, 15, 10 and 5. For the selection of the intervals, the 6 hours immediately preceding the peak events were excluded so that only the soil moisture antecedent to the flood wave of the event considered was considered.
To calculate cumulative precipitation, the rainfall prior to the events was simply summed within a certain considered range (30, 25, 20, 15, 10, 5 d). To calculate Evapotranspiration, on the other hand, the Penmann-Monteith equations were used according to the FAO approach.

3.3.2 Evapotranspiration (ET)

The Evapotranspiration process is constituted by two distinct parts:

- Evaporation: a fraction of water present in the soil is transformed into water vapors thanks to incoming solar energy. This phenomenon occurs in the upper horizon of the Root Zone;
- Transpiration: a fraction of water is incorporated by plants and then it is released into the atmosphere as water vapors. This phenomenon involves the whole Root Zone.

Evapotranspiration is calculated with the Penmann-Monteith equations according to the FAO approach. The main idea of this methodology is to separate the dependences on the climate, vegetation and water availability in order to better study the ET process.
So, by applying the FAO method there are three steps to follow:

1. Computation of the reference potential evapotranspiration ET_0 that a reference crop produces during its growing season in the absence of water stresses under the actual climate conditions;
2. Computation of the potential evapotranspiration ET_C that the actual crop produces in its actual live cycle in the absence of water stresses under the actual climate

conditions. This term is obtained as the product of the crop coefficient K_C and the reference potential evapotranspiration ET_0:

$$ET_C = k_C(t) \cdot ET_0$$

3. Computation of the actual evapotranspiration ET that the actual crop produces in its actual live cycle under the actual environmental conditions and the actual climate conditions. This term is obtained as the product of the water stress coefficient K_S and the potential evapotranspiration ET_C:

$$ET = k_S(s(t)) \cdot ET_C = k_S(s(t)) \cdot \ k_C(t) \cdot \ ET_0$$

ET_0 is a stochastic parameter that depends on climatic conditions. In this relationship it is assumed as constant within the various seasons and is derived through the meteorological conditions of Feltre, through the Penmann-Monteith equation:

$$\lambda \cdot ET_0 = \frac{\Delta}{\Delta + \gamma \cdot \left(1 + \frac{r_s}{r_a}\right)} R_n + \frac{\rho \cdot \frac{C_p}{r_a}}{\Delta + \gamma \cdot \left(1 + \frac{r_s}{r_a}\right)} VPD$$

From which the inverse formula is obtained

$$ET_0 = \frac{0.408 \cdot \Delta}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)} (R_n - G) + \frac{\gamma \cdot \frac{900}{T + 273} \cdot u_2}{\Delta + \gamma \cdot (1 + 0.34 \cdot u_2)} VPD$$

(Eq. 26)

The Penmann-Monteith equation is composed of two terms: the first is called equilibrium evapotranspiration and depends on the net solar radiation Rn and the heat flow that propagates towards the deep layers G; while the second refers to the non-saturation of the air and wind.

A series of values were calculated for each basin; for example, a summary table is given for the Fiorentina basin. The following formulas were applied below.

- the psychrometric constant (γ) is a function of the pressure detected at an altitude of z = 264 s. l. m.

$$p = 101.325 \cdot (1 - 2.26 \cdot 10^{-5} \cdot 264)^{5.256} = 98.19 \ KPa$$

$$\gamma = p \cdot 0.665 \cdot 10^{-3} = 0.07 \ KPa/°C$$

- Mean temperature (T)

$$T = \frac{T_{min} + T_{max}}{2}$$

- the wind speed at two meters above the ground (u_2), using the Arpav data provided at a height of five meters.

$$u_2 = u_z \cdot \frac{\ln\left(\frac{z}{z_0}\right)}{ln\left(\frac{z_{ref}}{z_0}\right)}$$

u_z = wind speed 5 meters above the ground

z = 2 m above the ground

z_{ref} = reference hight, in this case equal to 5 meters

z_0 = 0.2 m, roughness length, that change in function of the roughness class

- Air vapor pressure deficit (VPD), which corresponds to the difference between the partial water vapor pressures in condition of saturation e_s^* and in actual condition e_a

$$e_s^*(T) = 0.6108 \cdot e^{\frac{17.27 \cdot T}{T+273.3}}$$

$$e_s^* = \frac{e_s^*(T_{max}) + e_s^*(T_{min})}{2}$$

$$e_a = \frac{e_s^*(T_{max}) \cdot RH_{min} + e_s^*(T_{min}) \cdot RH_{max}}{200}$$

$$VPD = e_s^* - e_a \qquad \textit{(Eq. 27)}$$

- The slope of the curve of Clausius Clapeyron (Δ), obtained by averaging the value associated with the minimum and maximum temperature.

$$\Delta(T) = \frac{4098 \cdot e_s^*(T)}{(T + 273.3)^2}$$

$$\Delta = \frac{\Delta(T_{max}) + \Delta(T_{min})}{2}$$

- the heat flow (G) that propagates towards the deep layers, was assumed null.
- The Net Radiation (R_n), function of the global radiation that reaches the ground (R_s), of the Albedo α assumed equal to 0.23 and the energy emitted by the earth's surface (R_{nl}).

σ = Constant of Stefan-Boltzmann $= 4.90 \cdot 10^{-9}\ MJm^{-2}k^{-4}d^{-1}$

$$R_a = \frac{R_s}{0.18 \cdot \sqrt{T_{max} - T_{min}}}$$

$$R_{s,0} = (0.75 + 2 \cdot 10^{-5} \cdot z) \cdot R_a$$

$$R_{nl} = 0.5 \cdot \sigma \cdot (T_{max}^4 - T_{min}^4) \cdot (0.34 - 0.14 \cdot \sqrt{e_a}) \cdot (1.35 \cdot \frac{R_s}{R_{s,0}} - 0.35)$$

$$R_n = (1 - \alpha) \cdot R_s - R_{nl} \qquad \textit{(Eq. 28)}$$

Following these steps, the various ET_0 values for each basin were obtained.

At this point, soil moisture can be calculated as the difference between cumulative precipitation and evapotranspiration (ET_0), taking care to exclude the 6 hours before peak flow events.

3.3 Joint Distribution of flow peaks and antecedent soil moisture

Before being able to study the distribution of extremes, applying SMEVD and GEV distribution, it is necessary to relate peak flow and antecedent soil moisture, considering different time lag, in order to choose the one with the greatest correlation coefficient. Antecedent soil moisture is considered to be equal to net precipitation, calculated as the difference between cumulative precipitation and cumulative evapotranspiration, prior to flood events. The periods before the events are considered by assuming a window of fixed duration but varying between 5 and 30 days. This procedure makes it possible to identify the time range that closely approximates the hydrological response of a catchment. The hydrological response refers to the way a catchment reacts to weather events, such as rainfall, by creating an outflow of water (flow). There are two main types of outflows, which contribute to the formation of the flood wave: surface runoff, which usually occurs within hours, and sub-surface runoff, which can take up to weeks. The joint distribution is very useful in this case, because it allows several variables to be related at the same time, referring to the same area or subject. In this case, the variables are peak flow and soil moisture, and the area they refer to is the various basins considered.

The joint probability of two variables can be defined by the following formula:

$$P(x, y) = P[X < x \; e \; Y < y] \qquad \textit{(Eq. 29)}$$

to which corresponds a joint probability density

$$p(x, y) = \frac{P[X \in (x, x + dx) \; e \; Y \in (y, y + dy)]}{dx \; dy} \qquad \textit{(Eq. 30)}$$

Where X and Y are random variables.

The marginal distribution, on the other hand, refers to the probability of the variables contained in the subset of a collection of random variables. In this case, therefore, all relative frequencies, calculated with the joint distribution, referring to peak flow and soil moisture are added together. The results of the joint distribution are reported in chapter 4.2.

3.4 Three models in comparison

This chapter will explain the procedure for applying the three models SMEVD single class, moisture accounting SMEVD and the GEV distribution to events. These three models, which are explained in detail in Chapter 2.1, allow modelling the distribution tails of extreme events. The GEV distribution by definition only needs to be applied to annual maximum events, so it exploits a limited amount of data to study extreme events. In contrast, the SMEVD exploits a much larger number of data, selecting all peak flows, to make a more precise description of the extremes. What is to be demonstrated in this thesis is precisely the advantage of using the moisture accounting SMEVD. Before the methods can be applied, however, it is necessary to parameterize the distributions to the events.

Parameterization involves the fitting of distributions to events. There are many distributions that can be used for this step, such as the Gamma distribution, the Weibull distribution and the generalized Pareto distribution (GPD). In this case, the Gamma distribution was chosen as it has a better fit with the data. Through the parameterization step, what one wants to do is estimate the parameters of the distribution so that it best describes the data to which it relates. Various statistical techniques can be used, such as the Maximum Likelihood (ML) Method or the method of L-moments (see Chapter 2.2). In this case, the L-moments method was preferred as it allows a more robust estimation of extreme events and greater stability for small samples. The true similarity method, on the other hand, was found to tend to underestimate the most extreme events, giving more weight to intermediate events. The parameters of the Gamma distribution that are estimated through parameterization are two: α scale parameter and β shape parameter. The Gamma distribution was chosen and compared with the Weibull distribution and the generalized Pareto distribution (GPD) through the goodness-of-fit step (which will be described in Chapter 3.5). To quantify the goodness-of-fit of the distributions, the coefficients of determination R^2 were calculated, referring to the qq-plots (see Table 3). In some cases, the three distributions have very similar R^2 coefficients, except for the generalized pareto distribution (GPD) applied to the single-class SMEVD, which fails to adjust correctly to events. The Gamma distribution was chosen, since in most cases it is the distribution that best fits the events. The procedure followed is slightly different for

the moisture accounting SMEVD, as it requires the division of soil moisture into classes before applying the parameterization.

basins	Types of distribution	R^2		
		SMEVD single class	Moisture accounting SMEVD	GEVD
Boite	Gamma	0.16	0.69	0.86
	Weibull	0.15	0.37	
	Gp	-	0.62	
Cordevole	Gamma	0.37	0.82	0.92
	Weibull	0.36	0.71	
	Gp	-	0.81	
Fiorentina	Gamma	0.35	0.70	0.91
	Weibull	0.33	0.62	
	Gp	-	0.75	
Padola	Gamma	0.30	0.72	0.91
	Weibull	0.31	0.56	
	Gp	-	0.72	
Posina	Gamma	0.61	0.74	0.93
	Weibull	0.60	0.72	
	Gp	-	0.67	

Table 3 - Coefficients for determining the various models, types of distributions and stream basins.

In order to be able to apply the moisture accounting SMEVD, it is necessary to partitioning the soil moisture data sample into classes, which allows for a better fit of the SMEVD to the data. The division into classes therefore makes it possible to study soil moisture dynamics more specifically, creating classes of events with different moisture levels. Furthermore, this approach should allow a better description of the tail of distribution. The number of classes chosen for the various basins is three, class 1 for low values, class 2 for average values and class 3 for high levels of soil moisture. Once the various moisture classes have been created, the peak flow events associated with each moisture value are also divided into the classes, maintaining and not changing the relationship that exists between each moisture value and the associated flow events. The classes are not the same but change both in amplitude and in number of events within them. The class with low humidity values is the one with the highest number of events within it (so the most likely) and will allow a more precise description of the tail of the fit distribution.

Image 31 is an example of Gamma distributions applied to the three different soil moisture classes, referring to the Fiorentina basin. Graphs of all basins have not been included, as they are very similar to each other.

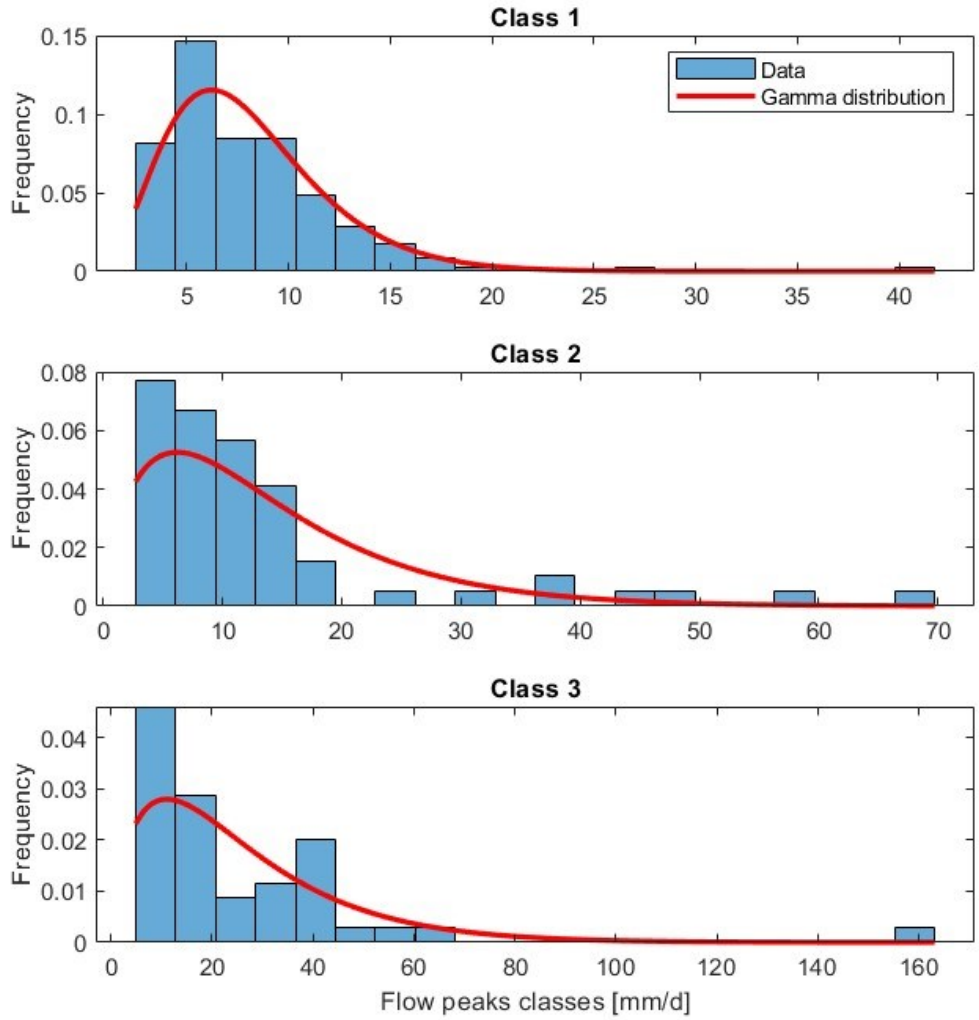


Figure 31 - Fit of Gamma distribution on peak flow rates divided into classes, defined by subdividing cumulative soil moisture data (referring to the Fiorentina basin).

According to the SMEV and GEV model definitions, once the parameters of all Gamma and GEV distributions have been calculated, it is necessary to calculate the associated cumulative density functions. In the theoretical equations, it is denoted as F(x) (see chapter 2.1). At this point, one has all the tools to calculate the cumulative density functions associated with the three distributions.

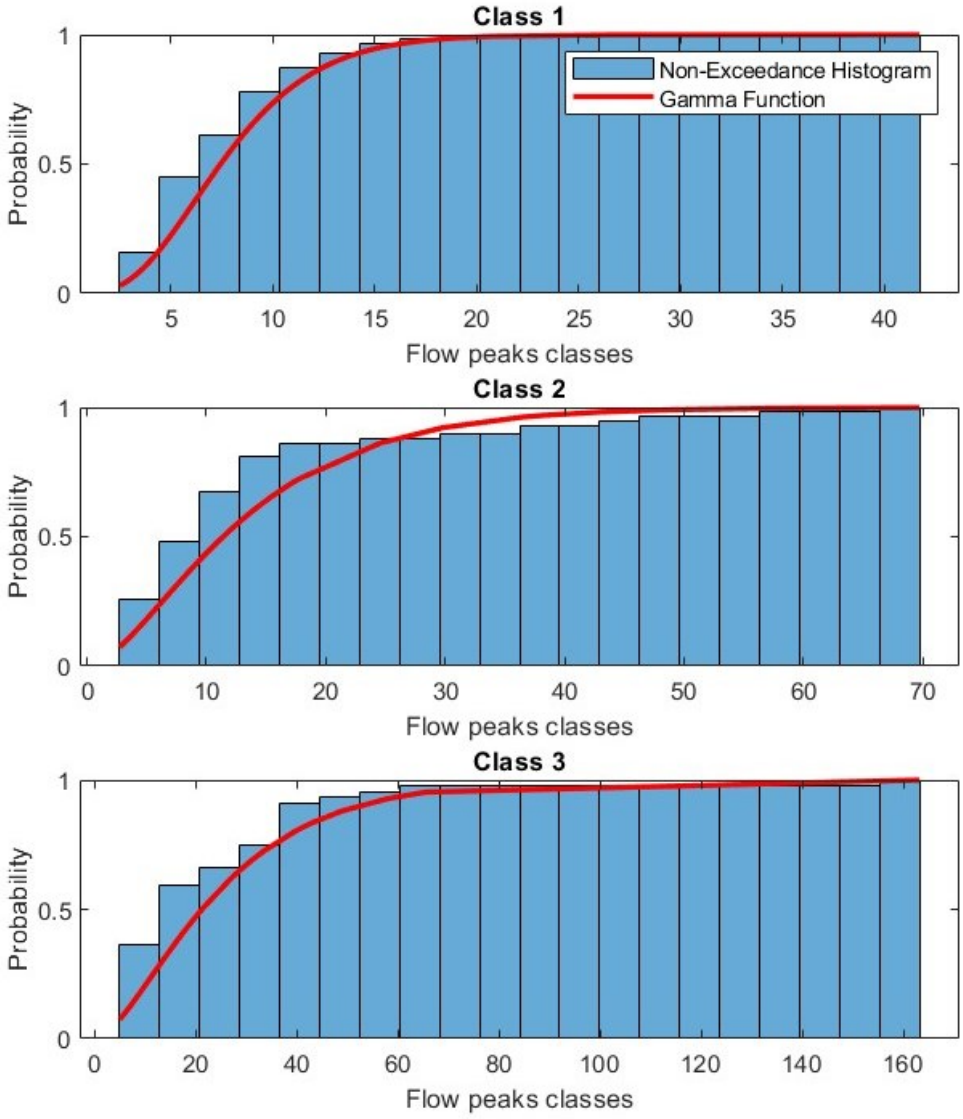


Figure 32 - Cumulative density function (CDF) of the Gamma distribution, applied to three soil moisture classes (for Fiorentina stream basin).

3.5 Goodness of fit

To verify that the distributions of the three methods fit the soil moisture data samples correctly, a graph called qq-plot was used (see Chapter 2.3.1). To calculate theoretical quantiles, the equation 16 in the generic form was used:

$$P(x_i) \cong G(\hat{x})$$

Specifically, to calculate the theoretical quantile associated with the empirical CDF, a Matlab code had to be used to invert and numerically solve the non-linear equation.

The Goodness of fit using the moisture accounting SMEVD, was done by considering theoretical distribution equal to the cumulative density function of the model. Considering $G(x)$ equal to this equation:

$$G(x) = F_1(x)^{n_1} \cdot F_2(x)^{n_2} \cdot F_3(x)^{n_3}$$

The G(x) distribution was fit using all peak flow data, while only annual maxima were used to make the Q-Q Plot. In other words, if we consider a period of 30 years, a maximum flow rate value is extracted for each year, until 30 values are obtained. The calculation of goodness of fit of SMEVD with a single class is similar to the goodness of fit of the moisture accounting SMEVD, the only difference that in this case a soil moisture class is considered. The theoretical distribution will be equal to:

$$G(x) = F(x)^n$$

In the end, it was calculated the goodness of fit of GEV distribution. In this case, the goodness of fit was calculated for the distribution with the same sample as the parameters of the distribution. The data sample corresponds to the maximum annual flow events.

3.6 Climate change scenarios

After testing the performance of the models through cross-validation, three climate change scenarios were estimated based on certain assumptions. To this purpose in the scenarios, the moisture accounting SMEVD is used, which allows greater control over the events as they are divided according to associated soil moisture. The first climate change scenario was assumed by considering a 50% decrease in the frequency associated with flow events (and thus also rain events) associated with the highest soil moisture class. Flow events with higher intensity are contained in this class. Then the events removed from the wettest class are reinserted into the class with intermediate moisture or the class with driest soil (it does not matter which class they are placed in). Events are removed randomly and the process repeated 1000 times in order to obtain an average climate change scenario. To represent this scenario through a graph, it was chosen to relate the return time (Tr) on the x-axis and the maximum flow events on the y-axis. In addition, the scenario referring to current climate conditions was included to compare the two cases. The climate change scenario has a confidence interval, within which the distribution can vary depending on the events removed. In the second scenario, instead of changing the number of events falling within the wettest class, the associated scale parameter of gamma distribution is changed. The scaling parameter of the wettest class is simply increased by 15%. The last scenario considers both the change in the number of events and increase in the scale parameter considered in the two previous scenarios.

4. RESULTS

This chapter brings together the main results of the analysis carried out in Chapter 3.

4.1 Hydrological Variables

One of the first aspects highlighted during the extraction of peak events is precisely the distribution of flow events over the years. Specifically, almost all the Piave River basins present the greatest number of events in the month of May. This is different for Astico and Fiorentina stream, which have two distinct peaks, one in May like the previous basins and one in November. Both months are transitional periods. May is a month that is often characterized by heavy spring rainfall due to the arrival of Atlantic disturbances and spring warming that can promote the development of thunderstorms and snowmelt. November is a month of autumn seasonality, Atlantic disturbances return frequently, and rainfall is often very heavy. In particular, the soil may already be saturated with water from previous autumn rains, reducing the absorption capacity and increasing flood events. Figure 38 shows several compound graphs, one for each basin. They consist of a histogram (represented in blue), which shows the frequency of events associated with each month, and a line of red, which depicts the number of events for each month.

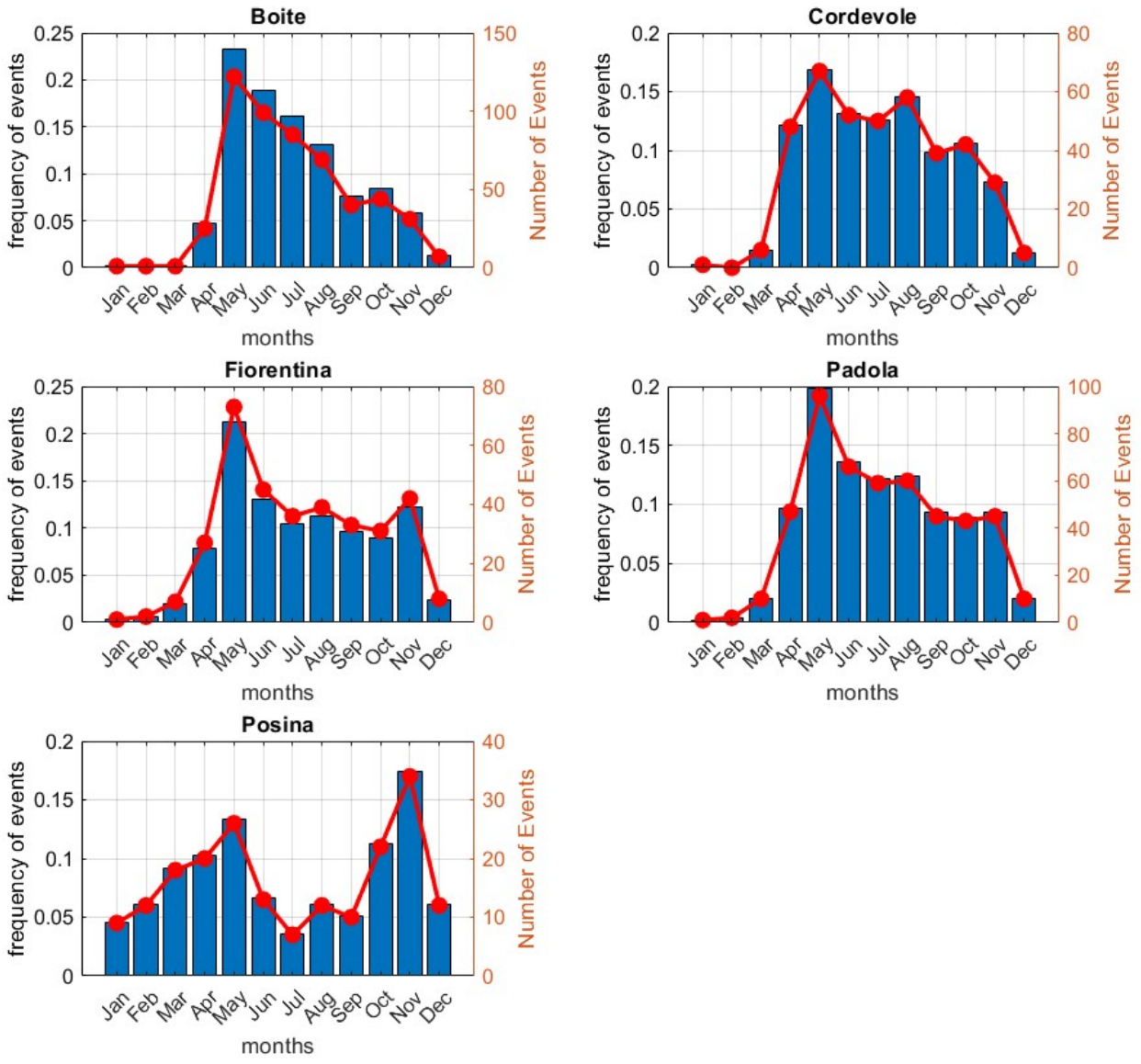


Figure 33 - Frequency distribution of flood events considering Seasonality.

