

BIVARIATE ANALYSIS AND SYNTHESIS OF FLOOD EVENTS FOR THE DESIGN OF HYDRAULIC STRUCTURES – A CASE STUDY FOR ARGENTINA

Master Science Thesis

Water Resources and Environmental Management

Leibniz Universität Hannover

Ana Claudia Callau Poduje

2802080

31/08/2012

Examiner: Prof. Dr.-Ing. Uwe Haberlandt and Dr.-Ing. J. Dietrich

Supervisor: Dipl.-Ing. Aslan Belli

**Master Thesis Assignment for
Mrs. Ana Claudia Callau Poduje (Mat.-No.: 2802080)**

**“Bivariate analysis and synthesis of flood events for the design of hydraulic
structures - a case study for Argentina”**

General Problem:

The design of hydraulic structures like dams and polders depends not only on flood peaks, but also on flood volume, duration and shape of the hydrograph. The classical univariate flood frequency analysis which is based on peaks only is not sufficient if retention of the hydraulic structure is of importance for the design. Also, it is not possible to define a single design hydrograph with a single return period, when the hydrograph is described by more than one random variable. A possible approach considering the most important two variables, flood peak and volume, is the bivariate statistical analysis and synthesis of flood events. Consequently the adequacy of a given hydraulic structure can then be tested in a simulation experiment as shown for instance by De Michele et al., (2005). Mrs. Callau has to carry out such a simulation experiment for one river basin in Argentina where sufficient data is available. The bivariate statistical approach should make use of the copula concept. As model hydrograph the Kozeny function could be used as described by Klein (2009). A comparison of the results of the univariate constrained model hydrograph by peak only and the bivariate constrained model hydrograph by peak and volume should be provided.

Tasks:

1. Carry out a literature research and state of the art review
2. Analyse the study region, collect data and define the problem
3. Select methods for hydrograph separation, bivariate statistics, synthesis of model hydrographs, reservoir routing, etc.
4. Carry out the analyses
5. Evaluate, compare and discuss the results considering current practice in Argentina
6. Write a thesis report about your work, which is to be delivered in form of 2 hard copies and 1 electronic pdf-file
7. Present and discuss your results in a colloquium

References:

- De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., Rosso, R., 2005. Bivariate Statistical Approach to Check Adequacy of Dam Spillway. Journal of Hydrologic Engineering, 10(1): 50-57.
- Klein, B., 2009. Ermittlung von Ganglinien für die risikoorientierte Hochwasserbemessung von Talsperren. PhD Thesis, Ruhr-Universität Bochum, Schriftenreihe Hydrologie/ Wasserwirtschaft, Nr. 25.

Issue date: 1.03.2012

Hand in date: 31.08.2012

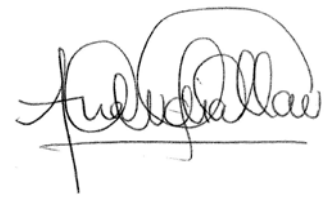
Consultant: Dipl.-Ing. A. Belli

Examiner: Prof. Dr.-Ing. U. Haberlandt

Dr.-Ing. J. Dietrich

DECLARATION

I declare that this research paper for the degree of Master of Science in Water Resources and Environmental Management, Faculty of Civil Engineering and Geodetic Science, Leibniz Universität Hannover hereby submitted has not been submitted by me or anyone else for a degree at this or any other university. That it is my own work and that materials consulted have been properly acknowledged.

A handwritten signature in black ink, appearing to read 'F. J. J. J. J.', written over a horizontal line.

Hannover, September 2012

ABSTRACT

A bibliographic review of the state of art of flood estimation techniques shows the importance of studying floods as multivariate events. The aim of this work is to discuss whether a multivariate analysis is necessary for designing dams or for assessing the adequacy of spillways belonging to existing dams. A multivariate approach offers a range of possible events associated to a joint return period, which can be used in a design stage. The multivariate criterion includes information related with the dependence structure linking the flood variables. This dependence is involved in the generation of random flood samples, which are used for risk assessment of existing dams.

Generally, peak flows and volumes are two statistically dependent variables. In this work, a flood event is considered as a multivariate event characterized by these two variables. A bivariate statistical frequency analysis is carried out to find a suitable model that adequately represents the data set of flood peak flow and volume. The dependencies of the flood characteristics are modeled with the copula method.

The copula model is used to generate 1000 random pairs of variables characterizing the flood and assess the risk of overtopping a dam. A set of pairs of values associated to a joint return period of 1000 years, provided by the copula, is used within a proposed dam design procedure. These pairs of variables are then transformed to hydrographs, applying the Beta distribution function to represent the shape of a flood hydrograph. The synthetic flood events are used to test the behavior of the dam, by routing them through the reservoir. The maximum water levels and outflows are considered as the most important factors in the risk assessment and design and are computed for each routed hydrograph. To show and discuss the advantages of the multivariate approach using copulas, the results are compared with values obtained using the univariate approach applied to the same variables independently.

A bivariate frequency analysis is carried out on the basis of the flood peak and volume series observed in the river Agrio basin, located in the province of Neuquén, Argentina. The catchment has a total drainage area of 7300 km² at the station Bajada de Agrio which provides 50 years of complete daily streamflow data. The selected watershed has two extreme flood events per year, produced by rainfall-excess during winter and snowmelt runoff combined with rainstorms during spring.

The results indicate that the estimated floods based on the bivariate frequency analysis applied to the case study results in higher values of maximum water elevations and outflows compared with the univariate approach. If the risk of an existing dam is to be assessed, the bivariate approach would result in a greater risk of overtopping the dam for a given dam height and spillway. If a dam is to be designed, the consideration of a joint return period would result in estimated design parameters, i.e. maximum water height and outflow, which are higher compared to the values estimated using the univariate models. In both cases the bivariate approach results in higher values of water levels and outflows, compared to the univariate approach. The final conclusion of this work is that incorporating multiple variables to describe flood events, results in values that are more critical both for the risk assessment and for the design of hydraulic structures, such as dams.

Key words: copula, dam safety, flood frequency, joint return period, probability density functions, synthetic hydrograph.

TABLE OF CONTENTS

DECLARATION	i
1. INTRODUCTION	1
1.1. Problem setting.....	1
1.2. Objective of the work	5
1.3. Structure of the work	6
2. LITERATURE REVIEW	8
3. MATERIALS AND METHODS	10
3.1. Flood identification.....	10
3.2. Statistical analysis	12
3.2.1. Box & Whisker plot.....	12
3.2.2. Double Sum analysis.....	13
3.2.3. Mann-Kendall test	13
3.2.4. Pettitt test	14
3.2.5. Grubbs and Beck test.....	15
3.3. Frequency analysis.....	15
3.3.1. Univariate frequency analysis	16
3.3.2. Multivariate frequency analysis using copulas	17
3.3.3. Return period.....	22
3.4. Synthetic hydrograph generation.....	26
3.5. Reservoir routing	32
4. STUDY AREA AND DATA.....	35
4.1. Description of the catchment	35
4.2. Hydrological data.....	37
4.3. Hypothetical reservoir	38
5. RESULTS AND DISCUSSION	40
5.1. Hydrograph separation	40
5.1.1. Base flow separation	40
5.1.2. Identification of independent events.....	41
5.1.3. Results of the hydrograph separation	41
5.1.4. Relationship between the resulting series.....	46
5.2. Extreme value statistics.....	47