Another interesting thing to observe is in graph 34 that represents the variation of evapotranspiration during the year, mediated over the years considered for each basin. In particular the Posina graph represents through thin lines, the evapotranspiration measured by each rainfall station considered. In bold is depicted the line that interpolates the evapotranspiration values measured by the various stations.

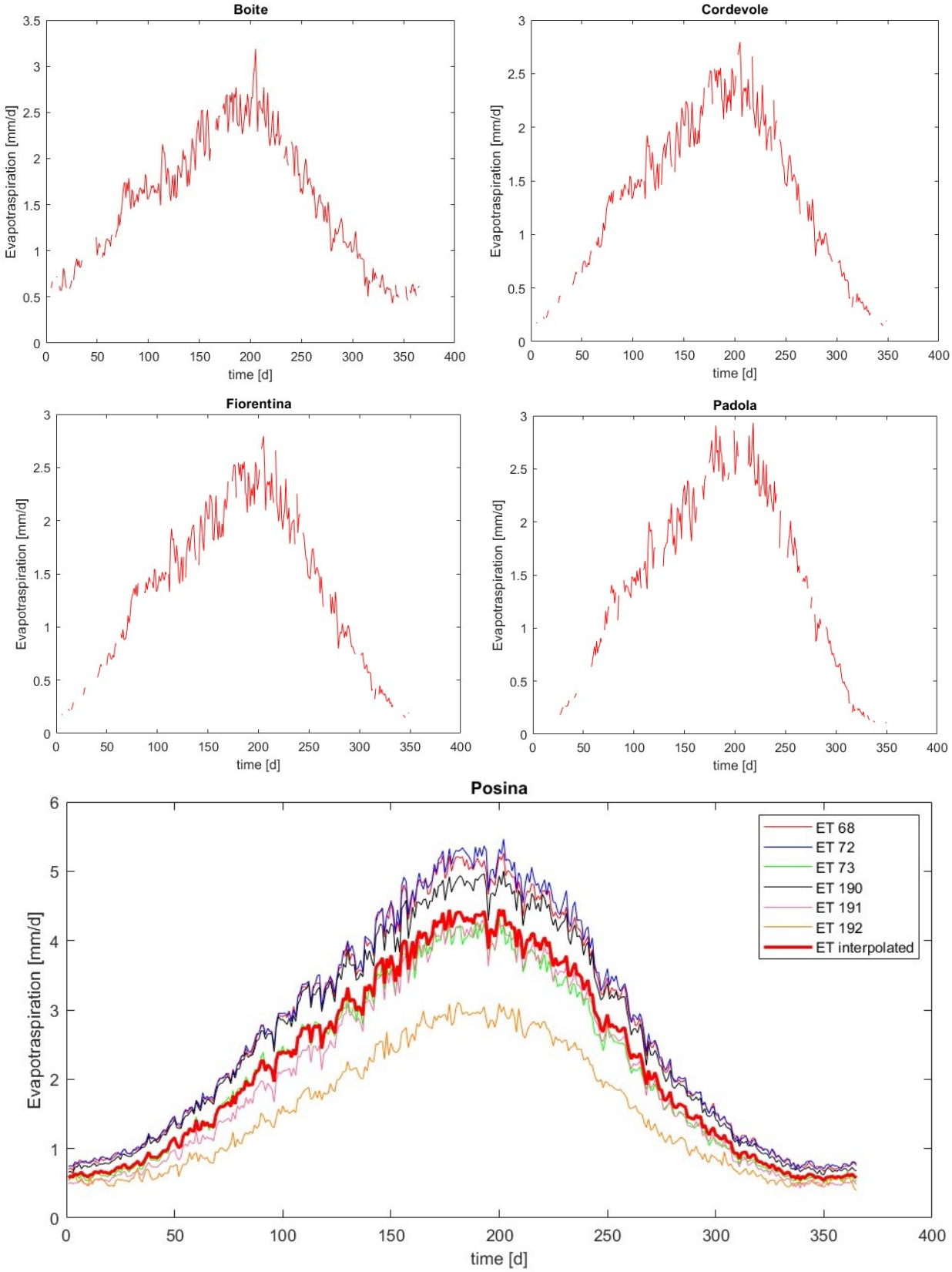


Figure 34 - Average evapotranspiration of various basins.

Table 4 shows the main parameters needed to calculate evapotranspiration (ET_0), in this case the count refers to the Fiorentina stream basin. Tables of all basins have not been included, as they are not very significant for the analysis in this thesis.

	Units of measure	Winter	Spring	Summer	Autumn	Annual
T_{min}	°C	-4.0	5.0	14.2	6.3	5.5
T_{max}	°C	6.7	18.1	28.1	17.6	17.7
T	°C	1.4	11.5	21.2	12.0	11.6
T	K	274.7	284.8	294.5	285.3	284.9
$u_{z\,(5m)}$	m/s	0.3	0.6	0.5	0.3	0.4
$u_{z\,(2m)}$	m/s	0.2	0.4	0.4	0.2	0.3
$e_s^*(T_{min})$	kPa	0.5	0.8	1.4	0.9	0.9
$e_s^*(T_{max})$	kPa	0.9	1.8	3.1	1.7	1.7
e_s^*	kPa	0.7	1.3	2.2	1.3	1.3
RH_{min}	%	23.0	18.0	27.7	27.3	24
RH_{max}	%	98.7	99.0	99.0	99.0	99
e_a	kPa	0.3	0.6	1.1	0.7	0.6
VPD	kPa	0.4	0.7	1.1	0.6	0.7
ΔT_{min}	kPa/°C	0.0	0.0	0.1	0.0	0.05
ΔT_{max}	kPa/°C	0.0	0.1	0.1	0.1	0.10
Δ	kPa/°C	0.0	0.1	0.1	0.1	0.07
R_s	MJ/m2/d	5.5	15.4	19.4	9.0	12.0
R_a	MJ/m2/d	9.3	23.6	29.0	14.8	19.0
$R_{s,0}$	MJ/m2/d	6.9	17.7	21.8	11.1	14.3
R_{nl}	MJ/m2/d	0.0	0.0	0.0	0.0	0.0
R_n	MJ/m2/d	4.2	11.8	15.0	6.9	9.2
ET_0	mm/d	0.8	2.7	4.0	1.6	2.1
ET_0	cm/d	0.1	0.3	0.4	0.2	0.214

Table 4 - main parameters required for the calculation of evapotranspiration. Example referring to the Fiorentina stream basin.

4.2 Joint distribution of flow peaks and antecedent soil moisture

Table 5 shows various linear correlation coefficients, calculated both by considering only cumulative precipitation and also including evapotranspiration in the joint distribution of flow peaks and antecedent soil moisture. The coefficients are higher or equal when evapotranspiration is also considered. This indicates that the linear regression line is better able to describe the trend of the points and therefore there is a greater correlation between the two variables. It can also be observed that the correlation coefficients while having an improvement considering evapotranspiration, remain rather low value, this may be due to the presence of snow, which goes to alter the hydrological response of the basin. The case with the highest correlation coefficient is when the cumulative soil moisture is considered, referring to the 5 days before the peak events. Therefore, for subsequent calculations, reference will always be made to the 5-day cumulative soil moisture.

Correlation coefficients													
		cumulative days											
		30		25		20		15		10		5	
		P	P - ET	P	P - ET	P	P - ET	P	P - ET	P	P - ET	P	P - ET
Basins	Boite	0.17	0.17	0.18	0.19	0.20	0.21	0.27	0.26	0.31	0.31	0.42	0.43
	Cordevole	0.37	0.41	0.39	0.43	0.41	0.45	0.46	0.49	0.49	0.52	0.56	0.56
	Fiorentina	0.39	0.40	0.40	0.43	0.43	0.45	0.48	0.50	0.50	0.54	0.56	0.55
	Padola	0.29	0.34	0.31	0.37	0.34	0.39	0.4	0.44	0.45	0.48	0.50	0.52
	Posina	0.44	0.48	0.49	0.52	0.54	0.56	0.56	0.58	0.57	0.59	0.65	0.66

Table 5 - Correlation coefficients of the various ranges of cumulative soil moisture.

Graphs 35-39 show the scatter points representing the marginal distributions between peak flow and antecedent moisture. Different colors indicate the joint distributions for different cumulative periods: 30, 25, 20, 15, 10 and 5 days. The respective linear regression line is also plotted for each distribution.

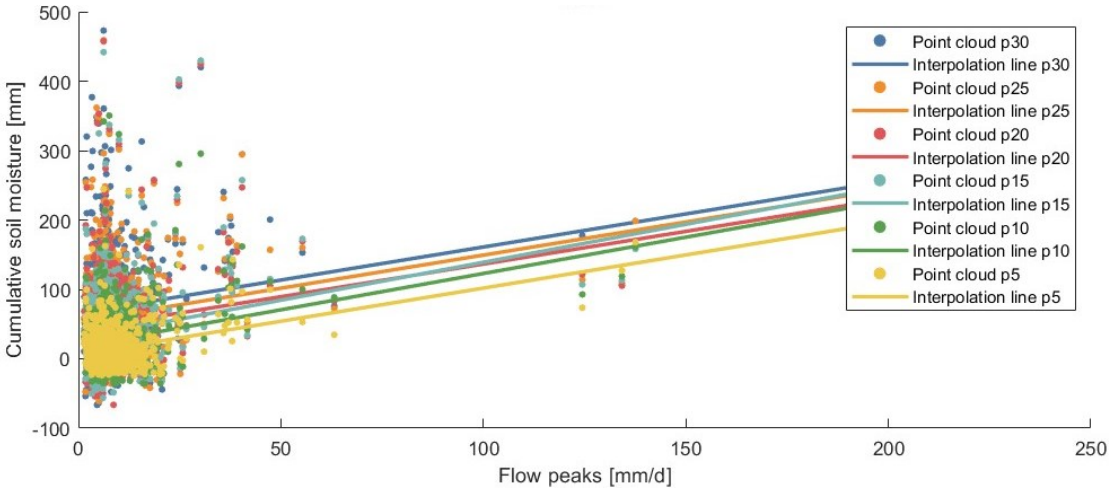


Figure 35 - Point cloud and interpolation lines referring to different cumulative soil moisture ranges. Referred to the Boite stream basin.

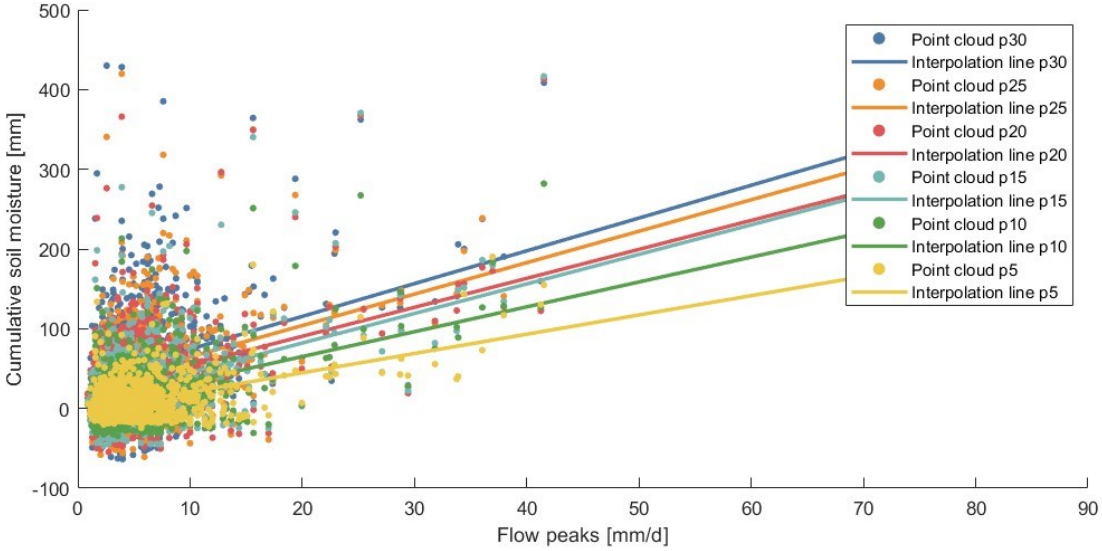


Figure 36 - Point cloud and interpolation lines referring to different cumulative soil moisture ranges. Referred to the Cordevole stream basin.

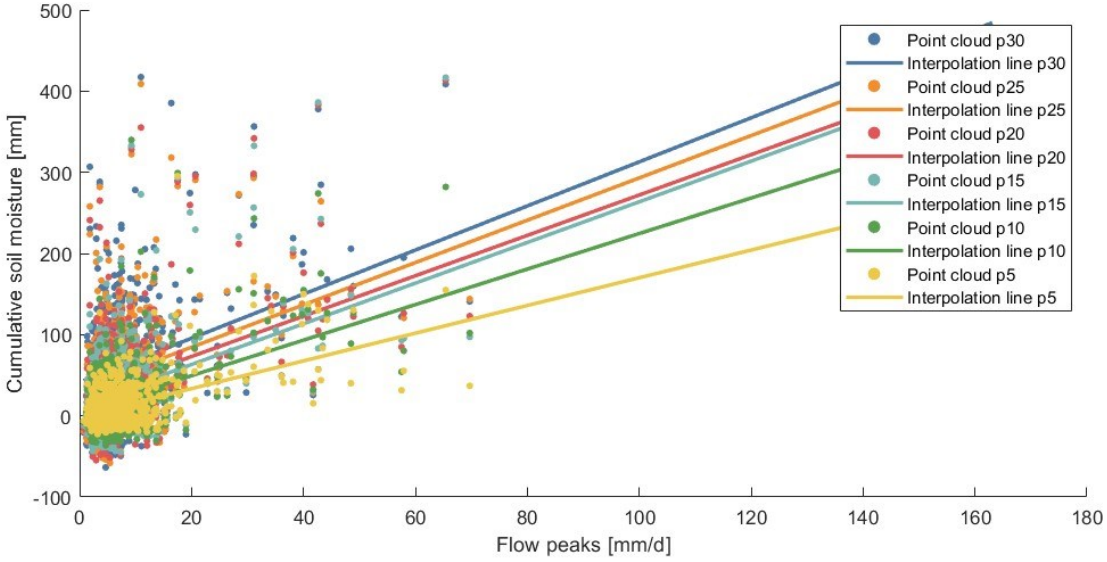


Figure 37 - Point cloud and interpolation lines referring to different cumulative soil moisture ranges. Referred to the Fiorentina stream basin.

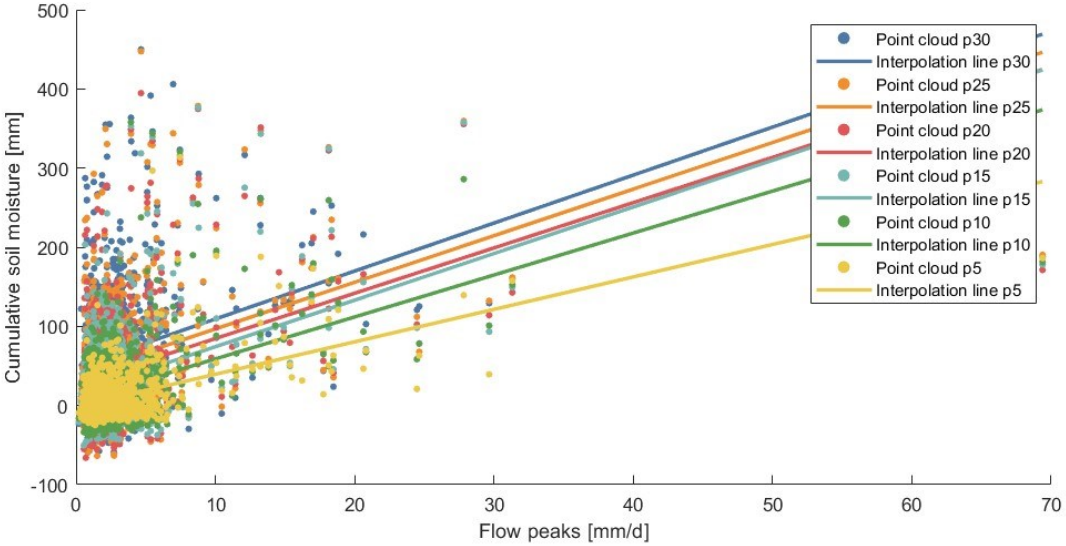


Figure 38 - Point cloud and interpolation lines referring to different cumulative soil moisture ranges. Referred to the Padola stream basin.

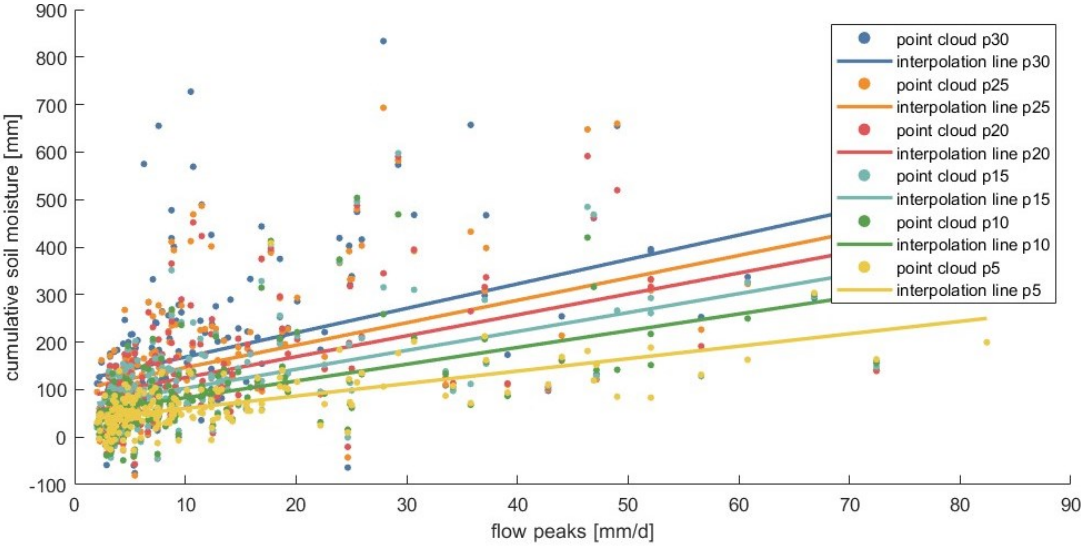


Figure 39 - Point cloud and interpolation lines referring to different cumulative soil moisture ranges. Referred to the Astico stream basin.

Figures 40-44 shows the complete joint distribution between flow peaks and cumulative soil moisture of the 5 days preceding events. The two data sets have been divided into classes, so that the graph can be visualized more clearly. The frequency of each class is shown with a color scale (logarithmic scale, to emphasize the difference in color between cells). As can be seen, the Astico stream basin has numerous extreme events characterized by high intensity, compared to the other basins. The basins of the other streams generally present one extreme event that differs greatly from the others; that event is associated with storm Vaia. The Astico basin does not present a high deviation between Vaia and the other events, probably due to the fact that the storm was less intense than in the sites belonging to the Piave basin.

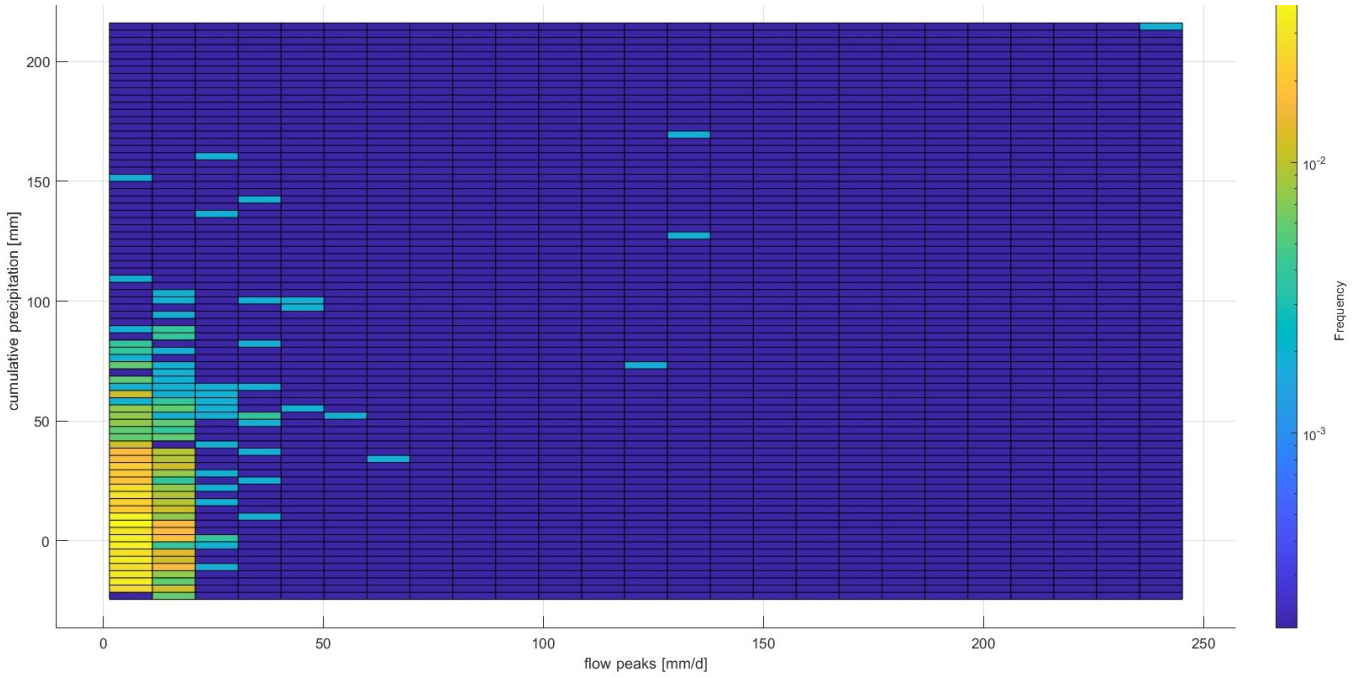


Figure 40 - Joint distribution of cumulative soil moisture present 5 days before flood events and peak flood events. Referring to the Boite stream basin.

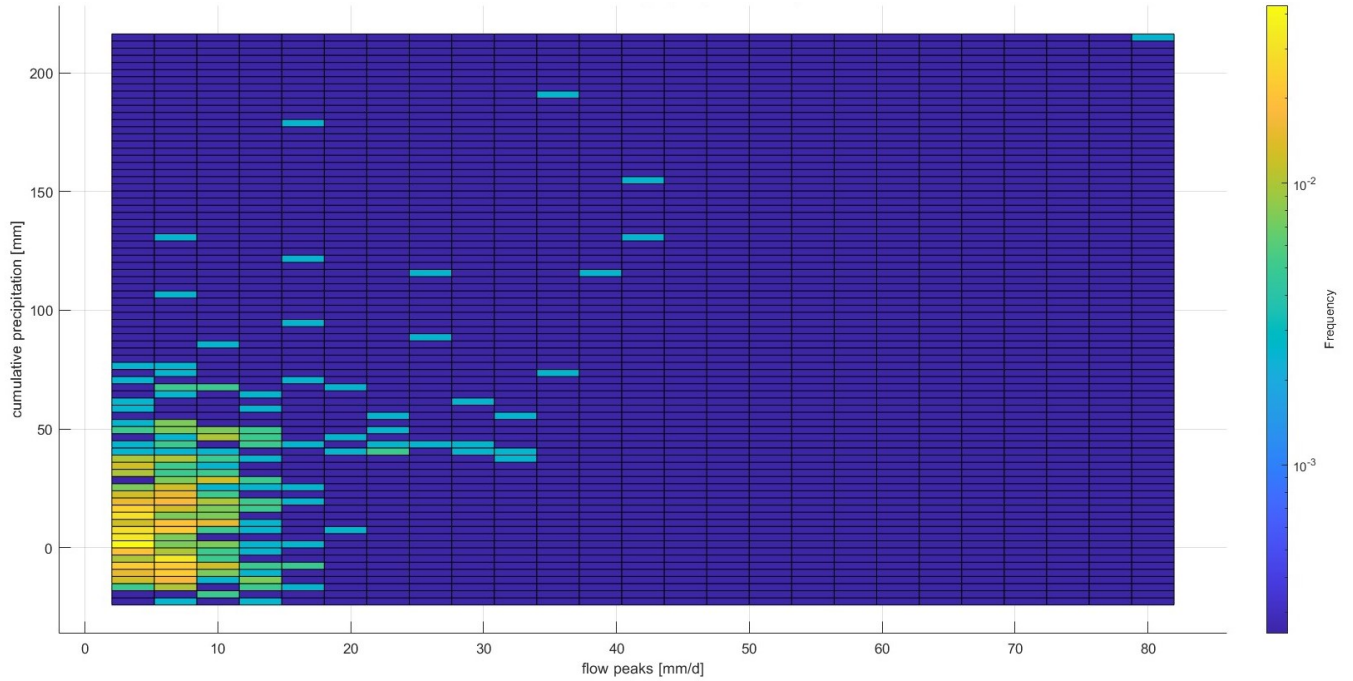


Figure 41 - Joint distribution of cumulative soil moisture present 5 days before flood events and peak flood events. Referring to the Cordevole stream basin.

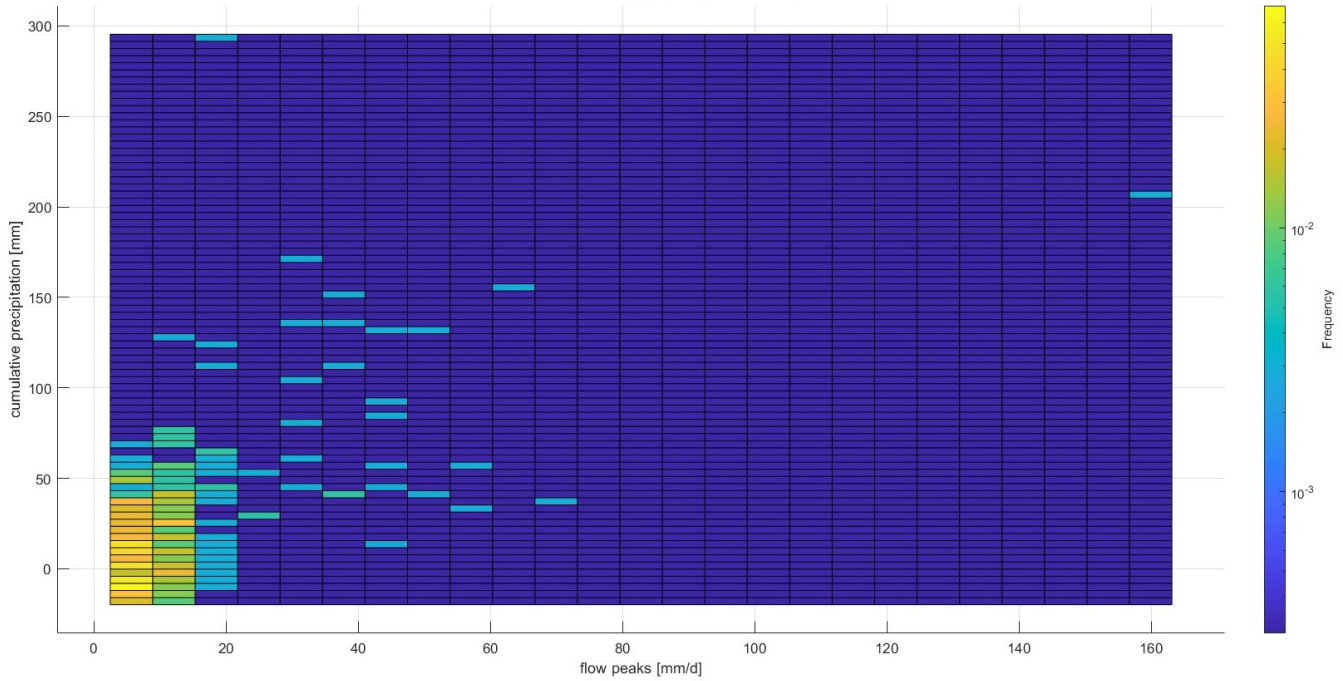


Figure 42 - Joint distribution of cumulative soil moisture present 5 days before flood events and peak flood events. Referring to the Fiorentina stream basin.

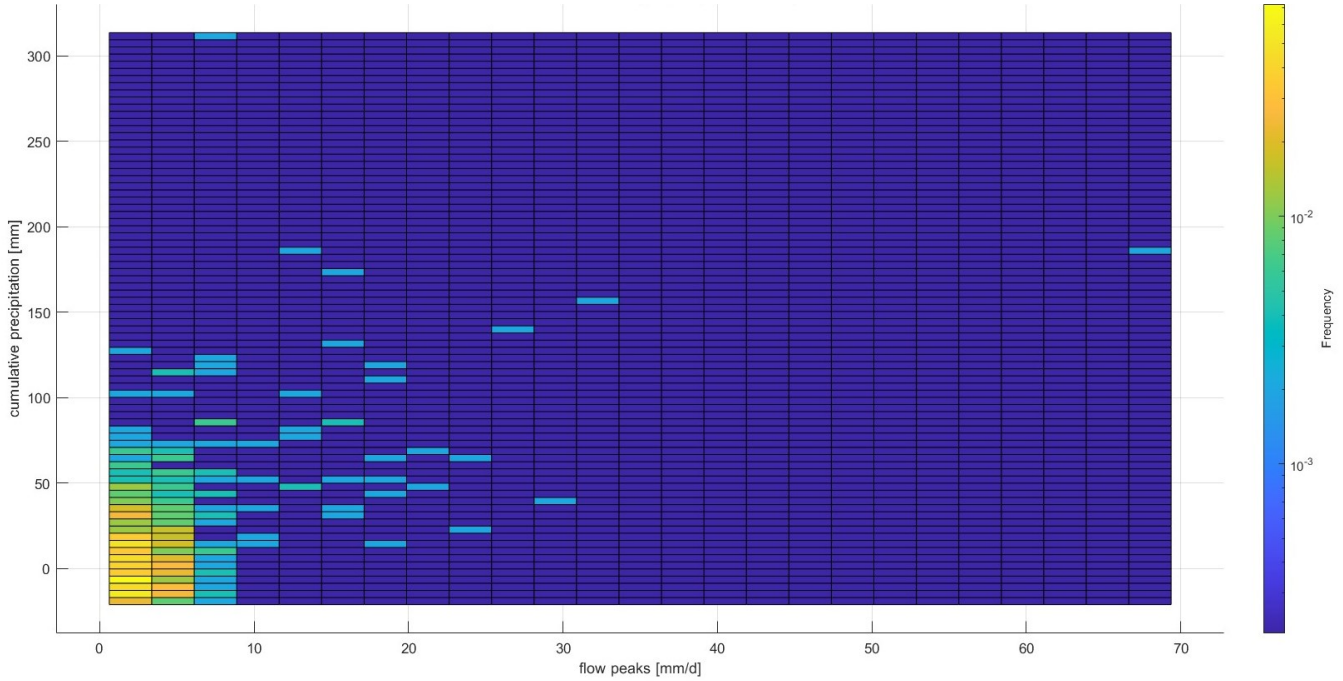


Figure 43 - Joint distribution of cumulative soil moisture present 5 days before flood events and peak flood events. Referring to the Padola stream basin.

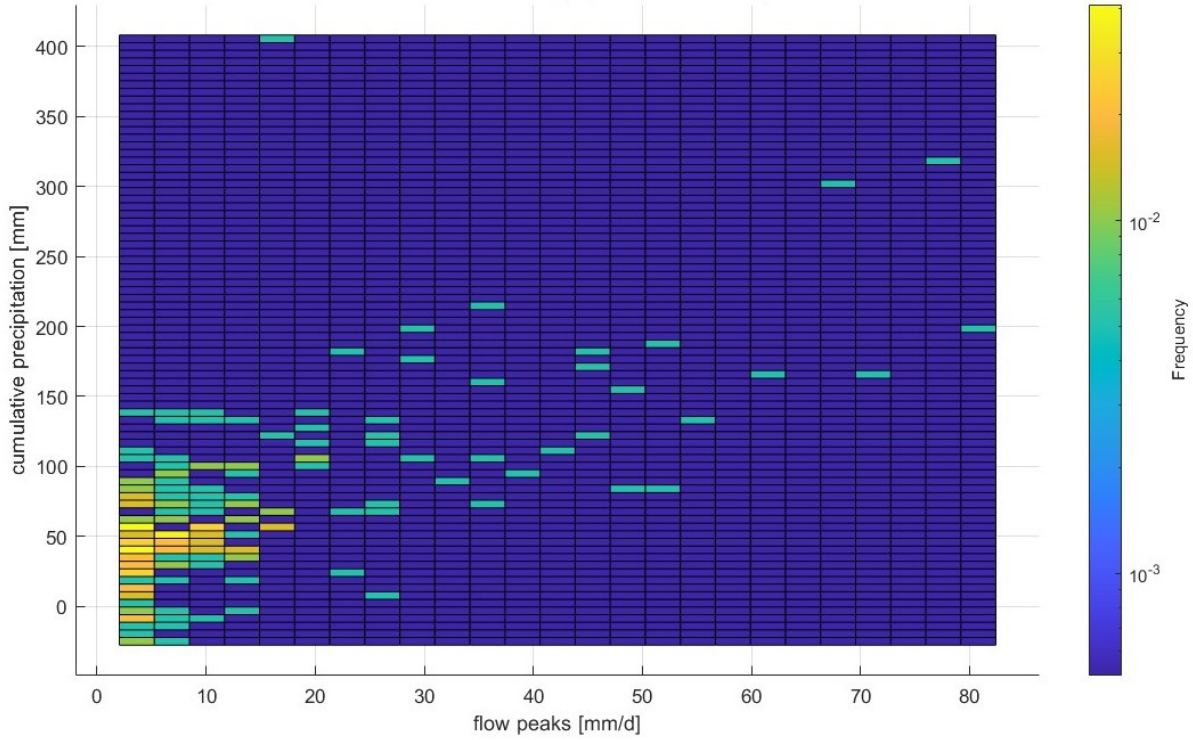


Figure 44 - Joint distribution of cumulative soil moisture present 5 days before flood events and peak flood events. Referring to the Astico stream basin.

4.3 Goodness of fit

Analyzing the qq-plots of the various methods, a significantly greater goodness of fit is found in the case of the GEV distribution, in all the basins considered. It can also be seen that the moisture accounting SMEVD fits the data much better than the SMEVD applied to a single class of data. This is because by increasing the number of params, the distributions can better describe the data. It is equally true that the drop in performance for moisture accounting SMEVD is mainly related to the most extreme event corresponding to storm Vaia. The model is not able to fit it correctly, given its great intensity compared to the other events. This thing can be appreciated by looking at coefficients of determination (see table 6). The closer the coefficient is to one, the greater the fit between the model and the observed data.

	R^2		
basins	**SMEVD single class**	**Moisture accounting SMEVD**	**GEV**
Boite	0.16	0.69	0.86
Cordevole	0.37	0.82	0.92
Fiorentina	0.35	0.70	0.91
Padola	0.30	0.72	0.91
Posina	0.61	0.74	0.93

Table 6 - R^2 coefficient of determination of qq-plots for various models and basins. Calculated using the Gamma distribution

Looking at graphs 45-49 of the qq-plots, one can obtain more information on the fit of models to data. Specifically, one can see that in the graphs referring to SMEV distributions, for equal probabilities, the theoretical quantile distributions underestimate quantiles for high values. While for low values, theoretical quantiles overestimate observed quantiles. The qq-plot referring to the GEV distribution fits the data better. The quantile most extreme and isolated from the others is related to the Vaia storm. The best fit between samples and distribution is for the Fiorentina and Padola basins. Whereas for the Boite basin the fit is worse, probably due to the high extreme events it is subjected to. The graph referring to the Astico stream, on the other hand, only has 14 years of data available, which is why the quantiles represented have a lower number.

It was noted that by setting a threshold to predominantly extracted peak flow events, e.g. 5, 10 or 15 mm/d, there is a progressive improvement in all models in the fit between the quantiles and the 45° straight line. By applying these thresholds, the number of events decreases dramatically, for example by setting the threshold equal to 10 mm/d for the Fiorentina stream, events decrease from 344 to 118. For this reason, it was decided not to raise the thresholds, so as not to affect the subsequent cross validation phase.

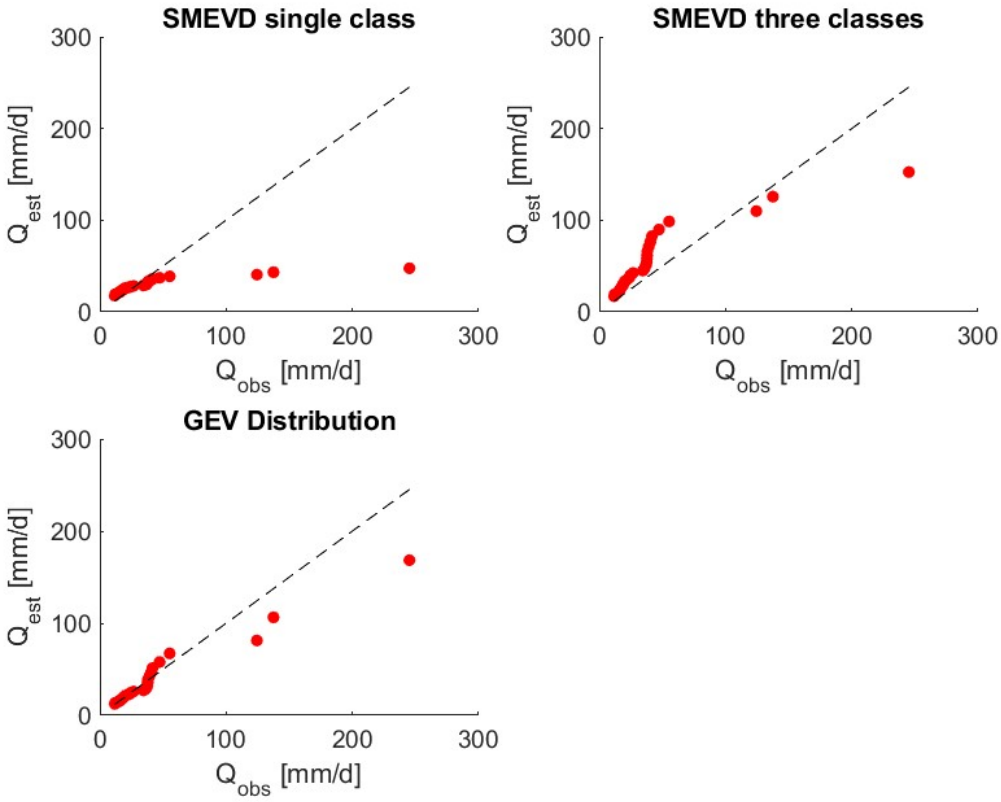


Figure 45 - qq-plot of the three models, referring to the Boite stream.

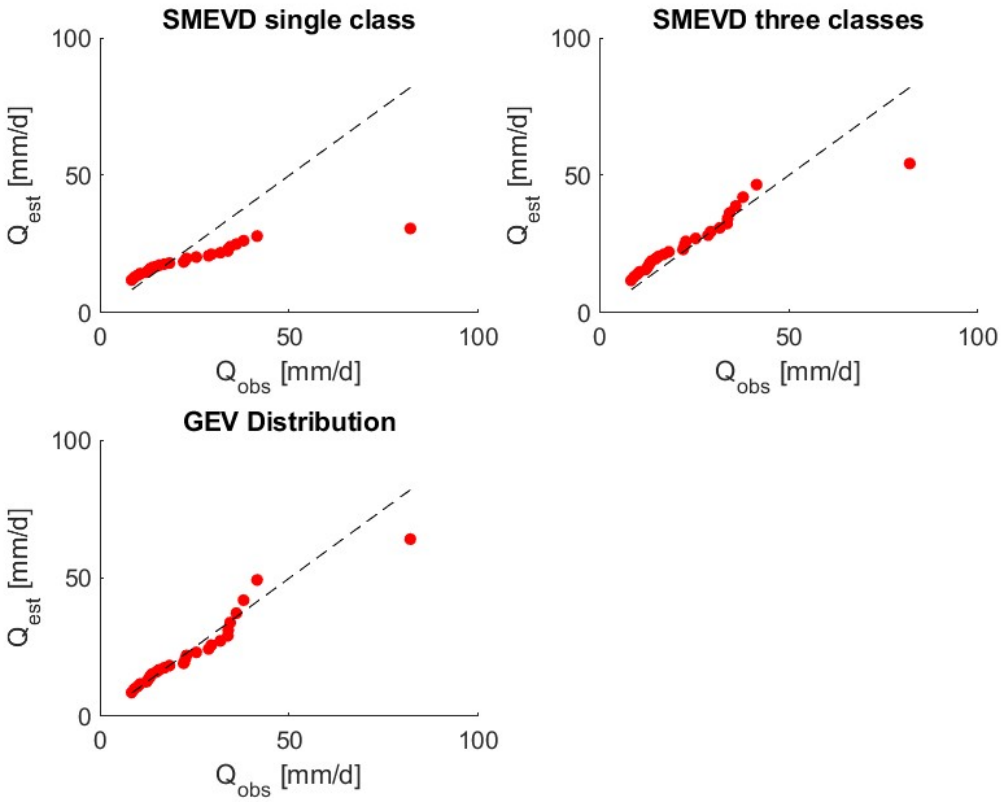


Figure 46 - qq-plot of the three models, referring to the Cordevole stream.

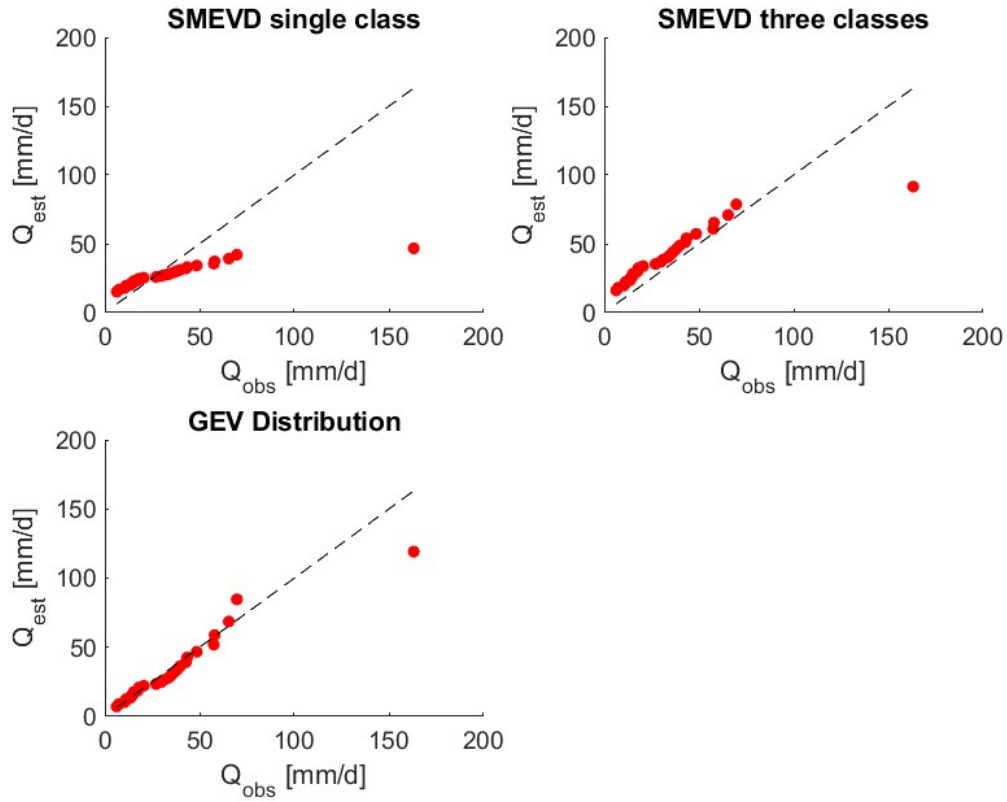


Figure 47 - qq-plot of the three models, referring to the Fiorentina stream.

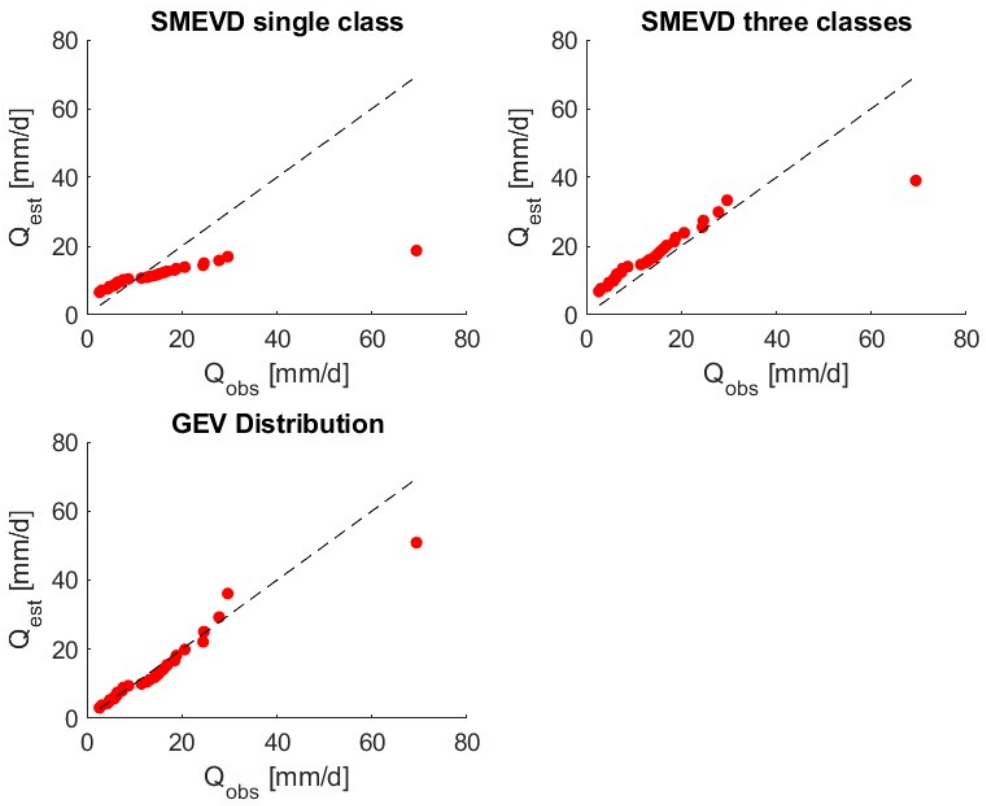


Figure 48 - qq-plot of the three models, referring to the Padola stream.

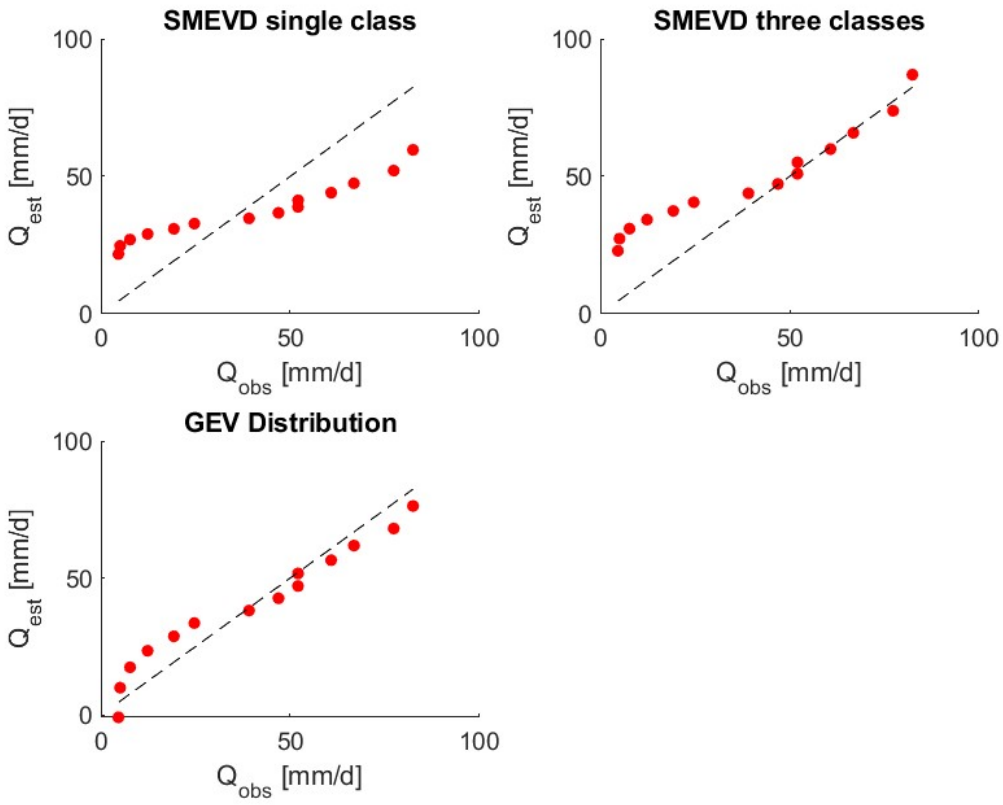


Figure 49 - qq-plot of the three models, referring to the Astico stream.