5.2.1.	Univariate frequency analysis	47
5.2.2.	Bivariate frequency analysis using copula	51
5.2.3.	Return period.....	55
5.3.	Synthetic hydrograph.....	56
5.3.1.	Analysis of the observed hydrographs	56
5.3.2.	Synthesis of hydrographs.....	62
5.4.	Reservoir routing	65
5.4.1.	Risk assessment for dam safety	65
5.4.2.	Dam design procedure	68
6.	SUMMARY AND CONCLUSIONS.....	72
7.	FUTURE WORKS.....	75
8.	ACKNOWLEDGMENTS	75
	REFERENCES	76
	APPENDIX A	79
	APPENDIX B	81
	APPENDIX C	83
	APPENDIX D.....	86

LIST OF FIGURES

Figure 3.1: Comparison between simulated samples extracted from different 2-dimensional copulas (Normal, Frank and Gumbel), for Kendall's tau of 0.80.....	19
Figure 3.2: Comparison between simulated samples extracted from different 2-dimensional copulas (Clayton, Galambos and Hüsler Reiss), for Kendall's tau of 0.80.	20
Figure 3.3: Comparison between simulated samples extracted from Clayton's 2-dimensional copulas, for Kendall's tau of 0.65 (left) and 0.80 (right).....	20
Figure 3.4: Joint return periods for a Clayton's 2-dimensional copula. Left: "OR" case; Right: "AND" case.....	24
Figure 3.5: Reservoir routing components: Inflow hydrograph, spilled flows and reservoir levels....	34
Figure 4.1: Map with general location of the study area.	35
Figure 4.2: Monthly values of precipitation and temperature in the upper catchment.	36
Figure 4.3: Monthly flows river Agrio.....	37
Figure 4.4: Map of the catchment area and location of the gauging station (red mark).....	37
Figure 4.5: Characteristic curves of the reservoir. Left: Surface-height curve; Right: Volume above spillway crest-height curve.	39
Figure 5.1: Components of the hydrograph.	40
Figure 5.2: Example of the resulting base flow separation for one year.	41
Figure 5.3: Example of the resulting independent flood events for one year.	41
Figure 5.4: Observed pairs of peak flow and direct volume series. Left: Winter events; Right: Spring events.....	43
Figure 5.5: Box Plots of total peak flow and direct volume series for both the winter and spring flood events.....	44
Figure 5.6: Double sum analysis of series observed during winter. Left: Peak flow; Right: Direct volume.....	45
Figure 5.7: Double sum analysis of series observed during spring. Left: Peak flow; Right: Direct volume.....	45
Figure 5.8: Relationship between flow and volume series for the winter flood events. Left: Peak flow; Right: Volume.....	46
Figure 5.9: Relationship between flow and volume series for the spring flood events. Left: Peak flow; Right: Volume.....	46
Figure 5.10: Comparison of two univariate models fitted to the peak flow series observed during the winter events. Left: Generalized Pareto model; Right: Weibull (3 parameters) model....	48
Figure 5.11: Univariate models selected to represent the flood characteristics of the winter events. Left: Generalized Pareto model fitted to the peak flows; Right: Generalized Pareto model fitted to the direct volumes.	49
Figure 5.12: Univariate models selected to represent the flood characteristics of the spring events. Left: GEV model fitted to the peak flows; Right: Weibull (3 parameters) model fitted to the direct volumes.	49
Figure 5.13: Synthetic series of peak flow and direct volume, random generation using the univariate distribution functions. Left: Winter events; Right: Spring events.....	50
Figure 5.14: Scatter plots of scaled ranks of the pairs of observed values (peak flow, direct volume). Left: Winter events; Right: Spring events.....	51

Figure 5.15: Ranked-pairs: observed during the spring events and generated using the two best copula models. Left: Clayton copula - Moments-like estimation. Right: Normal copula - Moments-like estimation.....	53
Figure 5.16: Ranked-pairs: observed during the winter events and generated using the same copula and different estimation methods. Left: Clayton copula – Pseudolikelihood estimation. Right: Clayton copula - Moments-like estimation	53
Figure 5.17: Synthetic series of peak flow and direct volume, random generation of pairs of values using the copula models. Left: Winter events; Right: Spring events.	54
Figure 5.18: Pairs of peak flow and direct volume values associated to the design return period of 1000 years, estimated by univariate and bivariate analysis. Left: Winter events; Right: Spring events.	56
Figure 5.19: Beta distributions representing the transformed hydrographs of the three extreme winter events in terms of the adimensional peak flows with values: ≈ 3 (1963), ≈ 6 (1984) and > 13 (2001)	58
Figure 5.20: Application of the Beta distribution to reproduce the maximum winter flood events. Left: Maximum peak flow, 2001; Right: Maximum direct volume, 1972.	58
Figure 5.21: Application of the Beta distribution to reproduce the maximum spring flood events. Left: Maximum peak flow, 1994; Right: Maximum direct volume, 2002.	59
Figure 5.22: Ranked-pairs: observed and generated using the best copula models. Left: Normal copula, winter events. Right: Gumbel copula, spring events.....	60
Figure 5.23: Observed and synthetic series of Beta parameters (a and b) corresponding to the winter floods. Left: Univariate analysis; Right: Bivariate analysis.....	60
Figure 5.24: Observed and synthetic series of Beta parameters (a and b) corresponding to the spring floods. Left: Univariate analysis; Right: Bivariate analysis.	61
Figure 5.25: Observed and synthetic series of adimensional time to peak and peak flow corresponding to the winter floods. Left: Univariate analysis; Right: Bivariate analysis..	62
Figure 5.26: Observed and synthetic series of adimensional time to peak and peak flow corresponding to the spring floods. Left: Univariate analysis; Right: Bivariate analysis..	62
Figure 5.27: Synthetic hydrographs generated with one pair of flood characteristics and different hydrograph shapes.	63
Figure 5.28: Synthetic hydrographs generated with one hydrograph shape and different combinations of flood characteristics.	64
Figure 5.29: Relationship of the adimensional time to peak and the maximum water elevations reached in the reservoir. Left: Winter events; Right: Spring events.	65
Figure 5.30: Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic winter floods.....	66
Figure 5.31: Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic spring floods.....	66
Figure 5.32: Empirical distributions of maximum spilled flows resulting from the routing of 1000 synthetic winter floods.....	67
Figure 5.33: Empirical distributions of maximum spilled flows resulting from the routing of 1000 synthetic spring floods.....	67
Figure 5.34: Analysis of the sensitivity of the shape parameters. Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic winter floods, considering only one shape of hydrograph for all the cases.	68

Figure 5.35: Random set of flood characteristics associated to a joint return period of 1000 years. Left: Winter flood events; Right: Spring flood events.	69
Figure 5.36: Empirical distributions of maximum reservoir level resulting from the routing of synthetic winter floods associated to a return period of 1000 years.....	69
Figure 5.37: Empirical distributions of maximum reservoir level resulting from the routing of synthetic spring floods associated to a return period of 1000 years.....	70
Figure 5.38: Empirical distributions of maximum spilled flows resulting from the routing of synthetic winter floods associated to a return period of 1000 years.....	70
Figure 5.39: Empirical distributions of maximum spilled flows resulting from the routing of synthetic spring floods associated to a return period of 1000 years.....	71

LIST OF TABLES

Table 3.1: Summary of 2-dimensional copulas: family, probability functions, parameter space and relationship of non-parametric dependence measure with parameter.	19
Table 3.2: Possible cases in base flow estimation approach.....	31
Table 4.1: Basic statistics of observed flood peaks series.....	38
Table 5.1: Basic information of the characteristics of the winter and spring floods.	42
Table 5.2: Cross-correlation coefficients for the peak flow and volume series.	42
Table 5.3: Sample statistics of the flood characteristic series: peak flow and direct volume.....	43
Table 5.4: Results on the Mann-Kendall (p-values) and Pettitt Tests (probabilities of a change point) applied to the peak flow and volume series.	45
Table 5.5: Goodness of fit of univariate models applied to the series of peak flow and direct volume. Cramer-von Mises statistics.	47
Table 5.6: Estimated parameters of the selected univariate models representing the flood characteristic series: peak flow and direct volume.	50
Table 5.7: Measures of the degree of dependence corresponding to the analyzed series.	52
Table 5.8: Goodness of fit of bivariate copula models applied to the series of peak flow and direct volume. Cramer-von Mises statistics and associated p-value.....	52
Table 5.9: Estimated variables associated to a design criteria of 1000 years return period.....	55
Table 5.10: Sample statistics of the shape parameter series: Adimensional time to peak (t_{peak^*}), adimensional peak flow (Q_{peak^*}) and the Beta parameters (a and b).....	57
Table 5.11: Selected univariate and copula models representing the shape parameters: a and b (Beta distribution parameters).....	59

1. INTRODUCTION

Hydraulic infrastructures along rivers, such as dams, levees or bridges, are designed in to safely operate during a flood event. An appropriate design of the flood control capacity of a reservoir can ensure the safety of a dam and avoid overtopping. The control capacity of the reservoir and maximum reservoir levels reached during a flood event are related with the adequacy of the dam's spillway.

In this study the hydrological safety of a dam is analyzed to check the adequacy of the spillway. The study is carried out using synthetic flood hydrographs and testing the behavior of the reservoir. Hydrological phenomena are multidimensional and often require the joint modeling of the random variables, for this reason in this work the flood events are characterized using peak flow and volume. The adequacy the spillway is tested by generating random pairs of peak flow-volume combinations. This generation is done using copulas and marginal distributions. The random pairs of characteristic values are transformed into hydrographs, which are routed through the reservoir. The maximum reservoir levels and outflows resulting from each of the synthetic hydrograph routings are computed.

The bivariate frequency of flow peak and volume is studied using copulas. First univariate distributions are fitted to the observed series separately. Then a copula is used to construct a bivariate distribution by modeling the dependence between the two variables and linking the two univariate marginal distributions. Finally the copula model is used to derive the information required for planning and managing flood mitigation facilities, like joint probabilities and return periods.

To show and discuss the advantages of the multivariate approach using copulas, the results are compared with values obtained using the univariate approach applied to the same variables independently. This work focuses on a method based in a range of possible flood events, instead of following a one-design-event methodology, and the results are presented as empirical frequencies. The approach followed in this work could be used both as an evaluation method to check the adequacy of an existing reservoir, and as a methodology for designing hydraulic structures.

1.1. Problem setting

A disaster is an event capable of causing damage or danger to human and animal life and/or property. The origin of the event can be natural or man-made, and they can usually not be eliminated, but certainly their impact can be mitigated. Flooding is an important damaging catastrophe among all natural disasters. Singh (1996) describes and accounts losses related with flood events worldwide, in terms of monetary units and amount of people displaced. In UNESCO (2003) it is reported that floods account for half of all water related disasters in the world from 1990 to 2001.

Maidment (1993) describes the complexity of the nature of the flood producing systems, which includes the interaction with the atmosphere, land geology and geomorphology, vegetation and soils, and the activities of people. Due to this complexity, the observed flow records are the best information for the estimation of future floods.

Flood conditions usually occur as a result of extreme hydro-meteorological events such as high rainfall and/or snowmelt. A flood consists in high water levels overtopping the natural or artificial banks of a stream or a river.

The management of floods involves different approaches that include controlling the flood, reducing the vulnerability to flood damages, and minimizing the loss after the flood. The flood control can be accomplished by modifying the land use or the stream. The stream can be modified using hydraulic structures like reservoir, levees, embankments, diversion works, channel improvement, etc. The vulnerability can be reduced by flood plain management: structural modifications, flood forecasting and warning, development policies, response planning, etc. The purpose of flood control is to reduce or eliminate the damage caused by the flooding of areas adjacent to rivers, or by the overtopping of engineering structures like dams, embankments or bridges. The failure of a dam caused by overtopping can cause a serious disaster. To avoid this type of disasters dams are designed to “stand safe” even in the presence of rare floods with estimate return periods of 1000 years or higher depending on the height of the dam and on the vulnerability down-streams.

Planning, design, and operation of flood related engineering works require the estimation of a primary design variable, known as the design flood. For many hydrological problems a detailed knowledge of flood event characteristics, such as flood peak, volume and duration, is necessary. The design flood is not only a physical term, its definition involves several dimensions that include public policy, political, legal, economic, and technical (Nagy et al., 2002). This design criterion could be peak flow, maximum water level, flood volume, or the entire flood hydrograph.

Special consideration should be given when designing important hydraulic structures like dams or flood control embankments. This type of works involves the adoption of design flood hydrographs according to the design standards to ensure the safety of the structure. These design standards are based either on the Probable Maximum Flood or on a given return period. The return period is the average time elapsed between two successive events that exceed a certain threshold. A proper estimate of the flood design and extreme inflow hydrographs is required in order to avoid undesirable problems that could threaten the safety of the structure. This is an important challenge for hydrologists and is related with both the safety and the economic viability of dams.

Design thresholds are defined in terms of the acceptable hydrological risk of the structure. In the case of a dam's hydraulic structures, these values could be a maximum reservoir level or/and maximum flow released through the spillway during the flood event. The geometry of a spillway is defined considering the peak of the outflow hydrograph and the risk of exceeding a given water level in a reservoir. The inflow hydrograph is routed through the reservoir and the peak is lowered by storage and release, resulting in the spilled flows.

The present work involves the analysis of flood events, and focuses on the adequacy of a reservoir behavior in presence of a flood event. Dams are hydraulic structures build on rivers to store water and regulate the natural flow regime of the river. The storage of the water is the reservoir, in which part of the flood waters can temporarily be retained to help reduce the flood peak. The passage of water masses through a reservoir is stretched over a longer period of time at a reduced flow rate.

Natural events are often characterized by the joint behavior of several random variables, which are usually dependent. The flood phenomenon is a multidimensional process, having as important hydrologic features the peak flood flow along with the corresponding volume and duration, as well as rate-of-rise and rate of recession. The flood peak value is an important factor to define the height

of a dam or to design a spillway. The flood volume may also be essential in the design of cases in which the reservoir storage capacity is small compared to the flood volume.

Many works show that there is a significant positive dependence between the flood peak and volume (Shiau (2003), Favre et al. (2004), Shiau et al. (2006), De Michele et al. (2005), Genest et al. (2007b), Nijssen et al. (2009), Chowdhary et al. (2011), Salvadori et al. (2011), Vandenberghe et al. (2012)). As a consequence, the relevant events should be defined in terms of two or more variables. A good approach for analyzing these complex events is through the joint distribution of several random variables, considering the correlation among them.

As mentioned before, the flood hydrograph can be characterized by random variables such as flood peak, volume, duration and time to peak. In a two-dimensional frequency analysis, two of the hydrograph parameters should be fixed, while the rest are obtained by some relationship with the fixed parameters or from a chosen hydrograph shape distribution. The question that arises is how many of the variables characterizing the flood should be considered in the frequency analysis and how the relationship between the variables should be included. In the following some of the arguments and discussions found in the bibliography related with this question are presented.