4.4 CDF of the models compared

In graphs 50-54, the cumulative density functions (CDF) of the various models are compared. Again, looking at the graphs, the SMEVD applied to a single class underestimates events with higher intensity, compared to the GEV, which describes them more accurately. The moisture accounting SMEVD, on the other hand, succeeds in describing the course of events better, coming closer to the GEV distribution. The comments on the various stream basins are the same as those made in Chapter 4.3 for goodness of fit.

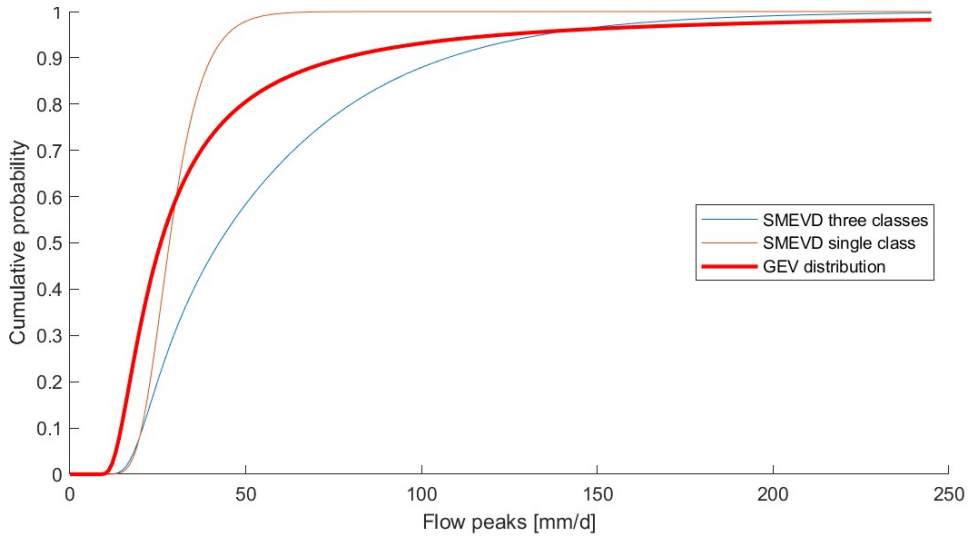


Figure 50 - CDF comparison of the three models. Referred to the Boite stream.

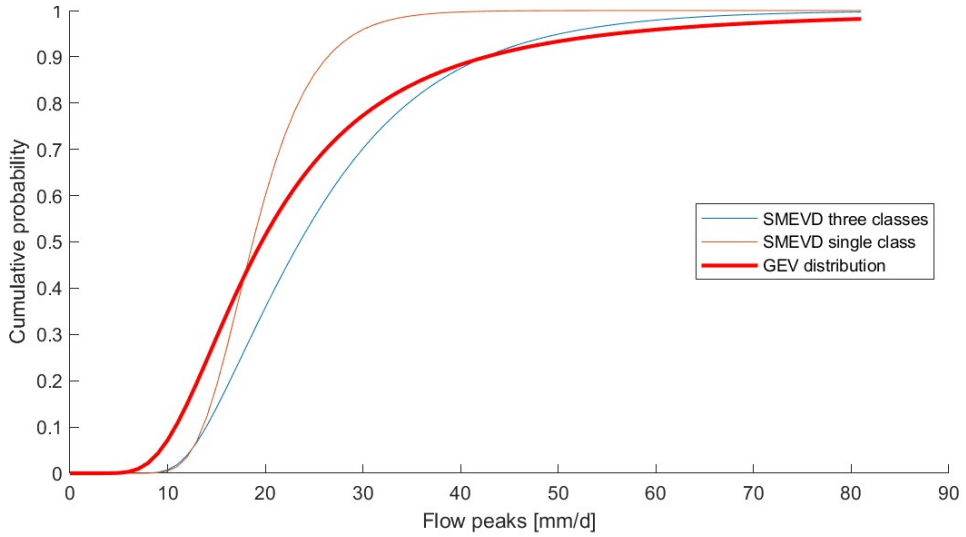


Figure 51 - CDF comparison of the three models. Referred to the Cordevole stream.

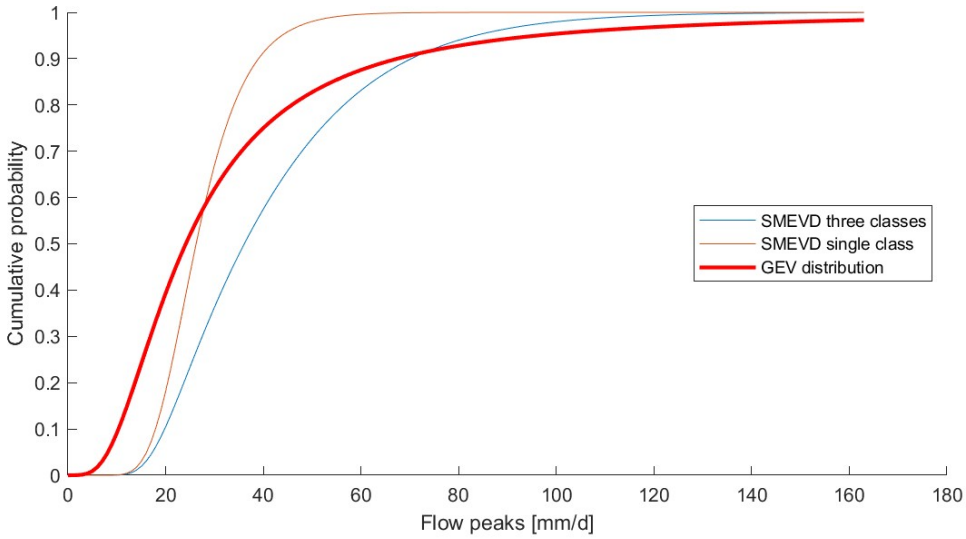


Figure 52 - CDF comparison of the three models. Referred to the Fiorentina stream.

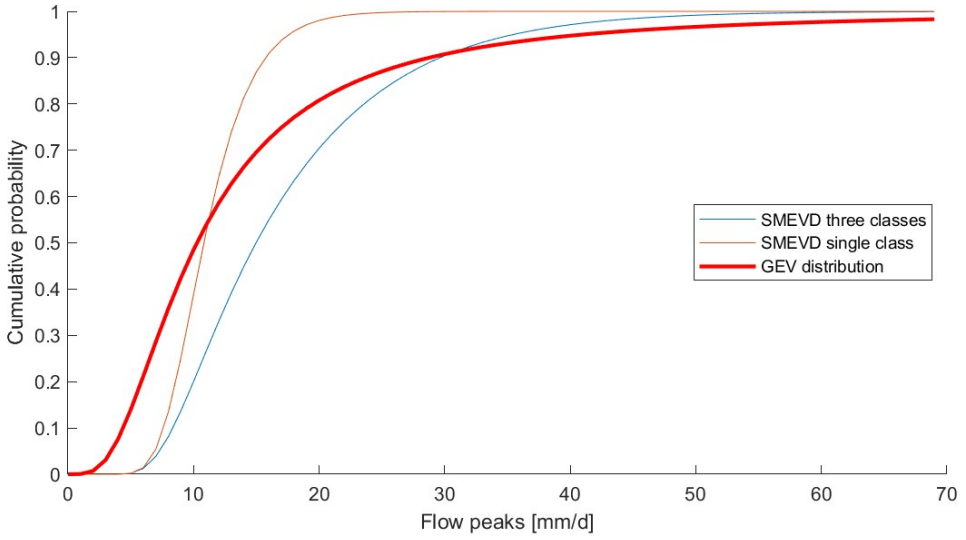


Figure 53 - CDF comparison of the three models. Referred to the Padola stream.

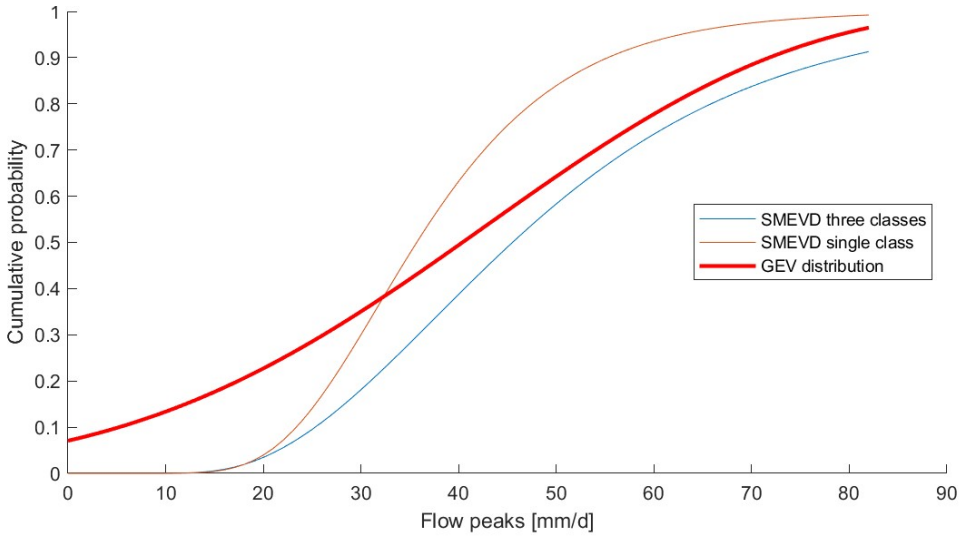


Figure 54 - CDF comparison of the three models. Referred to the Astico stream.

4.5 Cross Validation

Two types of graphs were chosen to analyze the results of Cross Validation, one using boxplots on relative errors and the other analyzing the distributions of relative errors through the violin-plot. In the first case, boxplots are produced for each return time considered for the validation phase. In this case, in figures 55-69 the scenario that considers 10 years was chosen to represent the model calibration phase. As can be seen, the models single-class SMEVD and moisture accounting SMEVD overestimate events for small return times. In contrast, for large return times, they slightly underestimate events. The GEV distribution, on the other hand, provides a more accurate calculation for events with low return times, but large variance for high return times. If, on the other hand, we look at the boxplots in Appendix chapter, obtained using the most extreme scenario of 3-year calibration, the graphs referring to single-class SMEVD and GEV distribution have numerous outliers (indicated by red crosses). In contrast, moisture accounting SMEVD has far fewer outliers, a sign that the model performs a more robust estimation of outliers.

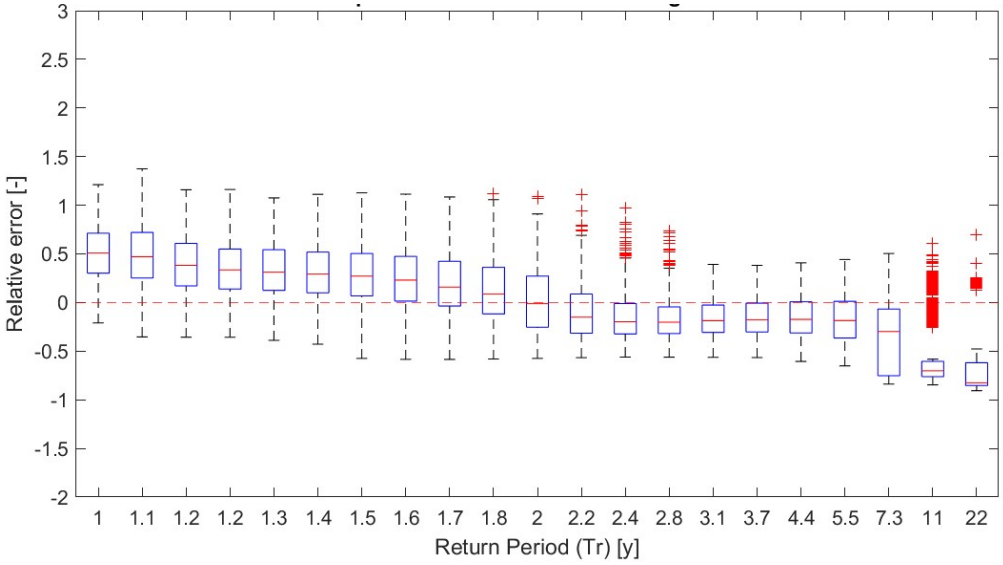


Figure 55 - Boxplot relative errors SMEVD single class. Referring to the Boite stream basin.

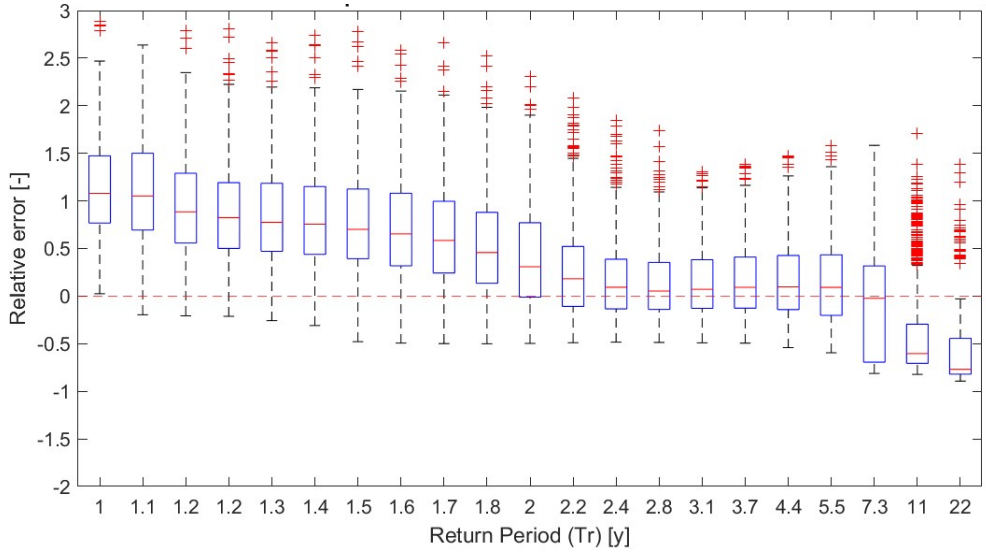


Figure 56 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Boite stream basin.

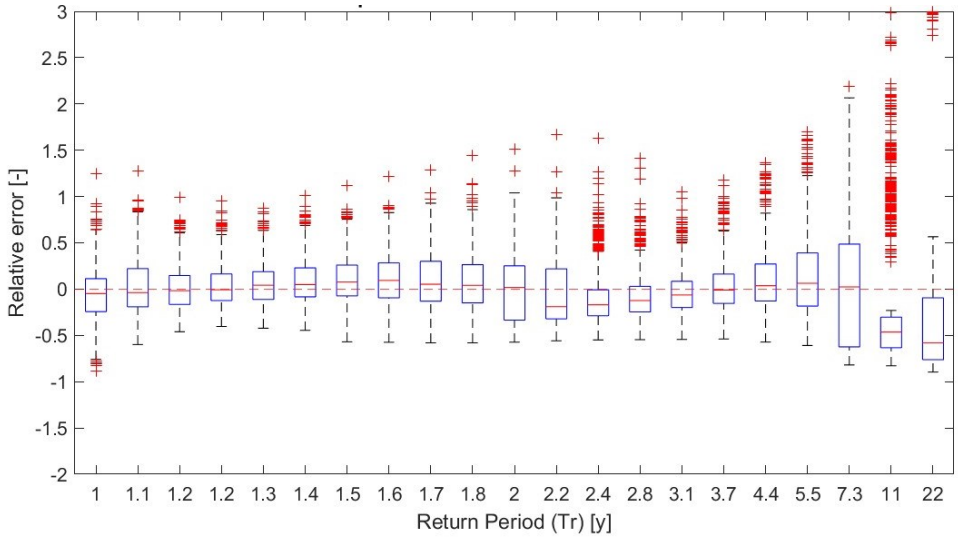


Figure 57 - Boxplot relative errors GEV distribution. Referring to the Boite stream basin.

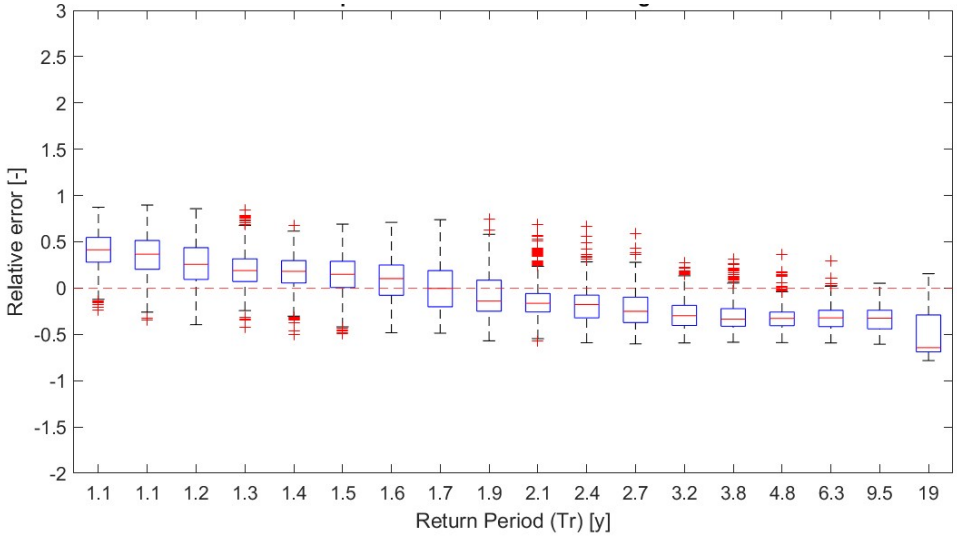


Figure 58 - Boxplot relative errors SMEVD single class. Referring to the Cordevole stream basin.

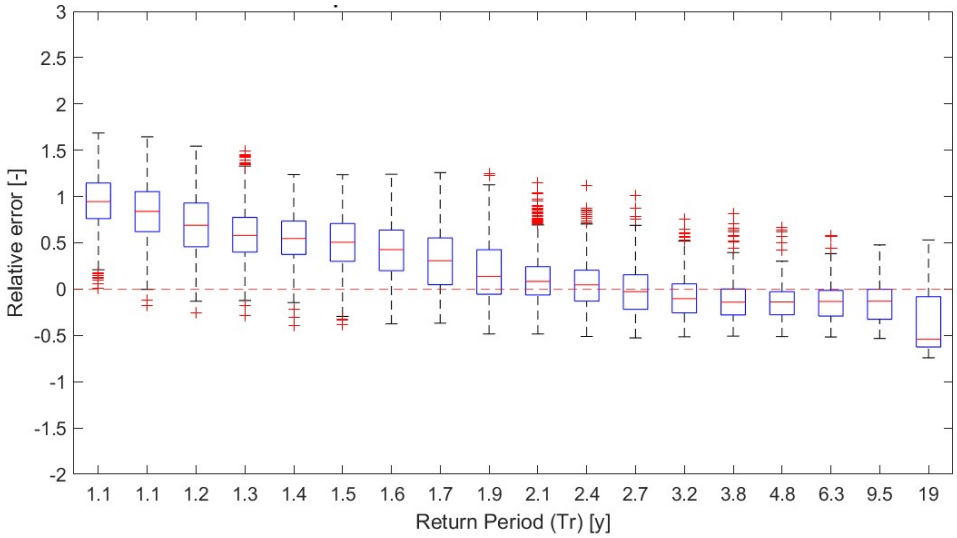


Figure 59 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Cordevole stream basin.

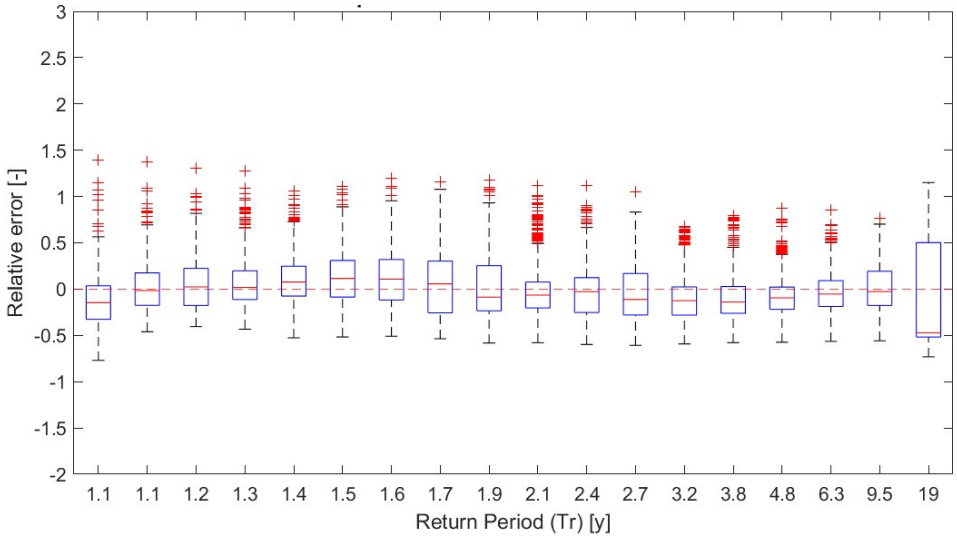


Figure 60 - Boxplot relative errors GEV distribution. Referring to the Cordevole stream basin.

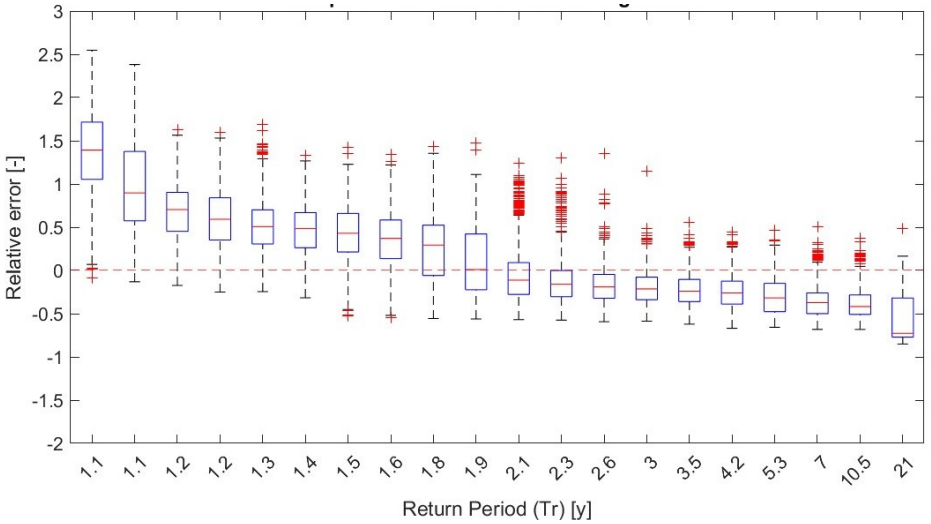


Figure 61 - Boxplot relative errors SMEVD single class. Referring to the Fiorentina stream basin.

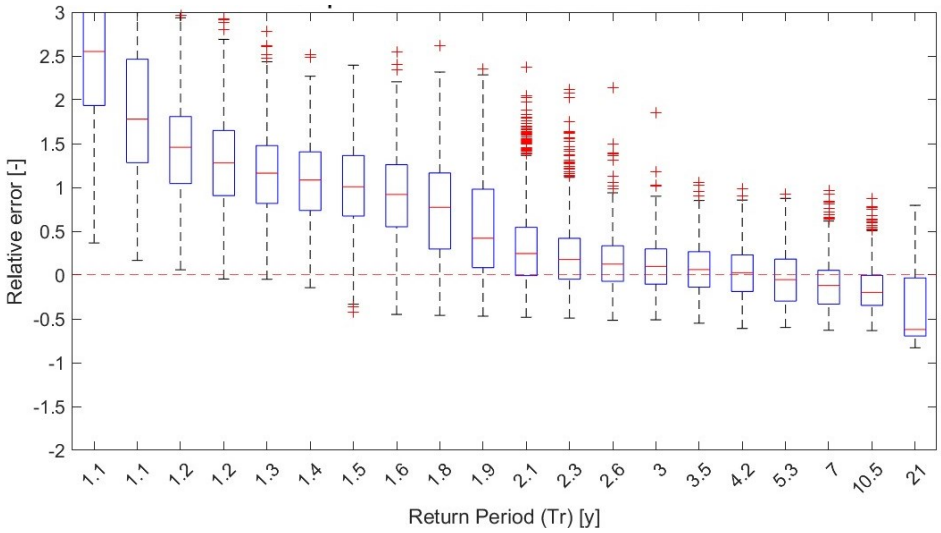


Figure 62 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Fiorentina stream basin.