Chowdhary et al. (2011) states that the decision to employ a univariate, bivariate or multivariate flood frequency analysis should be made on the basis of the objectives of any particular application. For a small or moderate sized flood protection structure a univariate analysis may be sufficient, while in situations where storage has a significant effect on flood attenuation, the peak flood discharge should be considered along with the duration and/or volume. According to Mediero et al. (2010), the relative influence of the inflow peak flow and volume when assessing the hydrological risk of a dam depends on the reservoir surface, spillway length and whether the spillway is controlled or uncontrolled. Generally a flood with a great peak flow will be more critical for dams with small reservoir areas, while floods with greater volumes should be carefully analyzed when dealing with dams with large reservoir areas. This statement depends on the length of the spillway crest and could be modified if the length is affected. The spillway geometry can be defined in terms of the risk of the dam, probability of exceeding a certain water level, or in terms of risk of flooding a location downstream from the dam, probability of exceeding an outflow discharge.

A univariate hydrological frequency analysis of complex events like floods or storms can only provide limited assessment of the events, because these events are characterized by some correlated random variables (peak, volume, duration, etc.). The hydrologic risk may be underestimated by using only the flood peak for the probability measure: an event with a very high peak and small volume may be stored in a reservoir while another event with a lower peak but higher volume may cause a spill from the reservoir.

Multivariate frequency analysis which consider peak discharges and other parameters, such as direct runoff volume or duration of hydrograph, provide a better characterization of the inflow hydrographs, reducing the uncertainty in flood analysis, compared with the univariate flood frequency analysis of the peak discharges (Goodarzi et al., 2011). A multivariate approach is a better approach to represent a hydrological extreme event than a univariate, since the event is characterized by multiple properties. However, a multivariate frequency analysis requires much more information and more sophisticated mathematical analysis. Therefore, a common method of representing the event is the bivariate frequency analysis.

Vandenberghe et al. (2012) describe some of the challenges that hydrological and hydraulic designers face when dealing with events characterized by multiple design variables. They discuss on ways of taking into account the dependence between the variables in the design, and how the joint return period should be defined and applied. They use copulas to construct multivariate distribution functions, arguing that they are able to model numerous types of dependence structures in a flexible way and highlighting their practicality in multivariate frequency analysis.

Goodarzi et al. (2011) made a risk analysis to assess the safety level of an existing dam in Iran. They made the overtopping risk analysis based on univariate and bivariate frequency analysis, considering peak flow and peak flow-volume respectively. Their results show the values of overtopping risk using the univariate analysis are lower than the values obtained using the bivariate frequency analysis. This statement is valid for the different values of initial water level considered in the study.

Besides the question of how many flood characteristics to include in the analysis, another important issue under discussion when designing or assessing the risk of an existing dam is whether to define a single flood event or to consider a range of possible events and decide which is more critical for the structure.

The traditional design of flood protection structures, like flood control reservoirs and polders, considers one single flood event with an associated return period. A broad range of different hydrological loads could be used to derive more efficient designs. When a multivariate analysis is followed, the selection of a single flood event reduces the amount of information that can be obtained from the approach.

A range of statistically similar design events, which are events that have the same return period, is derived from a multivariate approach. These flood events affect the structure in very different ways. Considering the case of a bivariate analysis, a return period is defined as design criteria of a structure, and several pairs of values of the variables under study are associated to that return period. These pairs of values are represented by an isoline corresponding to the selected return period. Out of the pairs of possible values, the points on the edges are less likely than pairs closer to the center of the isoline, and a most likely point of the isoline can be selected,. Following this criteria one pair of design values would be chosen. Another possibility is to ensemble a large number of pairs of values from the same isoline, i.e. associated with the same return period, and run simulations to assess the uncertainty of a specific design parameter.

As an example the range and likelihood of possible water heights in a reservoir associated to a given design return period could be assessed by routing a large number of pairs of flow peaks and volumes through the reservoir. Using only one design event, a single water height is obtained and the uncertainty cannot be assessed. Using a multivariate analysis makes it possible to incorporate uncertainty analysis in the design. The ensemble approach provides much more information on the possible outcome of design events.

A similar reasoning is presented in Shiau et al. (2006), who state that in bivariate frequency analysis, infinite many combinations of flood peak and volume result in the same return period, meaning that they all have the same likelihood of occurrence. However all these equal risk flood events may not produce the same magnitude of design safety factor, so further investigation is required in order to define the most critical case. The bivariate flood frequency analysis allows for a

better selection of the critical event related to a particular design criterion and offers a better assessment of risks, which cannot be achieved through a univariate frequency analysis.

The arguments presented and discussed in this chapter show the importance of studying and discussing whether a multivariate analysis is required in the design of dams for determining the spillway capacity and associated water levels, and for existing dams to assess the adequacy of spillways. As mentioned, the multivariate approach offers a range of possible events associated to a return period, which can be used in a design stage. The multivariate criterion includes information related with the correlation of the variables, and this additional data is involved in the generation of random samples of the variables of interest. These synthetic series are necessary for a risk assessment of existing dams.

1.2. Objective of the work

The design and assessment of hydraulic structures involves the delineation of flood profiles which requires the estimation of the discharge rates. This estimation can be carried out through different methods. In rivers with available gage records, the most commonly used procedure for the analysis of the flood data is the statistical frequency analysis.

The definition of hydrologic risk associated with many problems concerning reservoirs and flood protection works involve not only the peak flows but also the quantities of water (Bacchi et al., 1992). De Michele et al. (2005) states that although the peak of a flood is fundamental in assessing the hydrologic safety of dams and checking the adequacy of the dam spillway, the flood volume also has an important role in the definition of the spillway design flood and may influence significantly the hydrologic safety of the dam.

The mathematical theory of multivariate extremes is a relatively novel, but rapidly growing field (Salvadori et al., 2007b). An extreme event in a multidimensional context can be assimilated to the failure region in a structural design. A multivariate observation is extreme if it falls into some failure region which has a small probability of being reached. The multivariate frequency can be analyzed using classical multivariate models or based on copulas functions.

As mentioned by Chowdhary et al. (2011) the use of copula-based multivariate distributions for hydrological designs is in its initial stages. The approach using copulas is promising in comparison with the classical multivariate models since it takes into account a wide range of types of correlation which are frequently observed in hydrology. For instance, the dependence structure of the flood variables may differ from the Gaussian case, defined by the Pearson's coefficient of linear correlation. As was expressed by Shiau et al. (2006), the essential feature of the copulas is that a joint distribution of correlated random variables can be expressed as a function of the univariate marginal distributions. The advantage of using copulas to analyze multivariate problems is that the effects of the marginal distributions are separated to the effect of dependence. The description and modeling of the dependence structure between the random variables is independent of the marginal laws involved.

In this work, the risk of a flood event is assessed by considering the flood as multivariate event characterized by the following two variables: peak flow and volume. A bivariate statistical frequency analysis is carried out to find a suitable function to adequately represent the data set of flood peak flow and volume. The frequency is analyzed using copula models in order to study and discuss the advantages and limitation of this approach.

The application of the copula method involves the following steps: 1) determination of the marginal distributions of the two variables, using the conventional univariate statistical approach; 2) identification of an empirical joint distribution between the two variables; 3) identification of different copulas and estimation of their parameters; 4) determination of the joint probability distribution; and 5) application of goodness-of-fit statistics to compare the performance between the different joint distributions.

As mentioned in the previous section, the traditional approach for designing flood control reservoirs and polders considers only a single design flood with an associated return period. A sustainable design of flood protection structures requires considering a broad range of different hydrological scenarios, due to the fact that, as pointed by Klein et al. (2008), the natural variability of floods cannot be represented appropriately by a single design flood.

To show and discuss the advantages of the multivariate approach using copulas, the results are compared with values obtained using the univariate approach applied to the same variables independently. To incorporate uncertainty in the study, this work will focus on a method based in a range of possible flood events, instead of following a one-design-event methodology.

1.3. Structure of the work

This work is organized as follows. After the “Introduction”, the “Literature Review” section, gives a brief description of some of the publications related with the study of floods characterized by a set of random variables, in particular peak flow, volume and duration.

The next chapter “Materials and Methods” is divided into a set of sub-chapter describing the methods, equations and considerations taken into account in the different steps of calculation. It is arranged in a chronological way following the necessary steps to be fulfilled before going to the next calculation step. These steps involve identifying the flood events, selecting the samples of flood values that satisfy certain statistical criteria, analyzing the statistics of the samples, analyzing the frequency of the variables, generating synthetic hydrographs and routing the hydrographs through a reservoir.

In the chapter “Study Area and Data”, a brief description of the case study catchment is presented, as well as the available hydrological information used for the work. The last sub-chapter includes the description of a hypothetical reservoir assumed for the routing of the synthetic hydrographs.

The chapter “Results and Discussions” includes the results obtained in the different steps of calculations. The arrangement follows the chronology of the necessary calculations, as presented in the “Materials and Methods” chapter. The “Hydrograph Separation” includes the results regarding the base flow separation, identification of independent flood events, resulting series of variables characterizing the floods and the statistical relationship between these series. The “Extreme Value Statistics” sub-chapter describes the results of the univariate and bivariate frequency analysis as well as the resulting estimated variables associated to a selected return period. In the “Synthetic Hydrograph” sub-chapter the analysis of the observed hydrographs and generation of random shapes of hydrographs are presented. The last sub-section, “Reservoir Routing”, describes the maximum values of reservoir levels and outflows resulting from the routing of the synthetic hydrographs. The results are divided in two parts: the assessment of the risk of an existing dam and a proposed design methodology for a dam. The results obtained by the univariate and bivariate approaches are compared and discussed.

The conclusions and final remarks are given in the chapter called “Conclusions”, and some ideas regarding further studies are presented in the “Future Works” chapter. The next chapter is “Acknowledgments”.

All the details regarding the references mentioned in the work can be found in the “References” section. The last section of this work includes four appendixes, which are calculation flow charts describing the main steps of calculation followed in the different analysis. They are divided in the four different steps of calculation: frequency and return period analysis of the variables characterizing the flood; frequency analysis of the variables describing the hydrograph shape; synthesis of the flood variables and hydrographs and reservoir routing for the risk assessment case and the last one synthesis and reservoir routing for the design case.

2. LITERATURE REVIEW

There is an important number of works and publications involving multivariate statistical approaches regarding variables characterizing flood events. In the following paragraphs the contents of some of these publications are briefly described. Along this report, more publications are mentioned and discussed.

Klein et al. (2010) analyzed two reservoirs in Germany applying joint probability of the flood peak and corresponding flood volume. They argue that in flood protection structures the flood volume is very important and therefore its probability should be considered in the risk-based analysis in addition to the peak. The study corroborated that considering flood volume in addition to flood peak is important for risk-based planning and design.

Shiau (2003) studied the observed volume and peak flow series to characterize the extreme flood events in the river Pachang, in Taiwan. A bivariate extreme value distribution was used as a theoretical model to consider the joint effect of the two variables. The results show a good agreement between the theoretical models and the observed data providing more useful information than the univariate random variable frequency analysis.

Zhang & Singh (2006) used different flood characteristics of rivers in Canada and United States to study joint distributions based on the copula method. The flood peak-volume and volume-duration pairs of series were used. They compare the results of the copula method with other bivariate distributions (Normal and Gumbel mixed). The copula-based distribution was the best fit to the empirical bivariate distribution, this statement was derived from the graphical comparison and also from the values of root mean square errors, which were minimum for the copula case.

Vandenberghe et al. (2012) present an overview of the existing methods for defining joint return periods, focusing on how to calculate a design event for a given return period. They study and compare the following methods: regression analysis, bivariate conditional distributions, bivariate joint distributions, and Kendal distribution functions. A case study involving the definition of a design hydrograph is applied to evaluate the performance and differences between the methods. They conclude that for a given design return period, the method used to do the multivariate frequency analysis affects the resulting design event. When more variables are included in the analysis, it captures a more complex process generating extremes, so the design quantiles become smaller. Another important conclusion from this work is that the most likely event does not necessarily correspond to the most severe one for a structure, they state that selecting only one single event from a multivariate analysis is a drawback, since a full-ensemble-based design approach can give a lot more information of the process.

Shiau et al. (2006) used annual floods in a river in Taiwan, to analyze the bivariate frequency of flood peak and volume using copulas. They present results of probabilities associated to three extreme observed flood events. One event has the highest peak and highest volume. The second event has the second highest peak flow and the third flood case has the second highest volume. The first event has the highest univariate probabilities, as well as the highest joint probability among all the observations. In the second case, when the flood peak and volume are evaluated separately, the probability of the flood peak is very high, since it is the second highest peak, while the probability associated with the volume indicates that it is not a critical value. Different results are obtained when analyzing the third flood case. The probability of the flow peak indicates that it is not

a critical value, while the volume has a very high associated probability. When the flood control planning focuses on peak flows, the first two events should be considered as significant, but if the flood volume is also a safety factor in the design of flood mitigation facilities, ignoring the third flood event may lead to an underestimation. The bivariate evaluation gives the second and third event similar values of joint probabilities, indicating that the two events have similar magnitudes of risk when both peak flow and volume are considered as important factors in the flood control design.

Salvadori et al. (2011) discuss some issues of designing in a multivariate framework. They state that defining a design return period is not sufficient since the multivariate case generally fails to identify a unique realization to be used for design purposes. Additional considerations are required in order to select a characteristic realization over the critical layer of interest. This critical layer is defined as a surface for the trivariate case presented in their work, it would be an isoline for a bivariate case. They propose the introduction of a function that “weights” the realizations belonging to the critical layer. The function should be defined with a criterion that best fits the practical needs, and the design realization would be the one that maximizes the function. The problem with this approach is that in some cases a unique maximum cannot be found, and also different copulas have different geometries, so they can result in different design realizations. For this reason, they state that sometimes it is more appropriate to select a set of possible design realizations, rather than a single one, in order to evaluate the features of the phenomenon affecting the structure under design.

A very important statement regarding the correct use in practice of a multivariate approach is discussed by Salvadori et al. (2011). They say that in order to define a critical design realization, the stochastic dynamics of the phenomenon should be considered along with other information like the physical features of the structure under consideration or the environment in which it should be designed. As an example they mention the definition of a design storm characterized by its durations and intensity. A sewer system should be designed using critical storms of short durations and high intensities, because it is a fast responding structure. If a structure in the main river (catchment level) is under design, the critical situation corresponds to storms with long duration and low intensities. If the stochastic dynamics of the phenomena is the only consideration in the design, a typical critical storm associated to a given return period would be calculated, but this storm may not cause any problem in any of the two systems mentioned, and therefore they would be wrongly dimensioned.

3. MATERIALS AND METHODS

The materials and methods used in the different steps of analysis are explained in this chapter. A brief review and discussion of the existing methods is presented for each of the calculation steps, with a justification of the selected method. The sub-chapters are arranged following the sequence of calculations.

First the criterion of flood identification is presented, which includes the method of hydrograph separation between base and direct flow, and the identification of independent flood events. Then the statistical techniques applied to analyze the series of flow and volume are presented, to decide whether the series meet the statistical requirements to be used for a frequency analysis. The next sub-chapter includes the methods and criteria used to analyze the frequency of the series. The considered probability functions and methods of estimating the parameters are mentioned as well as the criteria of selection of the best model. The analysis for the univariate case is briefly explained, and more emphasis is put in the explanation of the copula model, applied to a bivariate case. This sub-chapter ends with an application of the frequency analysis with return period criteria, different ways of considering a joint return period are discussed. The next sub-chapter explains the criteria followed to generate synthetic hydrographs from the observed flood events. The last part of this chapter is about the reservoir routing, in which the considerations and criteria of calculation is described.