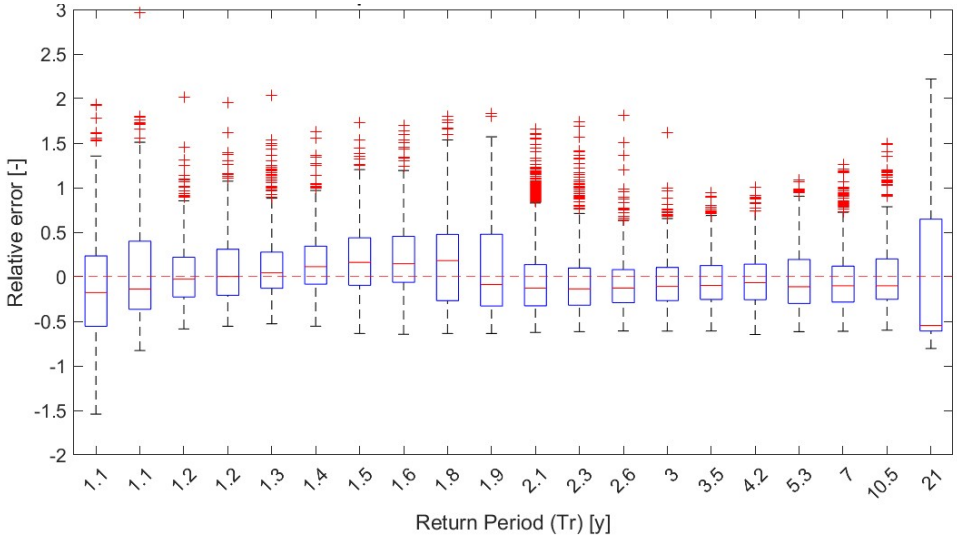


Figure 63 - Boxplot relative errors GEV distribution. Referring to the Fiorentina stream basin.

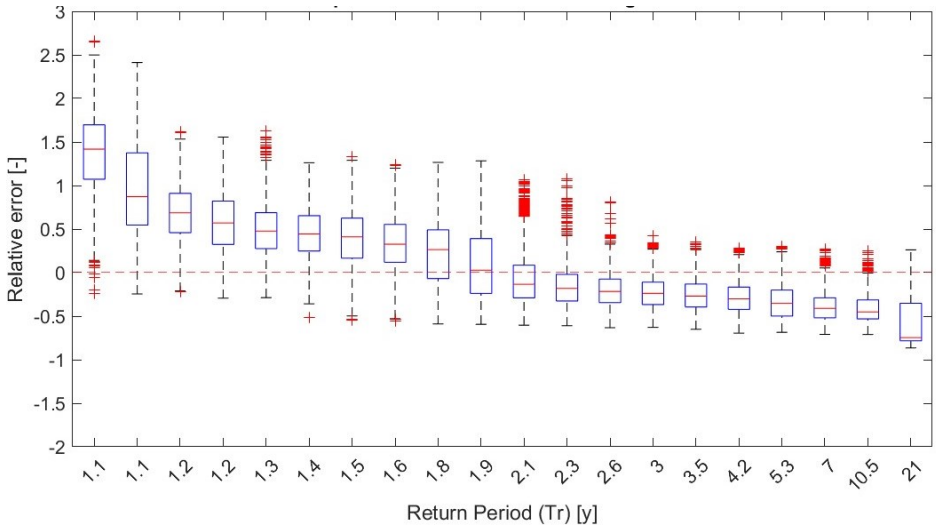


Figure 64 - Boxplot relative errors SMEVD single class. Referring to the Padola stream basin.

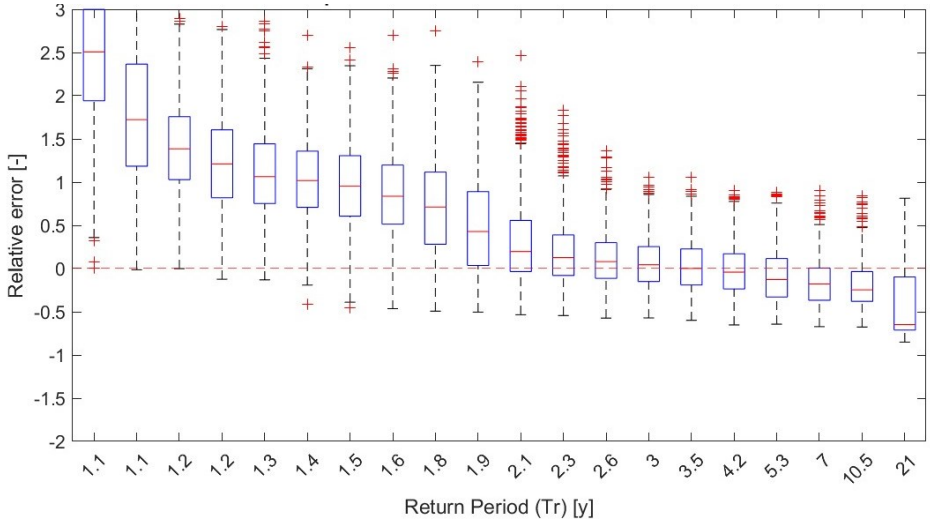


Figure 65 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Padola stream basin.

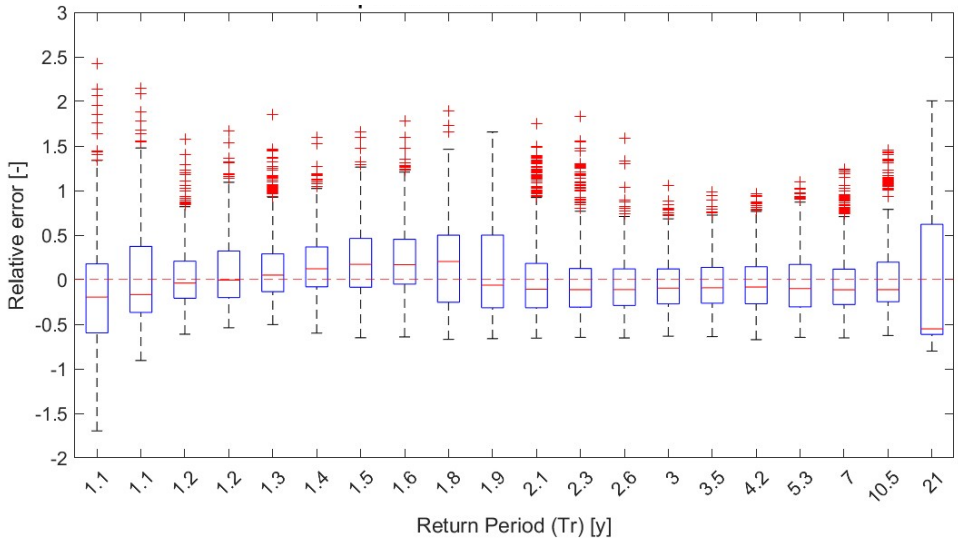


Figure 66 - Boxplot relative errors GEV distribution. Referring to the Padola stream basin.

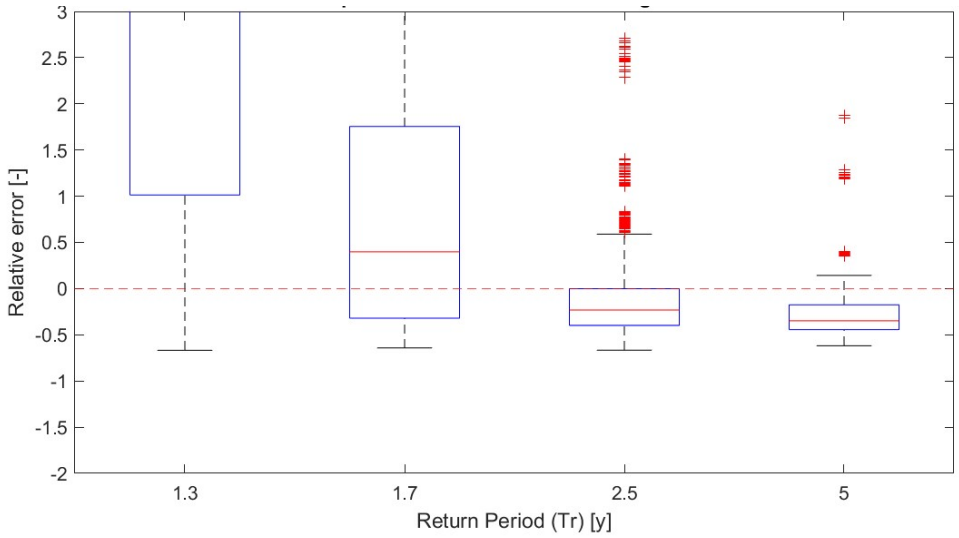


Figure 67 - Boxplot relative errors SMEVD single class. Referring to the Astico stream basin.

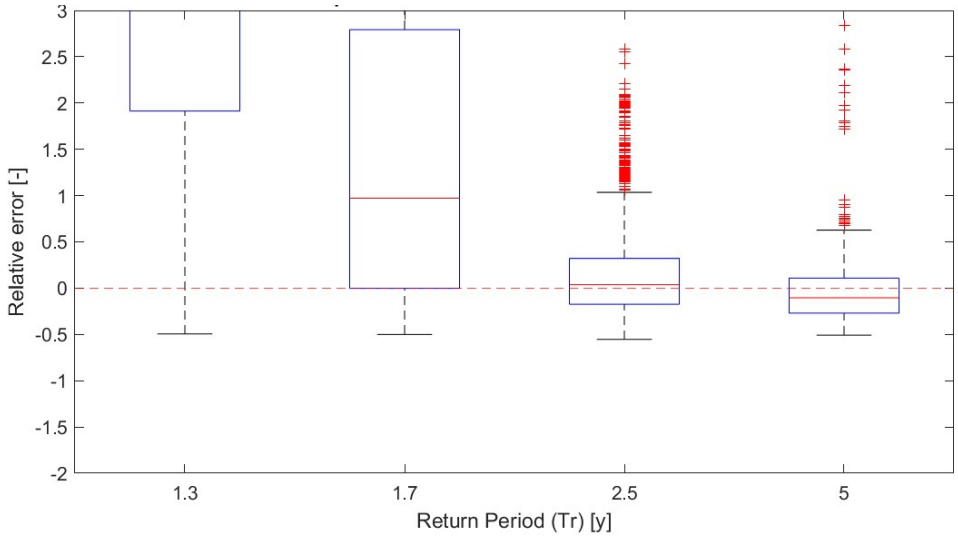


Figure 68 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Astico stream basin.

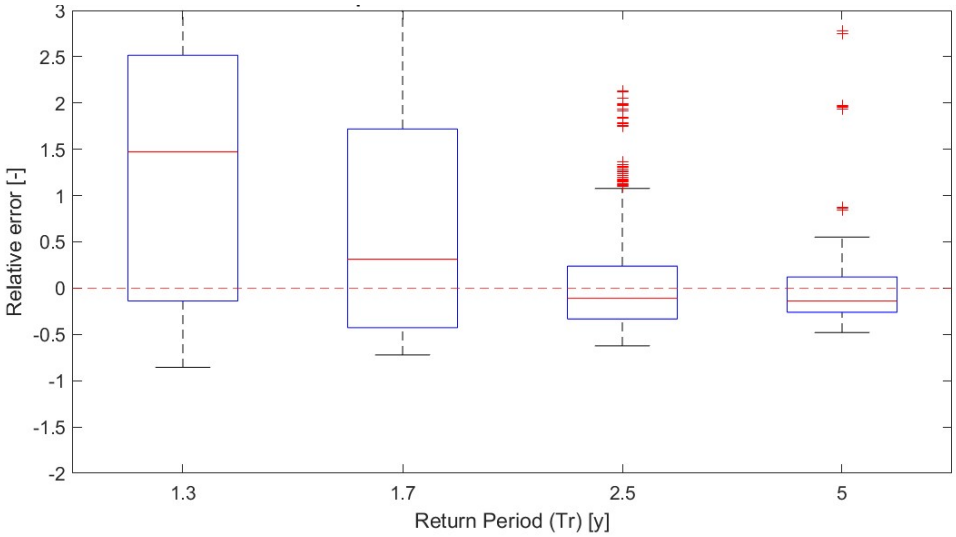


Figure 69 - Boxplot relative errors GEV distribution. Referring to the Astico stream basin.

In the second case, the relative error distributions of the various models were analyzed using violin-plots. Two graphs were produced for each basin, the first considering relative errors associated with the last return time, while the second considers relative errors associated with the last 4 return times. The final return time changes depending on the stream being considered (as the available historical data are not the same for all basins) and the years used for the calibration phase. For example, if we consider 10 years of calibration for the analyzed sites, then the return times will be as follows: Boite stream 22 years, Cordevole stream 19 years, Fiorentina stream 20 years, Padola stream 25 years and Astico stream 12 years. The graphs 70-74, referring to the last return time, allow the three models to be compared by considering 3, 5 and 10 years of events for the calibration phase. The models associated with 10 years of events for the calibration phase allow to observe bimodal distributions. The peaks characterized by high frequencies occur when storm Vaia is considered for the validation phase, indicating that models are unable to predict that event. The other peaks, on the other hand, are characterized by a lower probability and occur when the event associated with Vaia is not considered in the validation phase. Looking at these peaks, the model with the best predictive ability is the one using moisture accounting SMEVD. A special comment can be made about the basins of Boite and Astico streams. Looking at graphs 70 and 75 of the Boite torrent, no model fits the data correctly. This could be due to the unpredictable behavior of the Boite stream, which has many extreme events of great magnitude. Even looking at the graph referring to the last return time, no model seems able to predict event associated with storm Vaia. The graphs referring to the Astico stream, on the other hand, show that model using the GEV distribution manages to predict events better than models using the SMEVD. This may be due to the limited availability of data (14 years compared to 31 years for the other basins). The scenarios that consider 3 or 5 years for calibration, on the other hand, perform worse than the case analyzed with 10 years; this is because the distributions fit the data more imprecisely, given the low availability of events for calibration. However, these graphs allow to observe that the dispersion of relative errors is greater in the model using the GEV distribution. The graphs 75-79 show the same information but take the last four return times into consideration. In this case the moisture accounting SMEVD and the GEV distribution referring to 10 years of calibration, has a very similar error distribution.

By reducing the number of years of data for calibration, the GEV distribution also has a greater dispersion of errors.

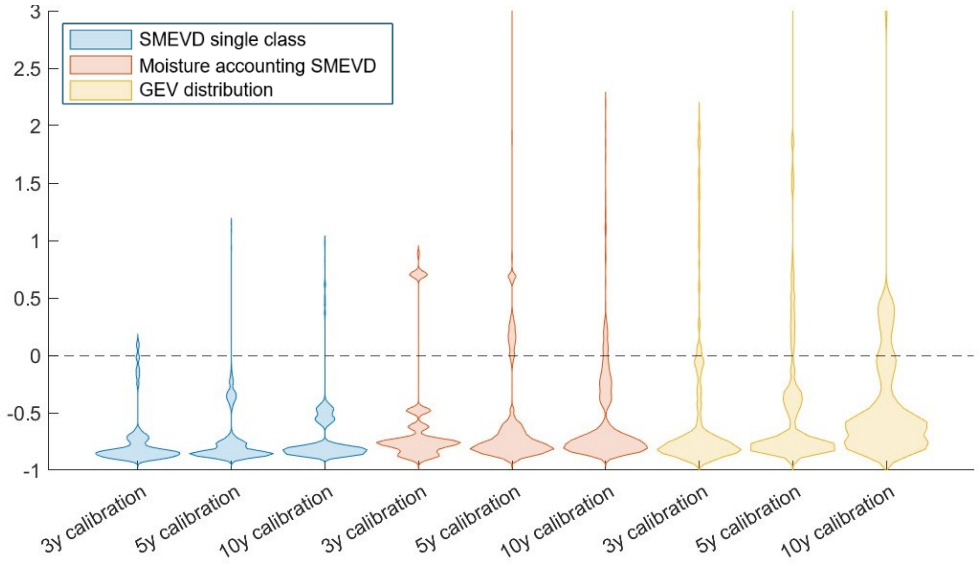


Figure 71 - Violin-plot containing the relative error distributions of three different models, referring to the last return time. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Boite stream basin.

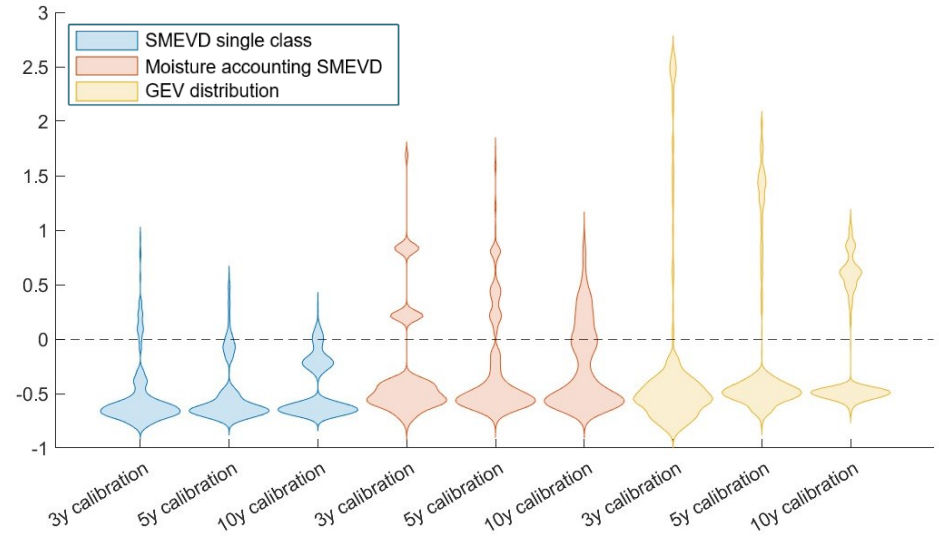


Figure 70 - Violin-plot containing the relative error distributions of three different models, referring to the last return time. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Cordevole stream basin.

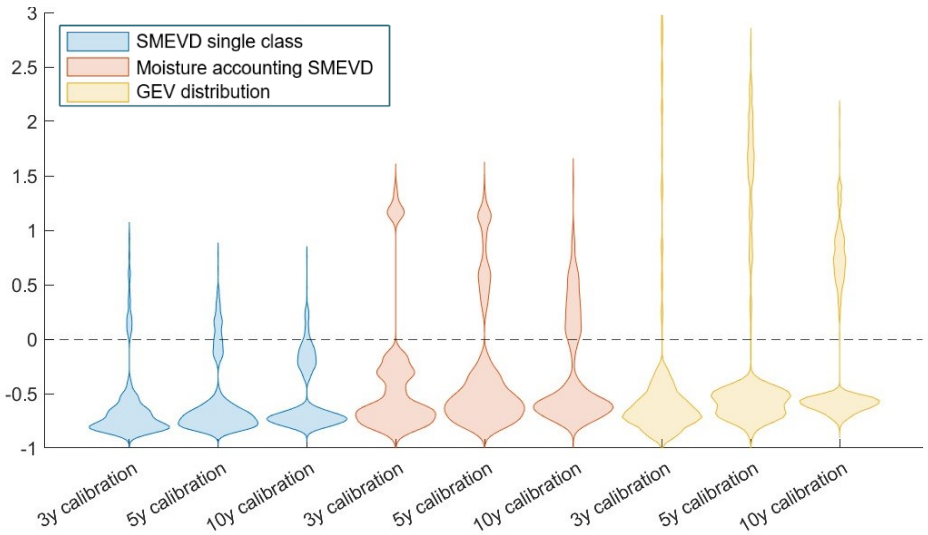


Figure 72 - Violin-plot containing the relative error distributions of three different models, referring to the last return time. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Fiorentina stream basin.

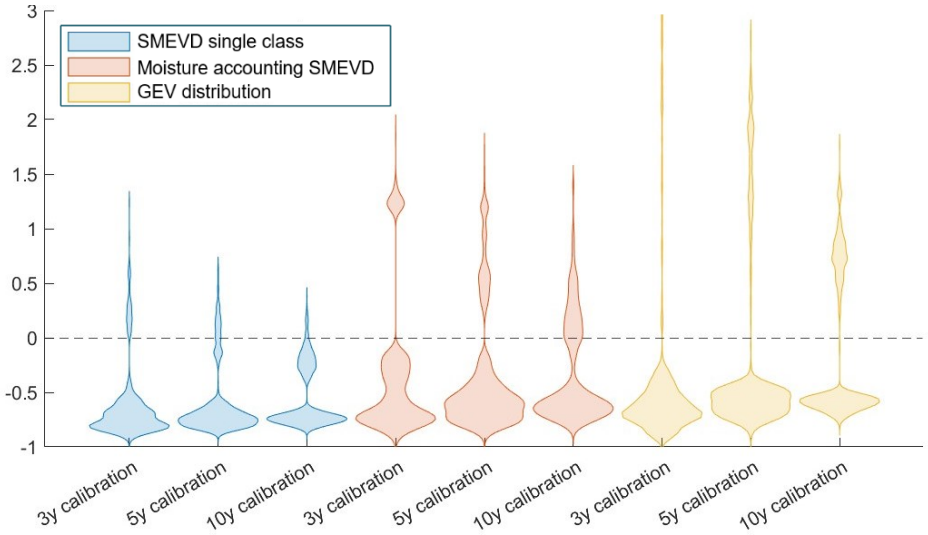


Figure 73 - Violin-plot containing the relative error distributions of three different models, referring to the last return time. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Padola stream basin.

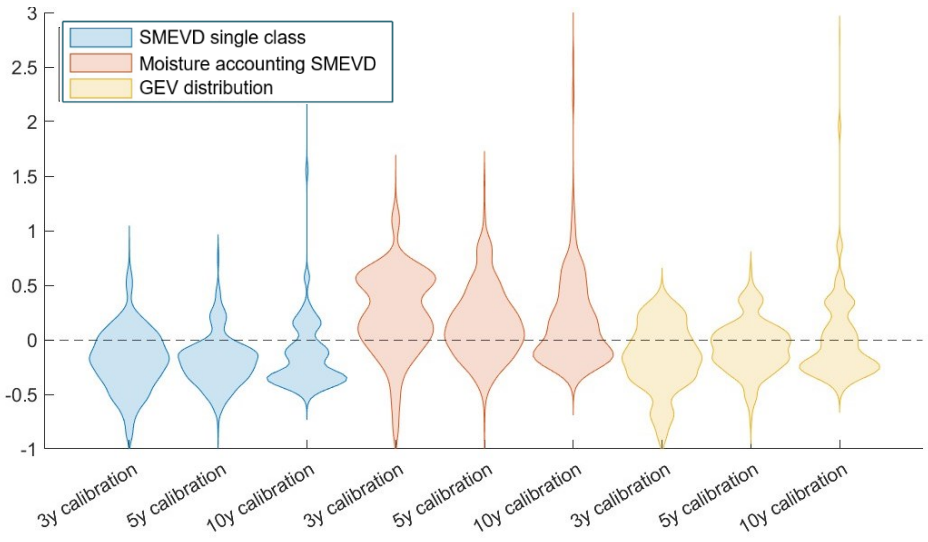


Figure 74 - Violin-plot containing the relative error distributions of three different models, referring to the last return time. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Astico stream basin.

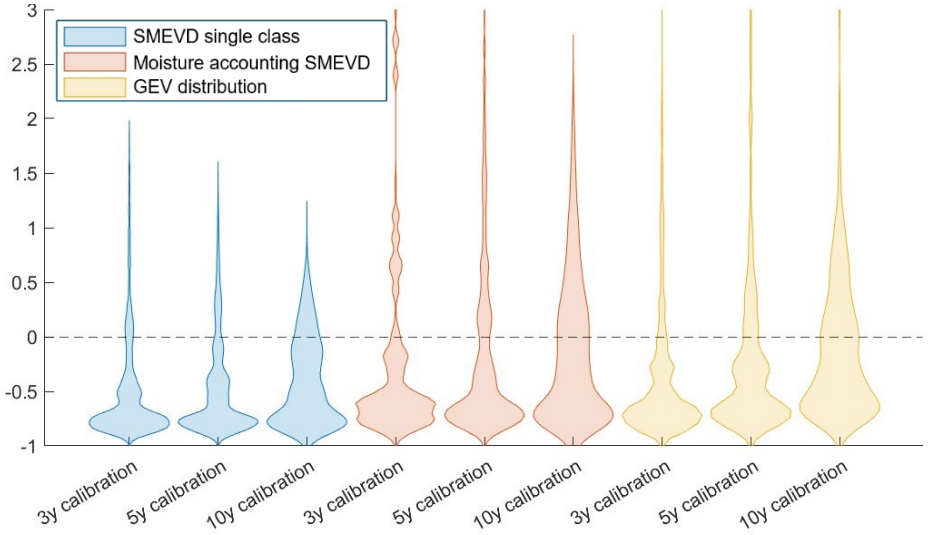


Figure 75 - Violin-plot containing the relative error distributions of three different models, referring to the last four return periods. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Boite stream basin.