3.1. Flood identification

The most significant characteristics of a flood event are the flood peak, flood volume and flood duration (Yue, 2000). In order to define them, a start and end date of flood runoff must be identified. In some studies, the starting date is defined by an abrupt rise of the hydrograph, and the end date by the flattening of the recession limb of the hydrograph (see e.g. Yue, 2000; Klein et al., 2010; Shiau, 2006; Klein et al., 2011). Bergmann et al. (2000) estimate the direct runoff of a flood event by the separation of the base flow from the total flow. Vandenberghe et al. (2012), identify hydrographs from synthetic direct runoff series, as the continuous sequence of non-zero direct discharge values including the annual peak. Grimaldi & Serinaldi (2006) define the flood event duration in terms of time interval for which the discharge exceeds a fixed threshold value, and the volume is the area of the hydrograph above the given threshold.

In this work, the abrupt rise of the hydrograph and the flattening of the recession limb were difficult to identify for all the case years studied. The approach proposed by Grimaldi & Serinaldi (2006) was discarded due to the fact that their results show that the strength of the relationship between the flood characteristics depends on the threshold value used to identify the flood event. For this reason the approach presented by Bergmann et al. (2000) was applied, i.e. the event was defined in terms of the separation of the base flow and identification of direct flow. In this way the starting date of the event is defined by the day preceding the peak flow for which the direct flow is equal to zero. Then the direct flow increases, passes the peak flow and decreases until it reaches a value of zero again, and this date is defined as the end date of the event.

The base flow separation was carried out by an automated technique called recursive digital filter technique and is described in Nathan & McMahon (1990). The technique is set up in the Baseflow Program developed by Arnold et al. (1995) and Arnold & Allen (1999). The input data are the daily flow values, and the output data are the base flows. The base flows are estimated by filtering the

surface runoff from the total registered flow. The equations for estimating the base flow and surface runoff are presented here (see (1) and (2)).

$$b_t = Q_t - q_t \quad [m^3/s] \quad (1)$$

$$q_t = \beta q_{t-1} + \frac{(1 + \beta)}{2} \{Q_t - Q_{t-1}\} \quad [m^3/s] \quad (2)$$

where Q_t are the total observed flow values, b_t are the base flows and q_t are the direct flows. β is the filter parameter which does not have units.

According to Nathan & McMahon (1990), the filter parameters that yielded the most acceptable baseflow separation were in the range 0.9-0.95. The value of 0.925 was determined to Arnold et al. (1995) to give realistic results compared to manual separation. The same value was used by Arnold & Allen (1999) resulting in acceptable results compared with field estimates of baseflow. In this work the 0.925 value of filter parameter was used in the separation of the baseflow and surface runoff.

The selection of the maximal flood events is an important first step for a multivariate analysis, as discussed by Chowdhary et al. (2011). The design of hydraulic structures involves extreme flood events that are generally associated with peak flows that cause inundation of flood plains or the overtopping of crest dams or levees. As pointed by Chowdhary et al. (2011) the negative effects of high volume or duration of the flood event are also important, but they typically come into play only when there is a primary failure due to high peak flows. Based on these criteria, the annual maximum flood events used in this work have been selected in terms of annual peak flows and their associated volumes.

A peak flow for each year is identified. Rivers with two-wave-shape annual hydrographs might have two annual peak flows, one for each period of the year (rain event and snow melt event). Each peak flow has an associated direct runoff volume resulting from the baseflow separation. It is important to point that the volume associated to the peak flow should be caused by the peak flow and not by other high flows that might be independent from the peak flow event. The dependency between the flows identified as direct runoff must be tested, in order to define the actual volume of the peak flow. In case of independency the high flows can be isolated and separated into single events, and only the volume from to the peak single event is considered.

Subsequent peaks can be considered dependent if the time between the events is lower than a critical time, or if the lowest inter-event discharge is higher than a specific low flow level. Willems (2009) proposes the separation between peaks to be 1, 2 or 3 times the recession constant of the quick flow (depending on how strong one wants to have the independence between subsequent quick flow components), and the low peak limit to be 5%, 10% or 15% of the peak flow (depending on the accuracy of the baseflow filter results). Bacchi et al. (1992) proposed a limit time value of 20 times the time to peak, and a low flow value of 20% of the peak flow. Other criteria presented by LAWA (1997) or Klein (2009) give a limit time value of 1 week, and different low flow limit values. According to Lang et al. (1999) the US Water Resource Council (1976) imposes a limit time between peaks of 5 days plus the natural logarithm of the basin area (in square miles). Cunnane (1979) states the following criteria: separation between flood peaks of three times the average time to peak and the low flow limit value of two thirds of the first peak. The average time to peak is defined as the mean value of the times of several (at least 5) clean typical flood hydrographs in the record. The same criteria is adopted by the UK Flood Studies Report (Natural Environment

Research Council, 1975), and it is suitable for a wide range of catchments in the UK, in terms of size, flood nature, etc..

These different criteria were analyzed to define the suitability of applying them to the study catchment. The method stated by Cunnane (1979) was found to be more suitable, due to the fact that the time interval between peaks is defined based on the selection of “clean” typical flood hydrographs, i.e. hydrographs without any previous flooding. The time interval proposed by Bacchi et al. (1992) was found to be a very high value for the study catchment, especially when analyzing the summer (snow-melt) events, in which the time to peak can be in the range of 10 or more days. The method proposed by the US Water Resource Council (1976) defines the time interval only as a function of the catchment area, but this criteria is not suitable for our catchment since the winter and summer events have a very different behavior, especially regarding the time to peak, and the catchment size is the same during the occurrence of both flood events.

The last step in the hydrograph separation procedure is to test the dependency between the flows identified as direct runoff. In case of dependency, the total direct runoff volume is considered for the volume estimation. In case of independency only the single event corresponding to the peak flow is considered for the volume estimation.

All the hydrograph separation results were visually inspected.

3.2. Statistical analysis

Hydrological events observed in the past can be used to interpret the phenomena and to estimate future events in terms of probabilities of occurrence; this process is called frequency analysis. In order to consider the results of the frequency analysis as theoretically valid, the data series must meet certain statistical criteria such as randomness, independence, consistency, homogeneity and stationarity (Bobée & Ashkar, 1991). In this work the different criteria are checked using non-parametric tests, which avoid assumptions regarding distributions to describe the process.

The graphical method called Box & Whisker Plot is applied to describe the statistical characteristics and data of the series used here. The homogeneity criterion implies that all the elements of the data series originate from a single population. This criterion is studied applying the Double Sum analysis. Stationarity means that the data series is invariant with respect to time, i.e. there are no trends, jumps or cycles. The Mann-Kendall test is applied to study possible trends in the series, and the Pettitt test to identify points of change. To detect outliers the Grubbs and Beck test is used. In the following sub-sections a brief description of these methods is presented. For further information please refer to the recommended bibliography.

3.2.1. Box & Whisker plot

The Box & Whisker Plot is a graphical way to display the distribution of a data set. It provides a visual summary that includes statistical information of the data set including: central tendency, variability, symmetry, range and presence of outliers. These plots are useful for examining the characteristics of a single data set and for comparing several data sets. In the study, this tool is applied in order to visualize the properties of every data set and to compare the characteristics between the summer and winter flood series. For more information regarding the interpretation of the Box Plot refer to Maidment, (1993).

3.2.2. Double Sum analysis

The Double Sum analysis is a simple graphical procedure for the assessment of inhomogeneous behavior of the data set. This analysis relies on a reference data set to be compared with the tested one. The reference data set is assumed to be consistent and homogeneous, and to have a linear relationship with the tested data set. Long simultaneous observations of both data sets must be available. Each data set is accumulated and then the two series are plotted according to the following equation (see (3)).

$$\sum_{t=1}^N X(t) \quad \text{vs.} \quad \sum_{t=1}^N Y(t) \quad (3)$$

The $X(t)$ are the values of the tested data set and the $Y(t)$ are the corresponding values belonging to the reference data set. N is the total length of the data sets.

In this work the tested data sets are the peak flow series and direct volume series. The reference data used to assess the homogeneity of the peak flow series is the mean flow of the corresponding period, i.e. April-September for the winter events, or October-March for the spring events. For the volume series, the total direct volume of the corresponding period was used as reference data.

3.2.3. Mann-Kendall test

The Mann-Kendall test is a non-parametric test used to identify possible trends of hydrological time series. It is a rank-based test used to test for a correlation between the ranks of observations and their time order. The test has the advantage that it avoids assuming the behavior of the tendency (linear or nonlinear). This test has been applied in several works to study the trend of hydro-meteorological variables. (Hirsch et al., 1982; Lettenmaier et al., 1994; Westmacott & Burn, 1997; Yue et al., 2003; Cunderlik & Burn, 2004; Hamed 2009).

The test assumes that the elements of the data serie (x_1, \dots, x_n) are independently identically distributed random variables. The null (H_0) and alternative (H_A) hypothesis of the test are the following:

H_0 : The data (x_1, \dots, x_n) are independently and identically distributed random variables.

H_A : The distributions to which x_j y x_k belong are not identical for every k and j with $k \neq j$ and $k, j < n$.

Rejecting H_0 means accepting a trend, while not rejecting it means there is no significant trend. The test statistic S is given by the following equation (see (4)).

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k) \quad (4)$$

where $\text{sgn}(x_j - x_k)$ takes the value of either 1, 0, or -1, if the difference $(x_j - x_k)$ is greater, equals or lower than 0, respectively. Under the null hypothesis, the test statistic S behaves following a Standard Normal distribution, with media and variance presented in the following equations (see the Equation (5)).

$$E(S) = 0$$

$$Var(S) = \frac{N(N-1)(2N+5) - \sum_t e_t(e_t-1)(2e_t+5)}{18} \quad (5)$$

where N is the sample size, t is the number of tied groups and e_t is the number of data in t -th (tied) group. The test statistic S can be associated with the normally standardized variable Z , with the following equations (see (6)).

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & si \ S > 0 \\ 0 & si \ S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & si \ S < 0 \end{cases} \quad (6)$$

In order to reject or not H_0 , the value of $|Z|$ must be compared with a critical value. This critical value ($z_{\alpha/2}$) is defined according to a significance level of α . H_0 is rejected if $|Z| > z_{\alpha/2}$. Otherwise H_0 is accepted, i.e. the trend is not significant for a level of α . The significance level used for this work is of 5%, and the corresponding critical value is 1.96.

3.2.4. Pettitt test

The Pettitt test is a non-parametric rank-based test used to identify a change in the serie (Pettitt, 1979). The advantage of this test is that the exact time of change does not need to be known. The test is considered to be robust to changes in distributional form and relatively powerful. This test has being applied in several works to study the trend of hydro-meteorological variables (Fealy & Sweeney, 2005; Seoane & Callau, 2009; Descroix et al., 2011).

The test assumes that the sequence of random variables (x_1, \dots, x_n) has a change point in r , if (x_1, \dots, x_r) have a common distribution function and (x_{r+1}, \dots, x_n) have another common distribution function different from the one belonging to the first sub-serie. The null (H_0) and alternative (H_A) hypothesis of the test are the following:

H_0 : The data (x_1, \dots, x_n) belongs to one distribution function, i.e. $r=n$.

H_A : There is a change point within the data so that $1 \leq r \leq n$.

Rejecting H_0 means accepting a change point, while not rejecting it means there is no significant change in the serie. The Pettitt test considers a time series as two samples represented by $x_1 \dots x_t$ and $x_{t+1} \dots x_n$. The test statistic K_n is defined as the most significant change point, and is found where the value $|U_{i,n}|$ is maximum, given by the following formulas (see (7)).

$$K_n = \max_{1 \leq i \leq n} |U_{i,n}|$$

$$U_{i,n} = \sum_{j=1}^i \sum_{j=i+1}^n \text{sgn}(x_i - x_j) \quad (7)$$

The approximate significance probability of a change point is estimated with the following formula (see (8)).

$$p(t) = 1 - \exp\left(\frac{-6K_n^2}{n^3 + n^2}\right) \quad (8)$$

where n is the sample size. The significance level used for this work is of 5%, meaning that to accept a change point, the significance probability has to be at least 95%.

3.2.5. Grubbs and Beck test

This test is applied to analyze the extreme values identified by the Box plot method. The test identifies low or high outliers by comparing the sample values with two different limit values estimated with the following equations (see (9) and (10)).

$$X_H = \exp(\bar{x} + K_N * S) \quad (9)$$

$$X_L = \exp(\bar{x} - K_N * S) \quad (10)$$

where \bar{x} and S are the mean and standard deviation of the natural logarithms of the sample to be tested. K_N is the test statistic and for a 10% of significance level has the polynomial approximation presented in the following equation (see (11)).

$$K_N \cong -3.62201 + 6.28446(N)^{1/4} - 2.49835(N)^{1/2} + 0.491436(N)^{3/4} - 0.037911N \quad (11)$$

where N is the sample size. Any sample values greater than X_H are considered high outliers, and those lower than X_L are low outliers. In this work the extreme values identified with the Box Plot were tested to assess whether they were outliers or just extreme values. For further information and applications of this test refer to Bobée & Ashkar (1991).

3.3. Frequency analysis

The objective of hydrological frequency analysis is to interpret past records of hydrological events in terms of future probabilities of occurrence. An estimation of unobserved flood events is conducted based on the observed flood records. Design floods are hypothetical or typical events that represent rare occurrences, expressed by their probabilities or return period as the degree of “rareness”.

Frequency Analysis is the process of using the sample information to identify the appropriate population. The population is a mathematical model that consists of a probability distribution function and its estimated parameters, and represents the relationship between the random variable and its likelihood of occurrence. The assumed population is used to estimate the probabilities and magnitudes. Frequency Synthesis refers to the use of the population for the purposes of estimation of either, a value of the random variable for some selected exceedence probability, or the exceedence probability for a selected value of the random variable. The exceedence probability is the probability that an event having a specified magnitude will be exceeded in one time period, which is most often assumed to be one year. Another term widely used is the return period which is defined as the average length of time between events having the same magnitude, and is the inverse of the exceedence probability. For further information refer to McCuen (1998).

A univariate frequency analysis is applicable when only one random variable is significant in the design process. If a given event is multivariate, i.e. described by a set of random variables, the

univariate frequency analysis cannot give a complete assessment of the probability of occurrence (Salvadori et al., 2007b). In this case a multivariate frequency analysis can provide a probabilistic assessment of the occurrence of critical events, enabling multivariate risk-based designs.

When a Copula method is applied to model a multivariate data set, the approach consists of the two following steps: 1) estimation of the marginal distributions of the data sets separately, and 2) estimation of the dependence function. This enables the derivation of multivariate probability distributions regardless of their dependence structure.

In this work marginal distributions are fitted to the different analyzed variables (flood peak and volume). The univariate frequency analysis and synthesis is developed using these marginal distributions. Then the dependence between the variables is analyzed and a Copula model is estimated to develop a bivariate frequency analysis and synthesis. The results from the univariate and bivariate approaches are used to compare the two criteria.