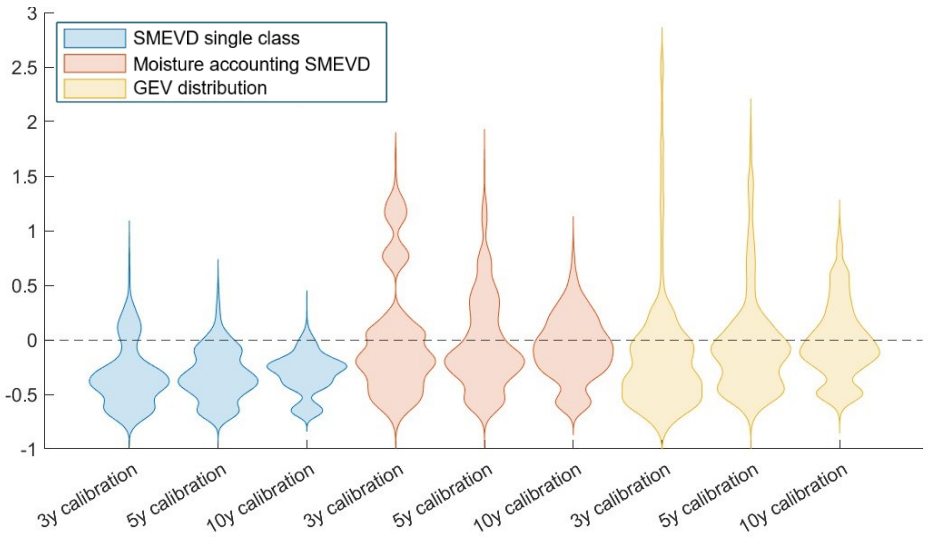


Figure 76 - Violin-plot containing the relative error distributions of three different models, referring to the last four return periods. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Cordevole stream basin.

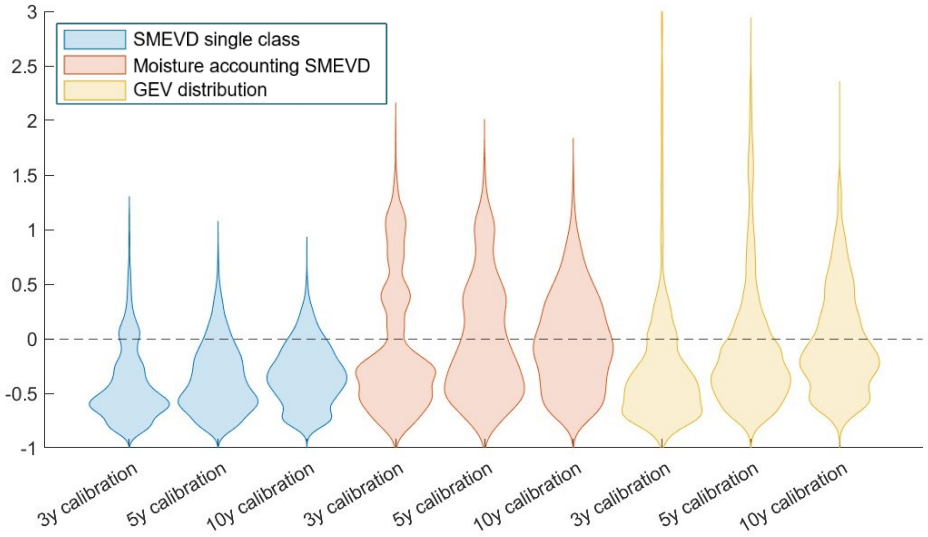


Figure 77 - Violin-plot containing the relative error distributions of three different models, referring to the last four return periods. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Fiorentina stream basin.

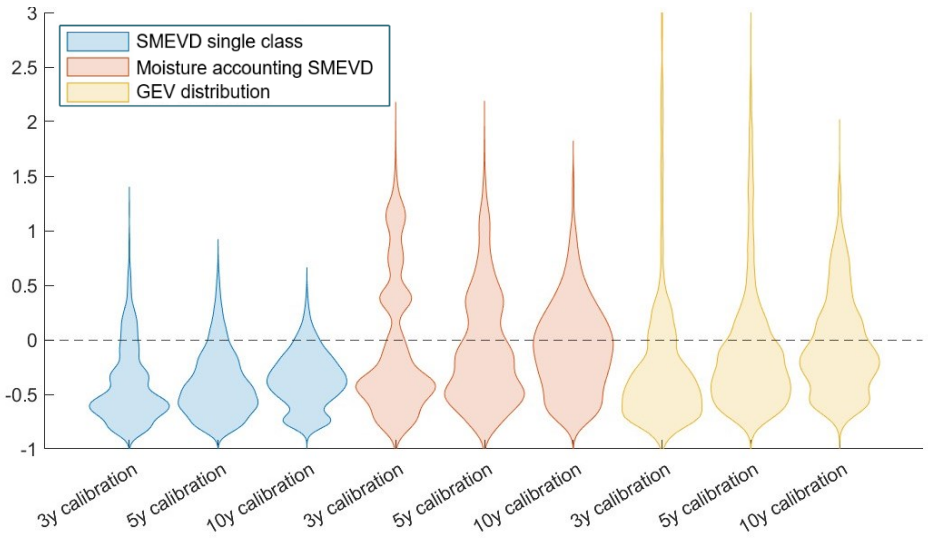


Figure 78 - Violin-plot containing the relative error distributions of three different models, referring to the last four return periods. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Padola stream basin.

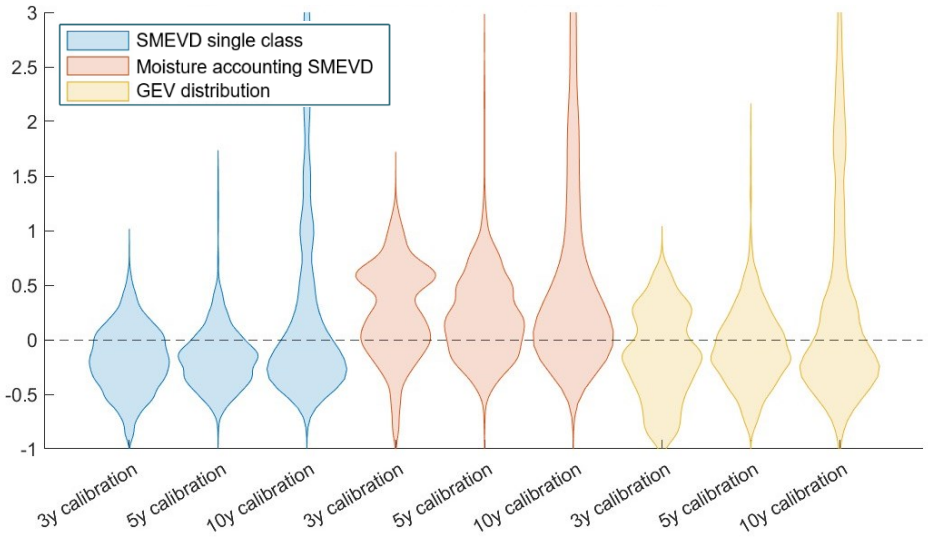


Figure 79 - Violin-plot containing the relative error distributions of three different models, referring to the last four return periods. Each model is represented considering 3, 5 or 10 years of events for the calibration sample. Referring to the Astico stream basin.

To test the performance of the models, considering the latest return times, the error indices mentioned in chapter 2.3.3, were used. As can be seen from table 7, in most cases the model that best manages to estimate events is the SMEVD applied to three classes. This is due to the fact that KGE and NSE have higher values (best case when values are equal to 1) and RMSE has values that are lower (best case when values are equal to 0). The Boite site presents the same problems as described when looking at the cross-validation graphs, which confirms the fact that no model can correctly predict flood flow events.

basins	Error indexes			
	Error types	SMEVD single class	Moisture accounting SMEVD	GEVD
Boite	NSE	-9.38	-58.74	-91.00
	KGE	0.03	-0.08	-0.70
	RMSE	41.99	43.93	34.67
	SD	0.38	0.53	0.27
Cordevole	NSE	-22.51	-14.60	-12.95
	KGE	0.36	0.48	0.30
	RMSE	11.19	11.74	10.26
	SD	0.29	0.41	0.20
Fiorentina	NSE	-8.53	-9.70	-9.14
	KGE	0.30	0.40	0.35
	RMSE	23.71	25.62	21.52
	SD	0.45	0.69	0.26
Padola	NSE	-9.46	-9.24	-8.54
	KGE	0.27	0.38	0.35
	RMSE	10.36	11.11	9.19
	SD	0.47	0.70	0.26
Posina	NSE	-3.28	-5.64	-3.89
	KGE	0.54	0.52	0.45
	RMSE	19.44	28.42	18.84
	SD	1.62	2.64	0.91

Table 7 - Error indices (NSE, KGE, RMSE and SD) to measure the performance of the models, for the various sites.

4.6 Climate change scenarios

This section discusses the results obtained from the three climate change scenarios assumed. Graphs 81-84 depict the first climate change scenario, referring to the different basins. It can be seen that by reducing the events in the highest soil moisture class by 50% and placing them in the driest soil class, a decrease in the flow rates of flood events for low to medium return times and an increase in the flow rates for high return times are induced. This is valid if one looks at the average case obtained from the random removal of events, considering 1000 iterations (in the graphs the average case is depicted with a red curve). The curve referring to the scenario of future climatic conditions has a confidence interval, which means that the forecast has a certain degree of uncertainty and can vary within the range indicated in the graphs by a pink area. In particular, graph 81 referring to the Boite torrent can be observed, which is characterized by a very wide confidence interval, meaning that it is an uncertain estimation. Graphs 85-88 depict the second climate change scenario. In this scenario, a 15% increase in the scaling parameter of the gamma distribution of the third humidity class is assumed. Compared to the previous scenario, the future climate scenario increases for both high and low return times. Furthermore, the uncertainty interval is reduced, since in this case the selection of events is not changed, but only the scaling coefficient of the gamma distribution. Graphs 89-92 depict the third climate change scenario, a hybrid case of the two previous scenarios, which assumes both a 50% reduction in events in the third moisture class and a 15% increase in the scaling coefficient. Combining the two assumptions, a more extreme climate change scenario is considered, which notes a greater increase in events with a medium to high return time and a lower reduction for events with a low return time, compared to the present climate condition. In all three scenarios there are blue circles, which correspond to the maximum annual events. Most of these events are associated with low return times, while the most extreme event is associated with storm Vaia and can have a return time varying between 200 and 1000 years depending on the site considered.

To compare the results of the various scenarios, a few return times (50, 100, 200, 500, 1000) were taken and the percentage increase or decrease of events associated with the future climate condition was calculated, compared to the current one. The results were collected in tables 8-10. Table 9 refer to the second climate change scenario and shows that applying a 15% increase to the scale parameter, the events associated with the selected return periods also undergo a 15% increase in magnitude.

basins	Climate Change Scenario 1				
	Tr = 50 y	Tr = 100 y	Tr = 200 y	Tr = 500 y	Tr = 1000 y
Boite	- 13.5%	- 7.3%	- 1.7%	+ 5.4%	+ 10.2%
Cordevole	- 10.1%	- 7.0%	- 3.7%	+ 0.3%	+ 2.8%
Fiorentina	- 11.7%	- 6.3%	- 1.2%	+ 4.8%	+ 8.4%
Padola	- 11.2%	- 6.6%	- 2.4%	+ 2.9%	+ 6.2%
Astico	- 9.1%	- 7.2%	- 5.8%	- 3.9%	- 2.8%

Table 8 - Comparison of events associated with future climate conditions, compared to current climate conditions, calculated for return times 50, 100, 200, 500, 1000. Future climate conditions are calculated using climate change scenario 1.

basins	Climate Change Scenario 2				
	Tr = 50 y	Tr = 100 y	Tr = 200 y	Tr = 500 y	Tr = 1000 y
Boite	+ 15.1%	+ 14.9%	+ 15.1%	+ 15.1%	+ 15.2%
Cordevole	+ 14.8%	+ 15.0%	+ 15.0%	+ 15.0%	+ 15.0%
Fiorentina	+ 15.2%	+ 15.0%	+ 15.1%	+ 15.1%	+ 15.0%
Padola	+ 15.0%	+ 15.4%	+ 15.1%	+ 15.1%	+ 15.0%
Astico	+ 15.0%	+ 15.1%	+ 15.1%	+ 15.0%	+ 15.0%

Table 9 - Comparison of events associated with future climate conditions, compared to current climate conditions, calculated for return times 50, 100, 200, 500, 1000. Future climate conditions are calculated using climate change scenario 2.

basins	Climate Change Scenario 3				
	Tr = 50 y	Tr = 100 y	Tr = 200 y	Tr = 500 y	Tr = 1000 y
Boite	- 0.8%	+ 6.9%	+ 13.4%	+ 19.4%	+ 26.9%
Cordevole	+ 3.2%	+ 7.4%	+ 11.1%	+ 15.3%	+ 18.1%
Fiorentina	+ 2.0%	+ 8.0%	+ 14.2%	+ 20.3%	+ 24.5%
Padola	+ 2.3%	+ 7.9%	+ 12.6%	+ 18.4%	+ 21.8%
Astico	+ 4.7%	+ 6.7%	+ 8.5%	+ 10.5%	+ 11.7%

Table 10 - Comparison of events associated with future climate conditions, compared to current climate conditions, calculated for return times 50, 100, 200, 500, 1000. Future climate conditions are calculated using climate change scenario 3.

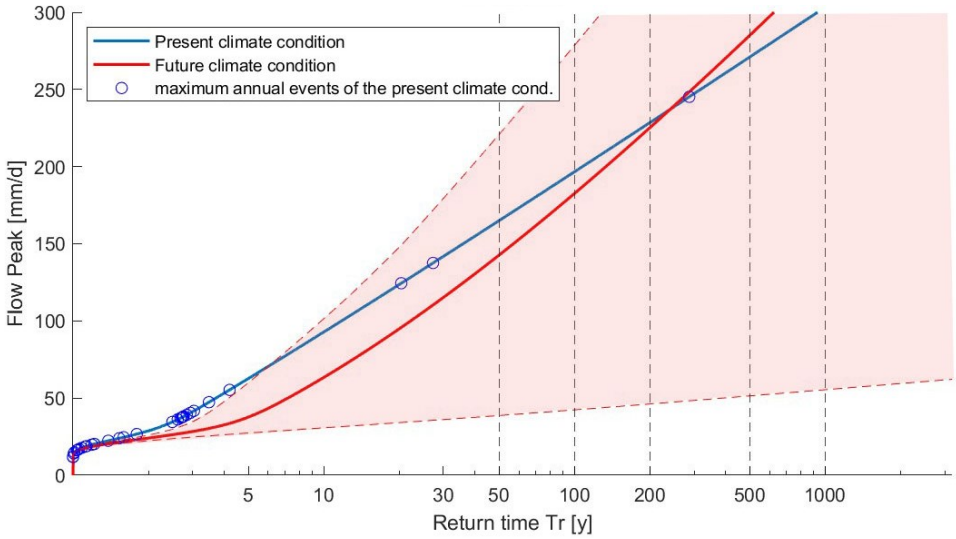


Figure 81 - Climate change scenario 1, achieved by removing 50 % of the events in the highest soil moisture class. The removed events were considered in the driest soil class. Graph referring to the Boite stream.

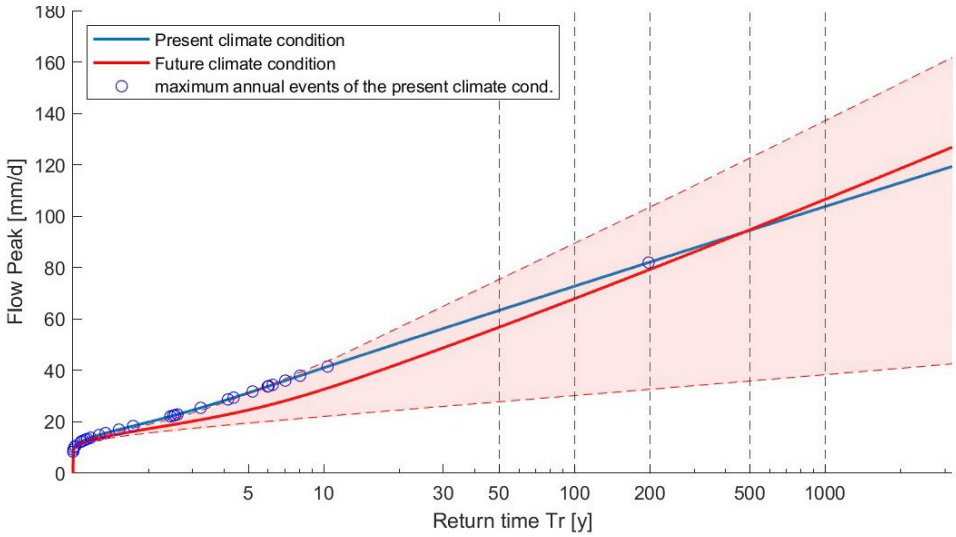


Figure 80 - Climate change scenario 1, achieved by removing 50 % of the events in the highest soil moisture class. The removed events were considered in the driest soil class. Graph referring to the Cordevole stream.

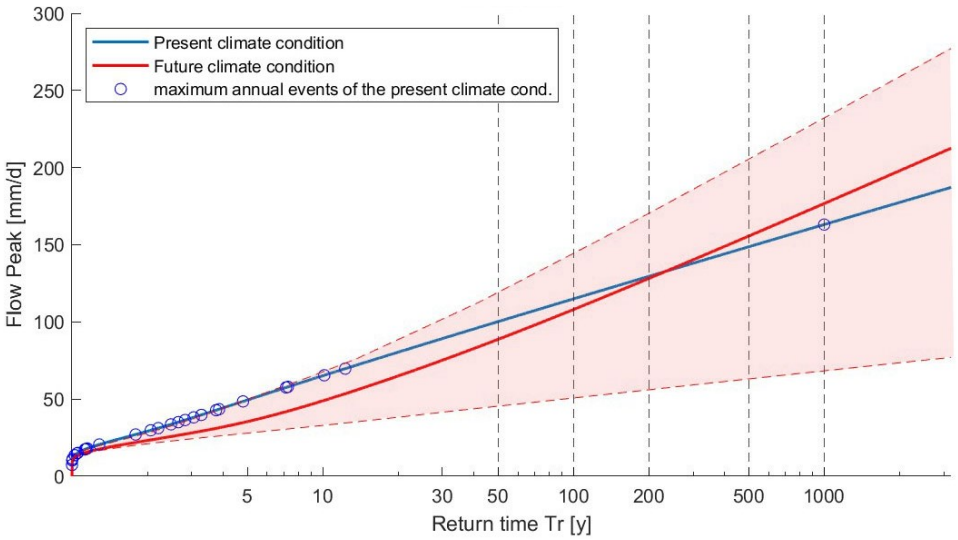


Figure 82 - Climate change scenario 1, achieved by removing 50 % of the events in the highest soil moisture class. The removed events were considered in the driest soil class. Graph referring to the Fiorentina stream.

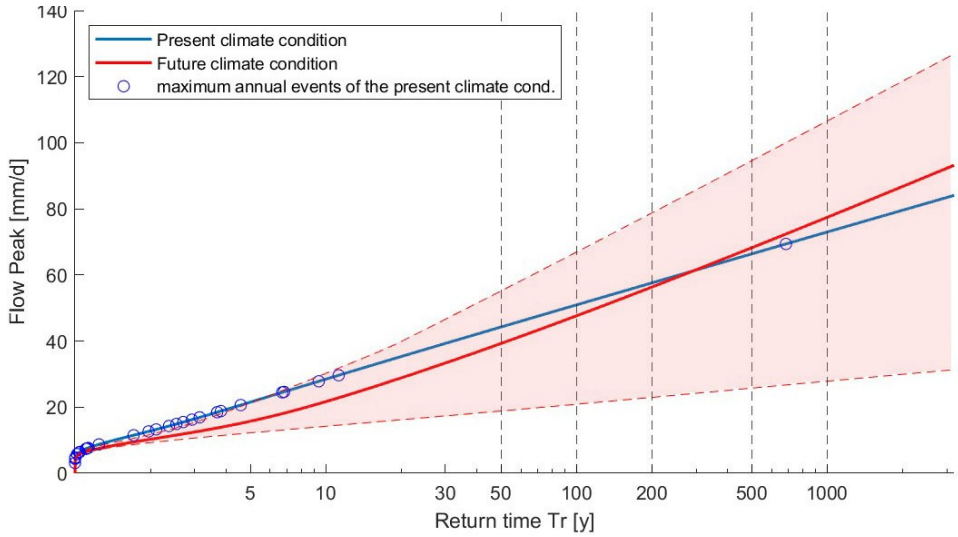


Figure 83 - Climate change scenario 1, achieved by removing 50 % of the events in the highest soil moisture class. The removed events were considered in the driest soil class. Graph referring to the Padola stream.

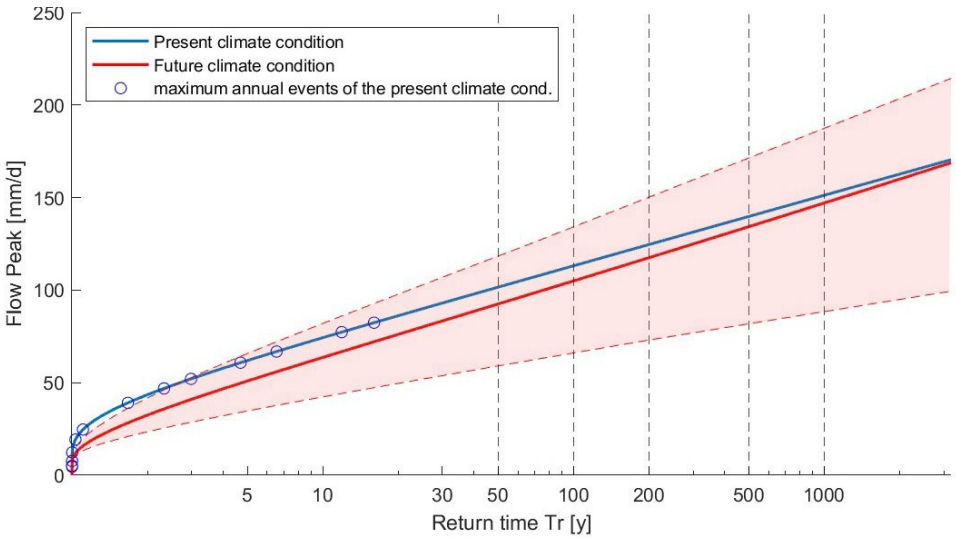


Figure 84 - Climate change scenario 1, achieved by removing 50 % of the events in the highest soil moisture class. The removed events were considered in the driest soil class. Graph referring to the Astico stream.

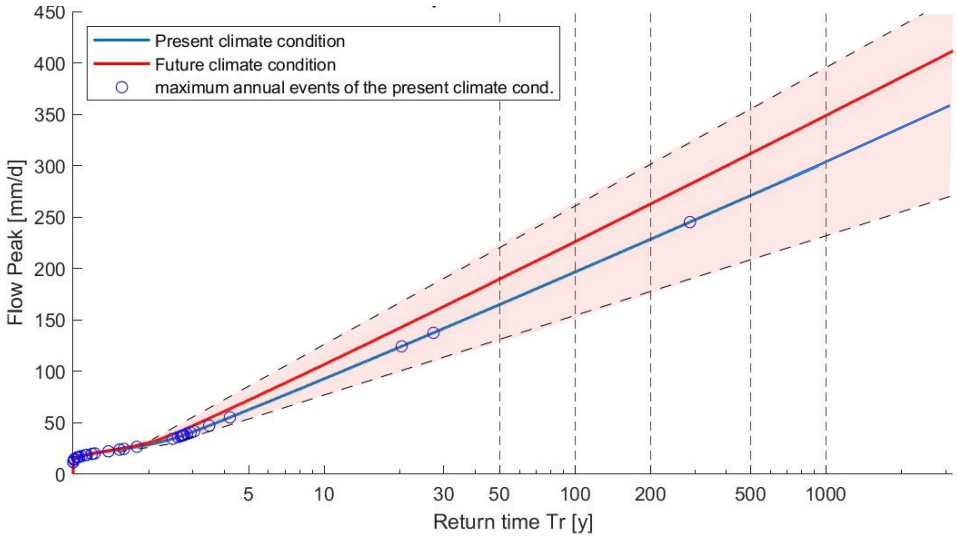


Figure 85 - Climate change scenario 2, obtained by increasing the scale parameter of the highest soil moisture class by 15%. Graph referring to the Boite stream.

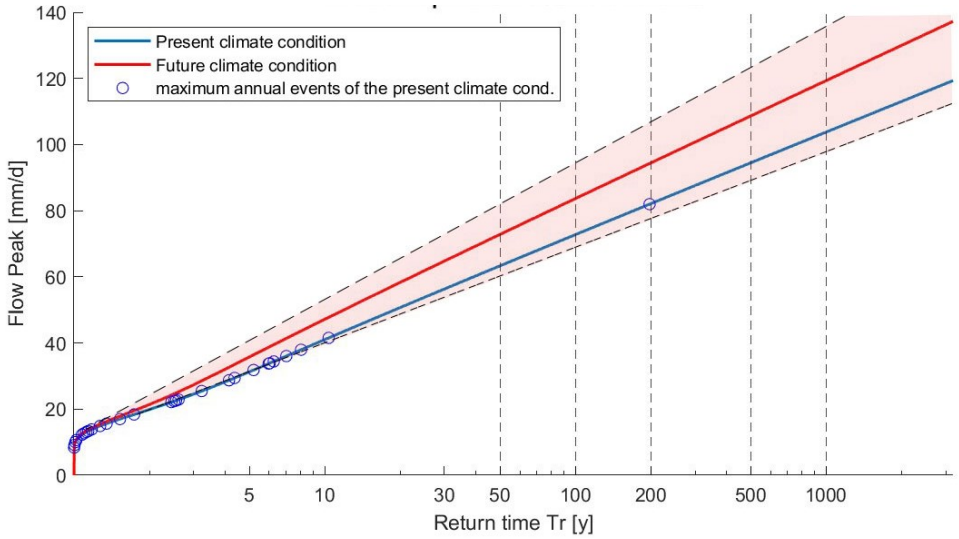


Figure 86 - Climate change scenario 2, obtained by increasing the scale parameter of the highest soil moisture class by 15%. Graph referring to the Cordevole stream.

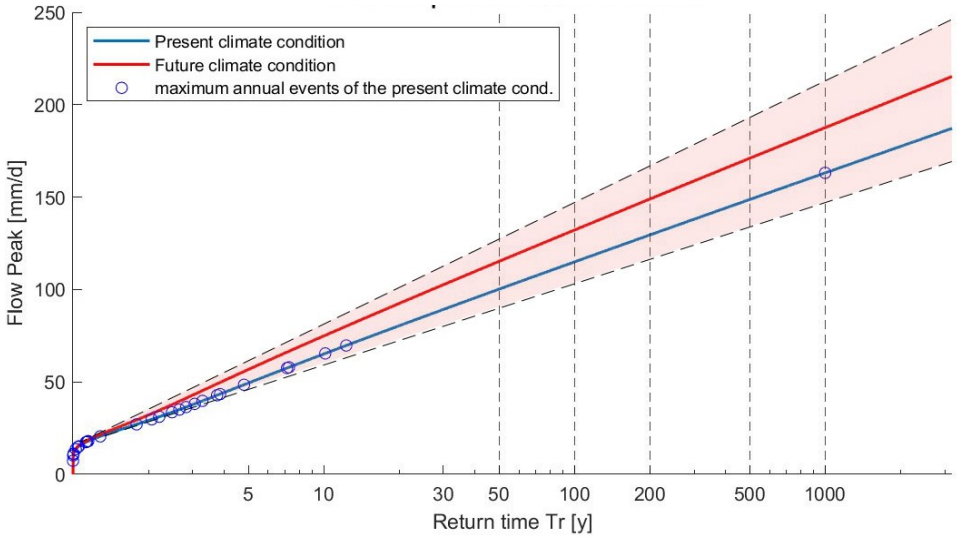


Figure 87 - Climate change scenario 2, obtained by increasing the scale parameter of the highest soil moisture class by 15%. Graph referring to the Fiorentina stream.

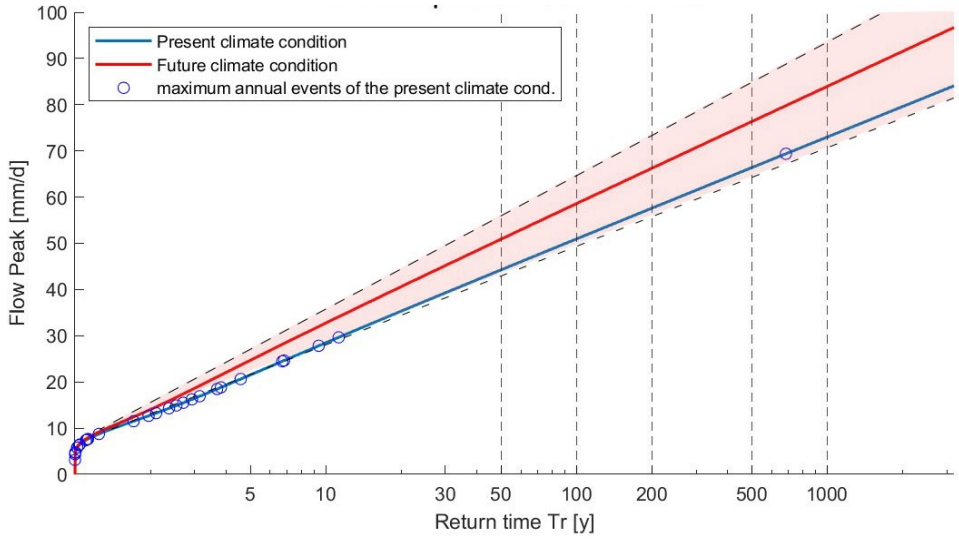


Figure 88 - Climate change scenario 2, obtained by increasing the scale parameter of the highest soil moisture class by 15%. Graph referring to the Padola stream.

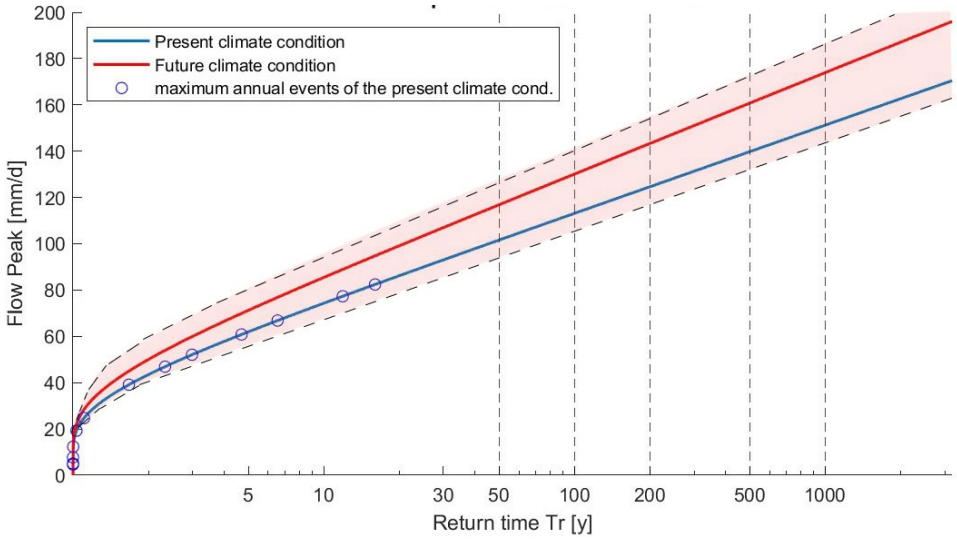


Figure 89 - Climate change scenario 2, obtained by increasing the scale parameter of the highest soil moisture class by 15%. Graph referring to the Padola stream.

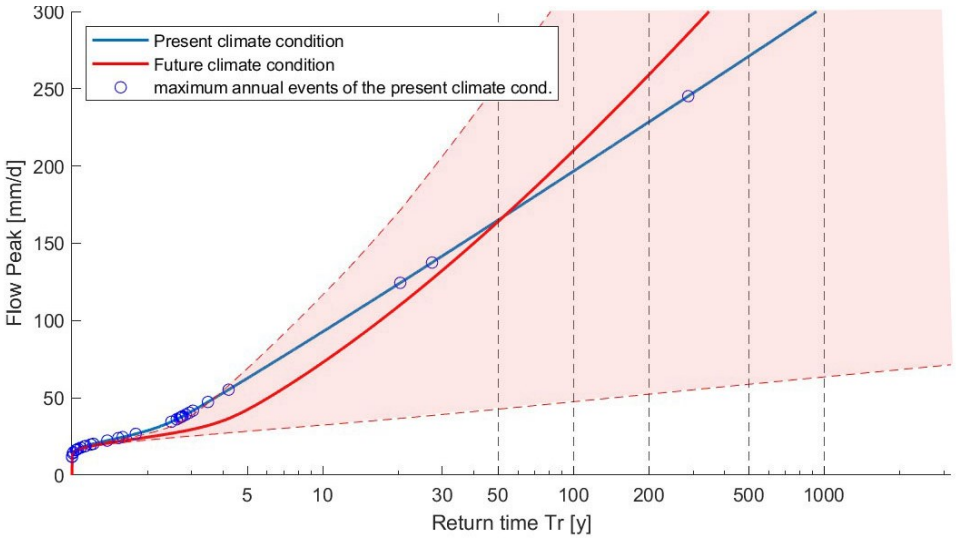


Figure 90 - Climate change scenario 3, obtained by combining the alterations implemented in scenario 1 and 2. Graph referring to the Boite stream.

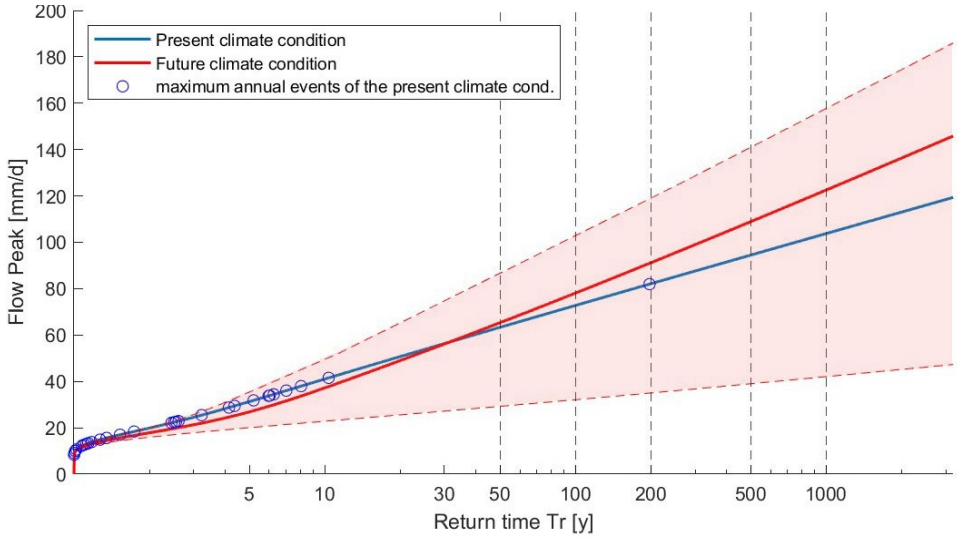


Figure 91 - Climate change scenario 3, obtained by combining the alterations implemented in scenario 1 and 2. Graph referring to the Cordevole stream.

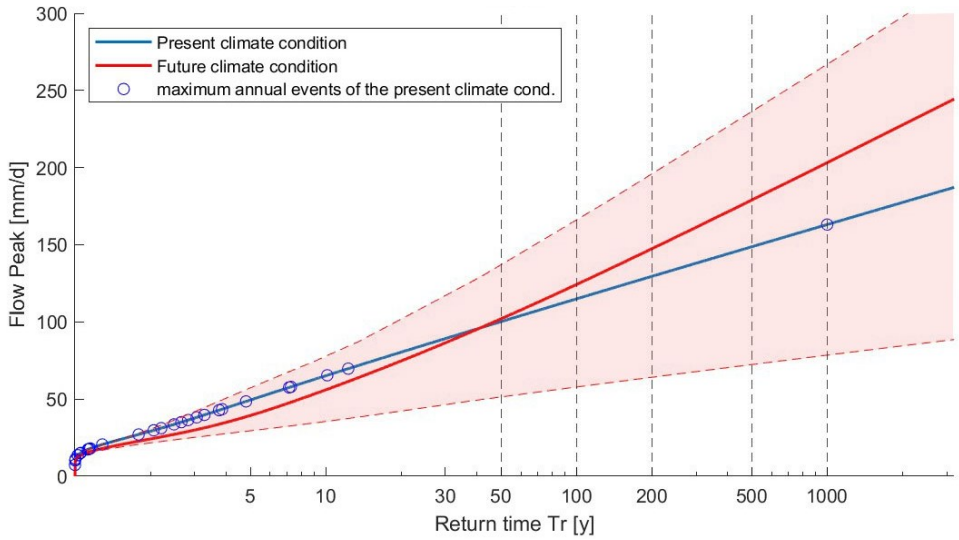


Figure 93 - Climate change scenario 3, obtained by combining the alterations implemented in scenario 1 and 2. Graph referring to the Fiorentina stream.

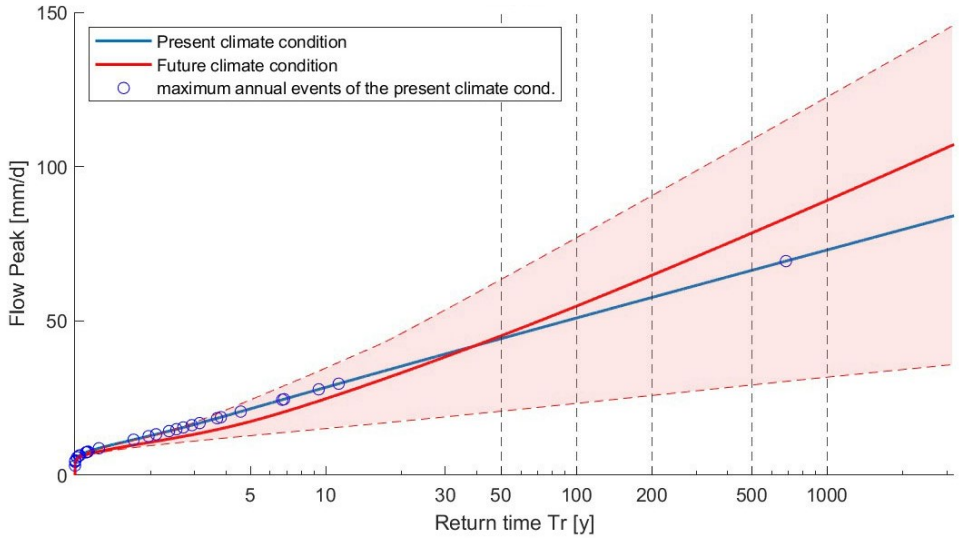


Figure 92 - Climate change scenario 3, obtained by combining the alterations implemented in scenario 1 and 2. Graph referring to the Padola stream.

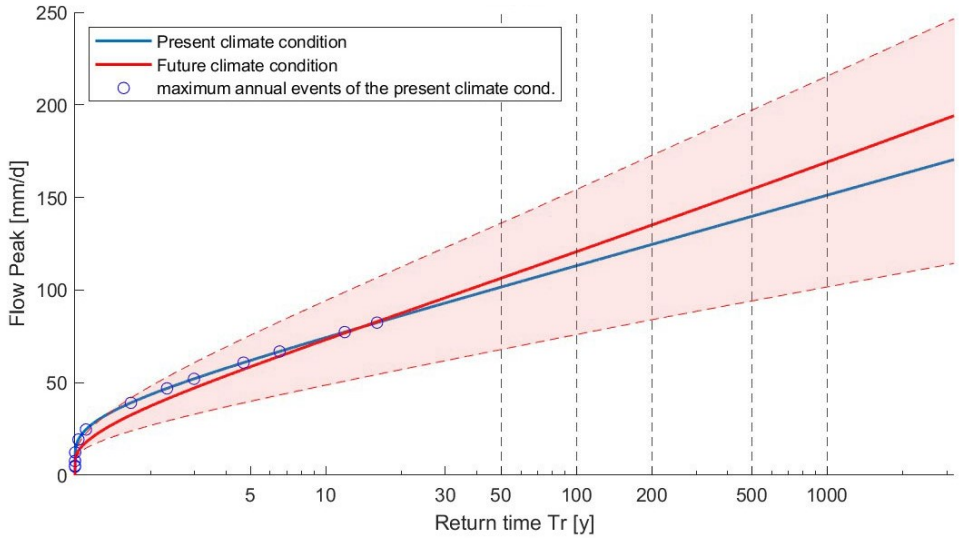


Figure 94 - Climate change scenario 3, obtained by combining the alterations implemented in scenario 1 and 2. Graph referring to the Astico stream.

5. CONCLUSIONS

In this thesis, three models for the analysis of extreme events were examined, with the aim of studying the impact of initial moisture conditions on the generation of flood events. From an in-depth analysis of the joint distributions of peak flow rate and antecedent cumulative soil moisture, it was possible to determine that the antecedent moisture calculated over the past 5 days is more correlated with the discharge as compared to the antecedent moisture calculated over longer time-windows. Subsequently, the three models were compared: one based on SMEVD considering a single class of data, another based on moisture accounting SMEVD, and the last based on GEV distribution. By comparing the goodness of fit of the models using the qq-plot, it could be seen that the model that best fits the data is the one using the GEV distribution. This is usually due to two reasons:

- Good flexibility, since the GEV distribution combines three types of extreme value distributions (Gumbel, Fréchet and Weibull);
- Asymptotic characteristics: The GEV distribution is derived from the classical extreme value theory, based on independent and identically distributed maxima (i.i.d.).

It is important to note that the performance for moisture accounting SMEVD only decreases for the event associated with storm Vaia. Cross-validation was used to measure the ability to predict unobserved events, which showed that the model based on the SMEVD applied to three soil moisture classes was able to predict flood flow events associated with high return times more accurately. The three-class SMEVD model performs better for the following reasons:

- Metastatistical and heterogeneous structure, which means that instead of assuming that the extreme data comes from a uniform and stable distribution over time, the SMEV incorporates the possibility that the source population has time-varying characteristics or comes from a combination of subpopulations.
- Better Representation of Extreme Tails, the SMEV allows for a more accurate description of the usually heavy tails typical of rare events;
- Good adaptability in contexts of climate change.

The SMEVD model based on multiple classes of data allows assumptions to be made about climate change, which is not possible with single-class SMEVD and GEV models. For the last step, three climate change scenarios were assumed using the SMEVD model. The first scenario assumes a 50% reduction of extreme events in the wettest soil class and a redistribution of these to the driest soil class. The second scenario assumes a 15% increase in the scale parameter of the gamma distribution associated with the wettest soil class. The second scenario assumes a 15% increase in the gamma distribution scaling parameter associated with the wettest soil class. Finally, the third scenario considers a combination of the variations of the first and second scenarios. The first scenario identifies an increase in the intensity of events associated with high return times and a reduction in events for medium to low return times. The second scenario allows an overall increase in event intensity of 15% to be observed. The third scenario is the most pessimistic, showing an increase in events associated with medium-high return times and a reduction in events with low Tr. The major limitations and challenges associated to the presented analysis are the following:

- Need for a wide range of data to carry out a more precise analysis (problem encountered with the study of the site associated with the Astico torrent, which had only 14 years of data).
- Fitting difficulties and reduced performance of the SMEV and GEV models with respect to extremely high magnitude events, particularly in storm Vaia events, were found at almost all of the sites analyzed.

In conclusion, this thesis highlights the advantages of using the antecedent moisture SMEVD as a tool to explore the non-obvious overall flood impact of drier conditions and more intense rainfall events likely brought about by climate change. This tool can be further developed and applied to a wider range of conditions than was done here and to account for different changes in rainfall regimes associated to global warming.

APPENDIX

The following are the boxplots of relative errors, associated with Cross Validation. The graphs are obtained considering 3 years for the calibration phase.

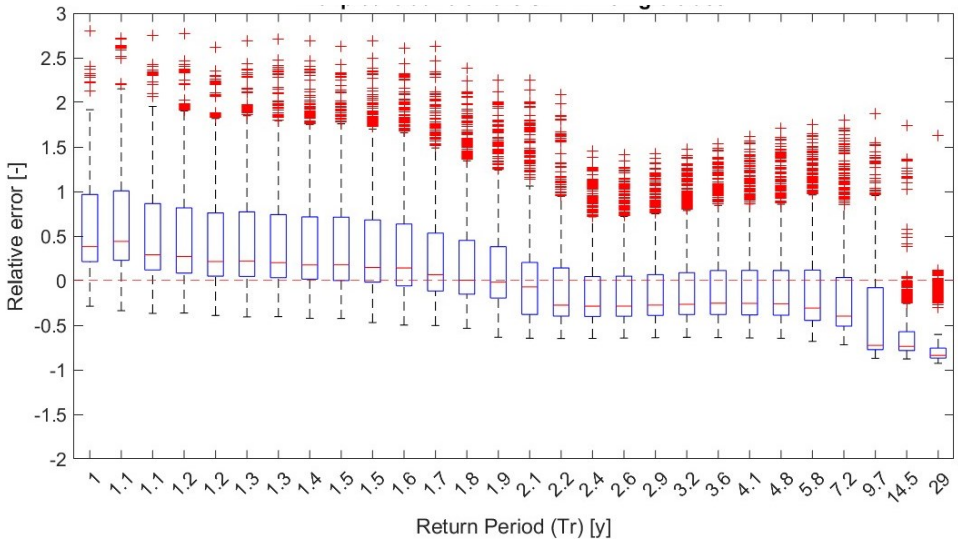


Figure 95 - Boxplot relative errors SMEVD single class. Referring to the Boite stream basin.

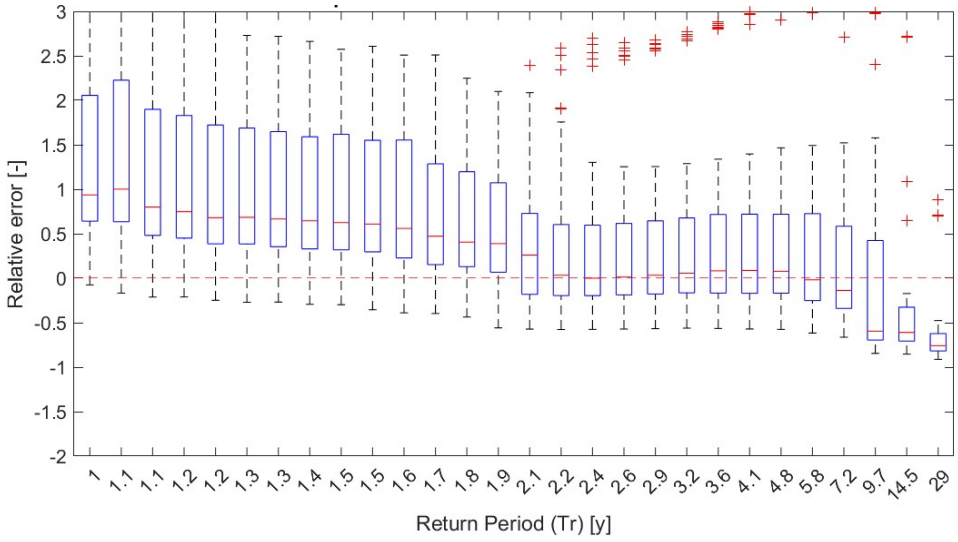


Figure 96 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Boite stream basin.

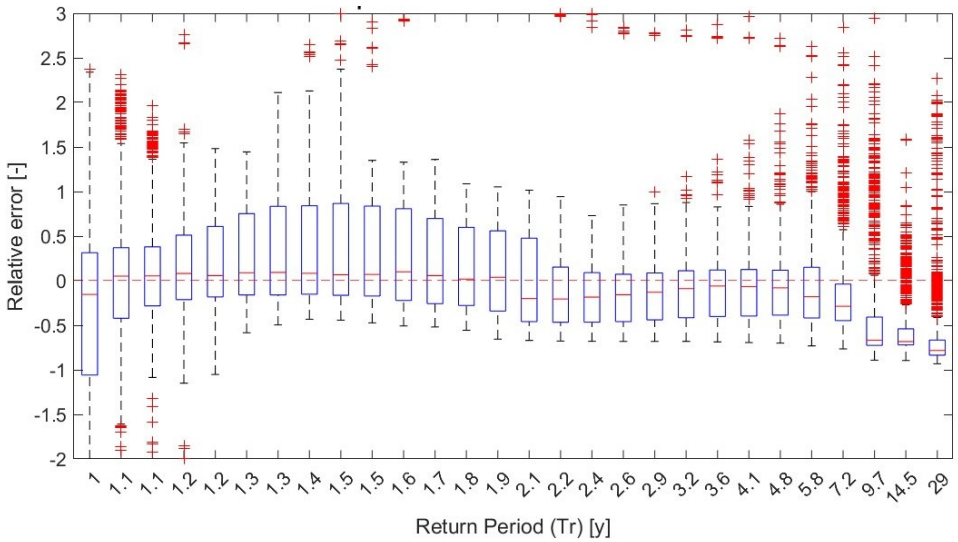


Figure 97 - Boxplot relative errors GEV distribution. Referring to the Boite stream basin.

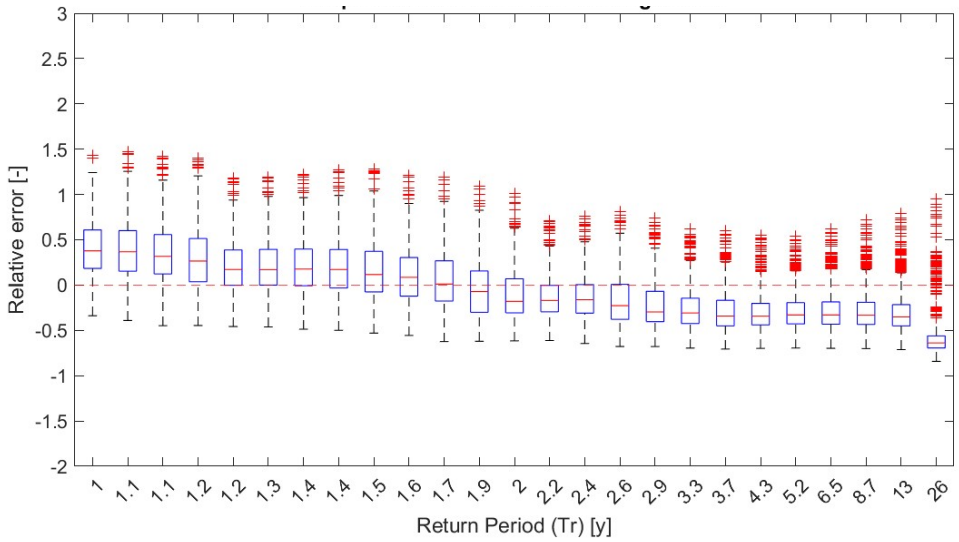


Figure 99 - Boxplot relative errors SMEVD single class. Referring to the Cordevole stream basin.

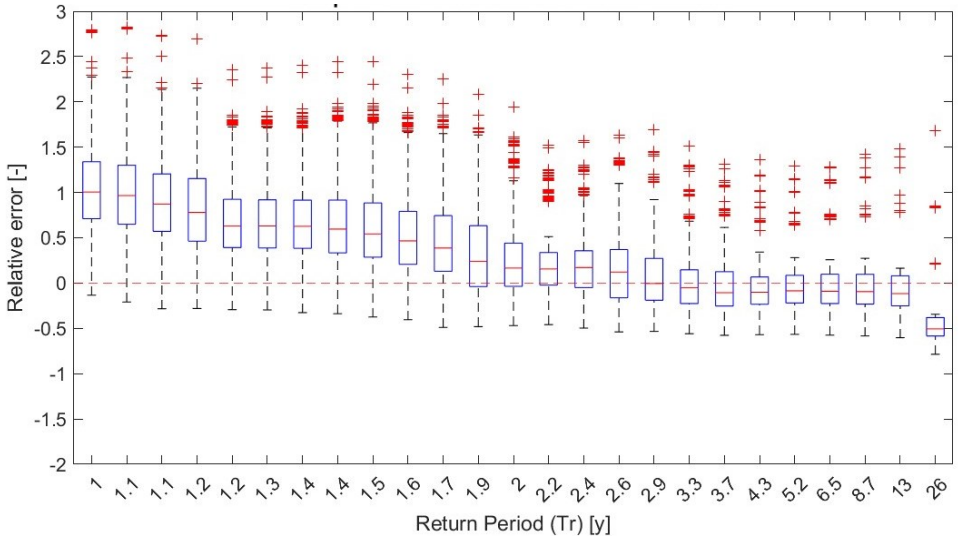


Figure 98 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Cordevole stream basin.

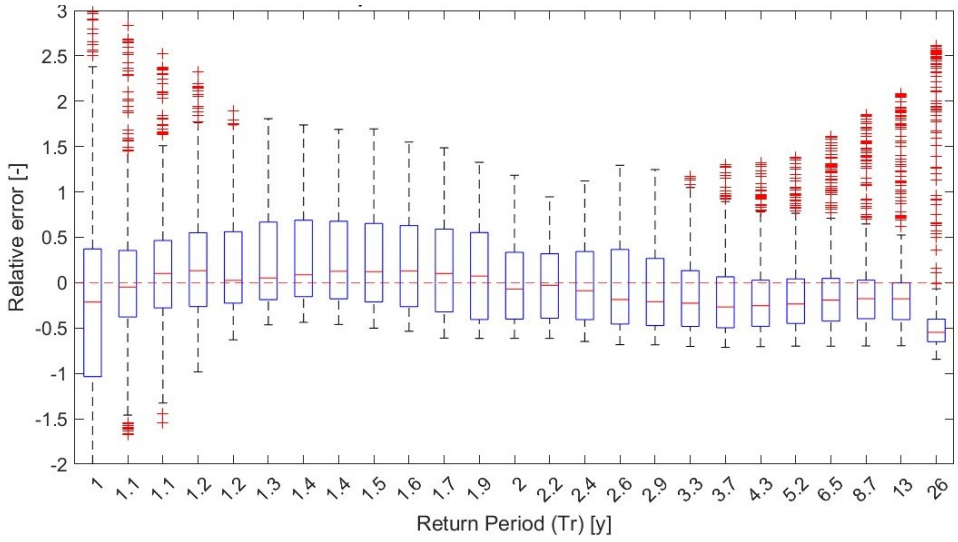


Figure 100 - Boxplot relative errors GEV distribution. Referring to the Cordevole stream basin.

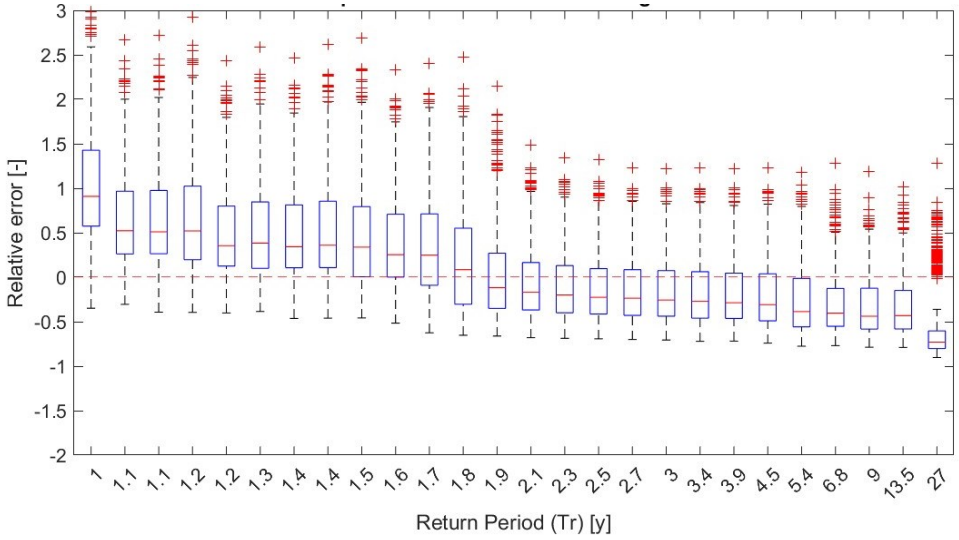


Figure 101 - Boxplot relative errors SMEVD single class. Referring to the Fiorentina stream basin.

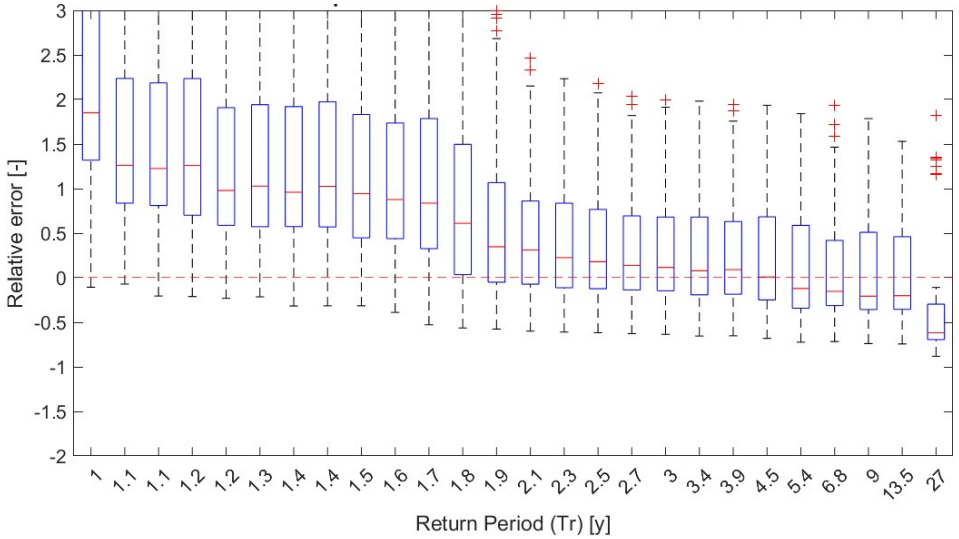


Figure 102 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Fiorentina stream basin.

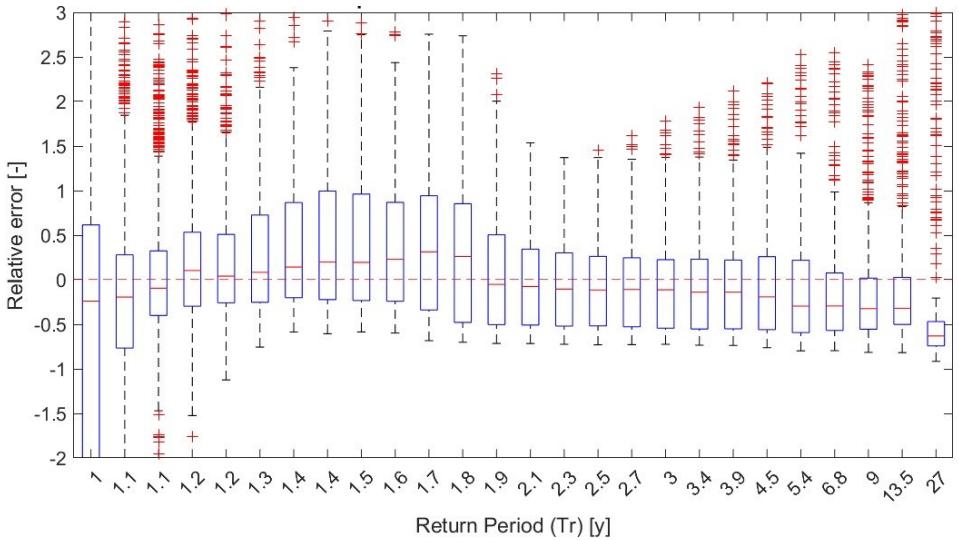


Figure 103 - Boxplot relative errors GEV distribution. Referring to the Fiorentina stream basin.

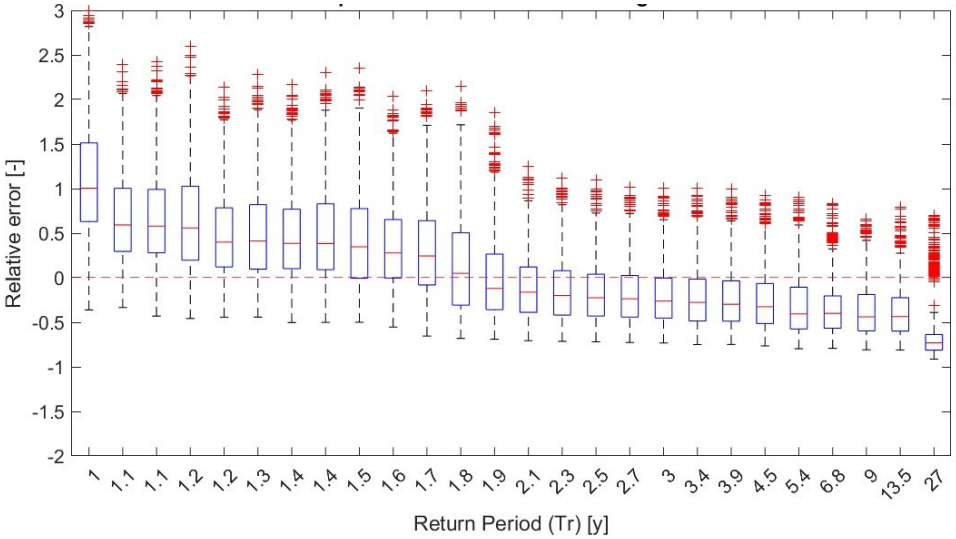


Figure 104 - Boxplot relative errors SMEVD single class. Referring to the Padola stream basin.

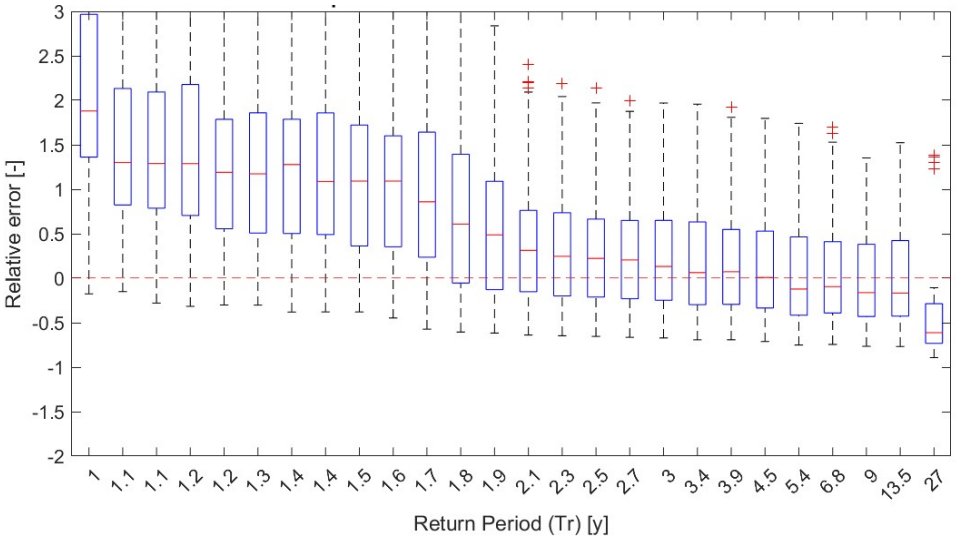


Figure 105 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Padola stream basin.

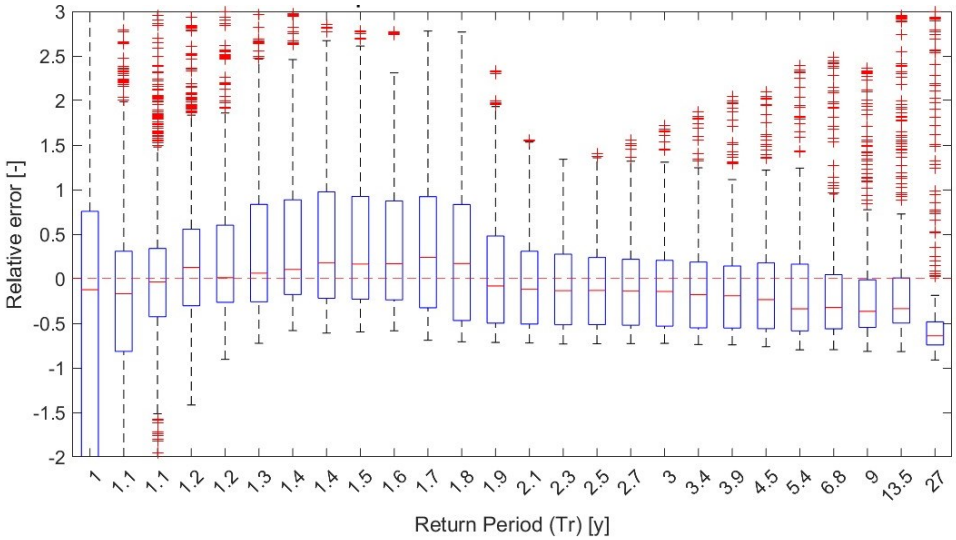


Figure 106 - Boxplot relative errors GEV distribution. Referring to the Padola stream basin.

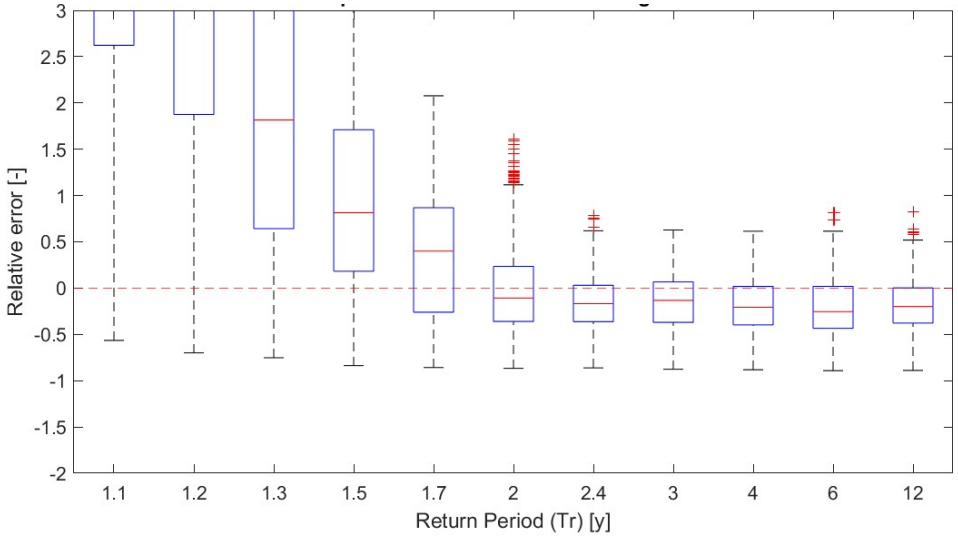


Figure 107 - Boxplot relative errors SMEVD single class. Referring to the Astico stream basin.

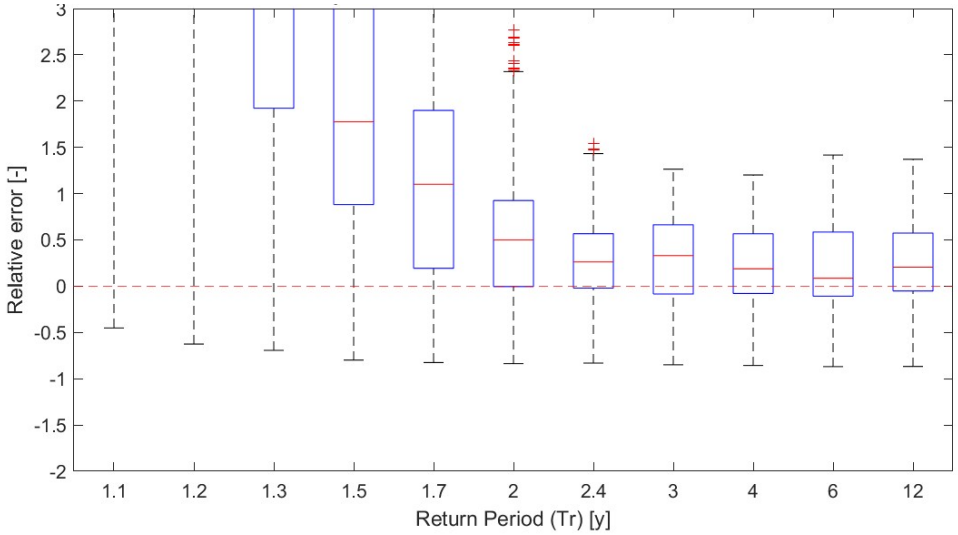


Figure 108 - Boxplot relative errors of moisture accounting SMEVD. Referring to the Astico stream basin.

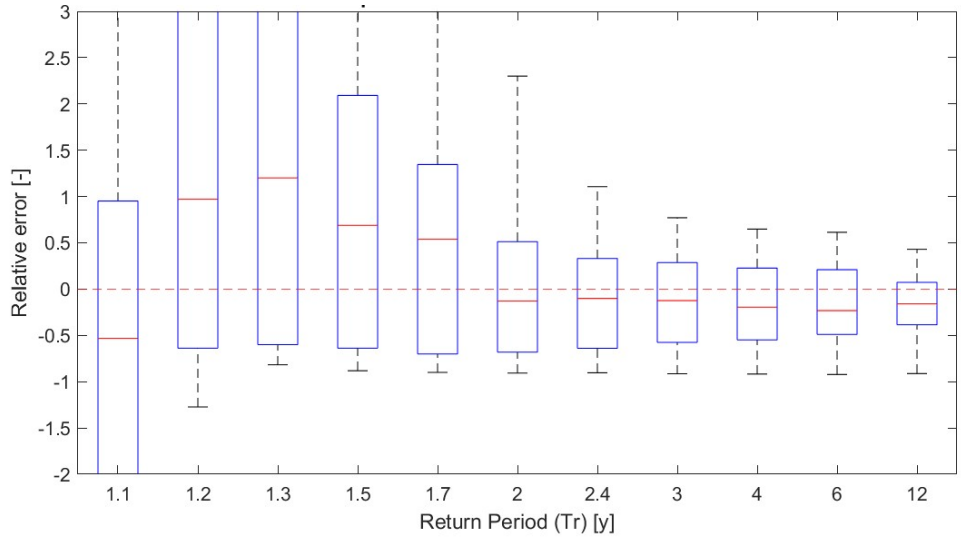


Figure 109 - Boxplot relative errors SMEVD single class. Referring to the Astico stream basin.

REFERENCES

- Alberto Calvelli (1996-1997), Thesis abstract in "Ancient Topography", https://www.antiqui.it/gis/gis4.htm#:~:text=I%20poligoni%20di%20Thiessen%20risultano,dei%20territori%20controllati%20dalle%20citt%C3%A0
- Assocodipuglia Associazione Regionale Consorzi Difesa Puglia, System implementation methodology for spatializing weather data, https://www.agrometeopuglia.it/sites/default/files/2017-09/Spazializzazione%20-%20Metodologia.pdf
- Davide Nardini (2020), Pulp Learning, "Train, Validation, Test: cosa sono e come si usano nel Machine Learning", https://pulplearning.altervista.org/train-validation-test-cosa-sono-e-come-si-usano-nel-machine-learning/
- E. Zorzetto, G. Botter, M. Marani (2016), "On the emergence of rainfall extremes from ordinary events", https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016GL069445
- Gareth James, Daniela Witten, Trevor Hastie and Rob Tibshirani (2013), Springer ,"An Introduction to Statistical Learning"
- Gianluca Botter (2012), "STATISTICS and DATA ANALYSIS LECTURE NOTES"
- Gianluca Botter, Enrico Bertuzzo and Andrea Rinaldo (2010), AGU Publications, "Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox", https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2009wr008371.
- Gianluca Botter, Stefano Basso, Ignacio Rodriguez-Iturbe and Andrea Rinaldo (2013), Proceedings of the National Academy of Sciences (PNAS), "Resilience of river flow regimes", https://www.pnas.org/doi/full/10.1073/pnas.1311920110#:~:text=The%20hydrological%20resilience%20of%20river,and%20the%20related%20socioeconomic%20impact.
- IPCC, general information, https://www.ipcc.ch/
- ISPRA, Extreme events,https://www.isprambiente.gov.it/it/attivita/cambiamenti-climatici/eventi-estremi

- J. Laherrère & D. Sornette (1998), “Stretched exponential distributions in nature and economy: “fat tails” with characteristic scales”, https://link.springer.com/article/10.1007/s100510050276

- J. R. M. Hosking (2020), jstor, “L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics”, https://www.jstor.org/stable/2345653

- Jamile Vasconcelos, George Lima, Marwan Wehaiba El Khazen, Adriana Gogonel & Liliana Cucu-Grosjean (2023), “On vulnerabilities in EVT-based timing analysis: an experimental investigation on a multi-core architecture”, https://link.springer.com/article/10.1007/s10617-023-09277-5

- Marco Marani1, Enrico Zorzetto and Arianna Miniussi (2019), Università di Perugia, “Beyond traditional extreme value theory: lessons learned from rainfall and hurricane intensity”, https://warredoc-unistrapg.org/wp-content/uploads/2019/01/Marco-Marani_Beyond-traditional-extreme-value-theory.pdf

- Marco Marani and Massimiliano Ignaccolo (2015), ScienceDirect, “A metastatistical approach to rainfall extremes”, https://www.sciencedirect.com/science/article/pii/S0309170815000494

- Maria Francesca Caruso (2022/2023), “Extremes in the hydrological cycle: A metastatistical framework applied to extreme coastal flooding and droughts”, https://www.research.unipd.it/handle/11577/3485021

- Paras Varshney (2023), built in, “Q-Q Plots Explained”, https://builtin.com/data-science/q-q-plot

- Rashid Akbary (2022/2023),“Adjustment of Depth-Duration-Frequency Curves Under Changing Climate”, https://thesis.unipd.it/handle/20.500.12608/50850

- S. Basso, M. Schirmer and G. Botter (2016), AGU Publications, “A physically based analytical model of flood frequency curves”, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016GL069915

ACKNOWLEDGMENTS

Vorrei dedicare questo spazio a ringraziare tutte le persone che hanno contribuito alla realizzazione di questo lavoro, rendendo possibile il raggiungimento di questo traguardo. In primo luogo, il mio più sentito ringraziamento va al mio relatore, Gianluca Botter e al correlatore Marco Marani, per il supporto e i consigli che mi hanno dato durante tutto il percorso. La loro pazienza è stata fondamentale per le mie continue domande. Un sincero grazie anche alla mia famiglia, che mi ha sostenuto in ogni fase di questo cammino accademico. Il loro affetto e incoraggiamento sono stati per me un sostegno importante. Un grazie speciale a Maria Francesca Caruso e Pietro Devò, che mi hanno assistito con grande disponibilità nella risoluzione dei problemi legati al codice Matlab, e ai colleghi dello studio i4 consulting per la calorosa ospitalità e il supporto durante tutto il periodo del tirocinio. Non posso dimenticare i miei amici, in particolare i compagni dell'università e la compagnia di Abano, che hanno condiviso con me momenti di confronto, motivazione e anche di svago, rendendo questo percorso più leggero e piacevole. Infine, un pensiero speciale va a tutti coloro che, direttamente o indirettamente, mi hanno offerto ispirazione e strumenti per superare le sfide incontrate durante questi anni.

Grazie di cuore a tutti.