All the estimations for univariate and Copula analysis and synthesis presented in this work have been done using R (R Development Core Team, 2012) a free software environment for statistical computing. The packages used are the Fitdistrplus package (see Delignette-Muller et al., 2010), the Lmomco package (see Asquith, W., 2010) and the Copula package (see Yan., 2007; Kojadinovic & Yan, 2010; Hofert & Maechler, 2011).

3.3.1. Univariate frequency analysis

Frequency modeling is a statistical method classified as univariate when it deals with a single random variable. The goal of univariate prediction is to make estimates of either probabilities or magnitudes of one random variable.

Several possible distributions are considered for fitting the series of peak flow and direct volume. These distributions are: Normal, Exponential (2 and 3 parameters), Gamma, Weibull (2 and 3 parameters), Log-Normal (2 and 3 parameters), Generalized Extreme Value (GEV), Pearson, Gumbel and Generalized Pareto. The methods of estimating the parameters used in this work are the following: Method of Moments, L-Moments and Maximum Likelihood. The goodness of fit of the different models is assessed using the Kolmogorov-Smirnov (KS) and Cramer-Von-Misses (CvM) criteria. Further information regarding these methods can be found in any book of hydrological statistics (see Maidment, 1993).

The KS and CvM statistics are estimated for the different models. The models are grouped as rejected and not rejected with a 5% of confidence level, according to the limit values presented by D'Agostino & Stephens (1986) and the values resulting from the Fitdistrplus package. For each of the analyzed series, a model is selected from the not rejected group by an inspection of the quantile plots (QQ-Plot), i.e. by comparing the quantiles of the data set and theoretical model. The model in which the theoretical quantiles best represent the sample quantiles is chosen.

The same methodology is followed to analyze the frequency of the series of parameters (a and b belonging to the Beta distribution) defined to characterize the shape of the hydrograph (see sub-chapter 3.4).

3.3.2. Multivariate frequency analysis using copulas

In a multidimensional analysis, the dependence between the different variables plays a fundamental role. Quantifying this dependence is a central theme in probabilistic and statistical methods for multivariate extreme values. As mentioned at the beginning of this work (see sub-chapter 1.1), many publications show that there is a significant positive dependence between the flood peak and volume, and hence they require a joint analysis.

As stated by Genest & Favre (2007a), the classical families of bivariate distributions, such as Normal, Log-Normal, Exponential, Gamma and Extreme-Value, have been traditionally used to study hydrological variables. The main limitation of this approach is that the two variables must be characterized by the same parametric family of univariate distributions. This restriction is avoided by the copula models, which are just beginning to make their way into hydrological applications. In this work a copula model is applied to study the joint behavior of the flood characteristics. This section contains some basic theory and the methods used to apply a copula model.

As explained by Klein et al. (2011) a joint distribution of correlated random variables can be expressed as a function of the univariate marginal distributions using a copula. A copula, denoted as $C(u_1, u_2, \dots, u_n)$, is a function that enables to model the dependency structure between the random variables, independently of their marginal distributions. The link between the copula and the multivariate distribution is provided by the theorem of Sklar with the following equation (see (12)).

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = C[F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)] \quad (12)$$

where $F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ is the joint cumulative distribution function with the continuous marginal distribution functions of the random variables: $F_{X_1}(x_1), F_{X_2}(x_2), \dots, F_{X_n}(x_n)$. The advantage of this approach is that the dependence between the variables under consideration can be appropriately modeled, independently of the choice of the marginal distributions.

The focus of the present work is to describe a flood using two characteristics, so the following descriptions and formulas will be based on the bivariate case, that is, two correlated random variables.

A copula is a joint distribution expressed in terms of univariate marginal distribution functions, so the following properties are verified (see (13) and (14)).

$$C(u, 0) = C(0, v) = 0 \quad (13)$$

$$C(u, 1) = u \text{ and } C(1, v) = v \quad (14)$$

where u and v denote two dependent cumulative distribution functions ($F_X(x)$ and $F_Y(y)$), and range between 0 and 1. The first property indicates that if either one of the marginal functions is 0, then the joint distribution will also be 0. The second property means that if either one of the marginal is 1, the joint distribution will behave as a univariate distribution of the opposite variable.

The empirical copula is derived from the random sample. In case of a bivariate sample with n pairs of values $(x_1, y_1), \dots, (x_n, y_n)$ the empirical copula is defined by the following formula (see Equation (15)).

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1\left(\frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v\right) \quad (15)$$

where R_i and S_i are the ranks of x_i and y_i of the sample belonging to (x_1, \dots, x_n) and (y_1, \dots, y_n) respectively, and $1(A)$ is the indicator function of set A . C_n is the best sample-based representation of the copula C . For any given pair of (u, v) , $C_n(u, v)$ is a rank-based estimation of the unknown quantity $C(u, v)$ (Genest & Favre, 2007a).

The copula is a theoretical characterization of the dependence in a pair (X, Y) , and is represented by the sample-based empirical copula. This dependence can be quantified using the two well-known nonparametric measures: Spearman's Rho and Kendall's tau, which are far superior to the classical Pearson's classical correlation coefficient, according to Genest & Favre (2007a). The Spearman's Rho coefficient is estimated with the formula (see (16)).

$$\rho_n = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^n R_i S_i - 3 \frac{n+1}{n-1} \quad (16)$$

where (R_i, S_i) are the pairs of ranks, and n the sample size.

The Kendall's tau is estimated with the formula (see (17)).

$$\tau_n = \frac{4}{n(n-1)} P_n - 1 \quad (17)$$

where P_n is the number of concordant pairs. Two pairs (X_i, Y_i) , (X_j, Y_j) are said to be concordant when $(X_i - X_j)(Y_i - Y_j) > 0$. Both estimators can be used to test for the independence between the two variables analyzed.

Chowdhary et al. (2011) argue that the fundamental objective of a copula model is to adequately represent the dependence structure of the data under consideration. The authors present a Copula test space based on admissible dependence range, for different copula models. Michiels & De Schepper (2008) developed a copula test space model that minimizes the fitting errors. They argue that an efficient copula test space can be constructed by taking into account information about the dependence. They present a restructuring of the copula space in terms of Kendall's tau, in which 29 different 2-dimensional copulas are analyzed and their applicability for different degrees of dependence is given.

Favre et al. (2004), Shiau et al. (2006), Genest et al. (2007b), Genest & Favre (2007a), Durante & Salvadori (2008), Klein et al. (2010), Wang et al. (2009), Chowdhary et al. (2011) and Klein et al. (2011) summarize several types of Copulas, and apply them to analyze multivariate hydrological data. The copulas employed to investigate the relationship between flood peak and volume are presented in the following table (see Table 3.1). These copulas are admissible in the whole range of positive dependence between the variables (see Michiels & Schepper (2008)), so they are suitable for modeling the dependence between flood peak and volume, as they show a positive dependence.

Table 3.1: Summary of 2-dimensional copulas: family, probability functions, parameter space and relationship of non-parametric dependence measure with parameter.

Copula	Family	$C_\theta(u,v)$	Parameter	τ
Normal*	Meta-Elliptical	$\Phi_\theta(\Phi(u)^{-1}, \Phi(v)^{-1})$	$-1 \leq \theta \leq 1$	$\frac{2}{\pi} \arcsin(\theta)$
Frank**	Archimedean	$-\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1) * (e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right]$	$\theta \neq 0$	$1 - \frac{4}{\theta} (1 - D_1(\theta))$
Gumbel***		$\exp \left[-\{(-\ln(u))^\theta + (-\ln(v))^\theta\}^{1/\theta} \right]$	$\theta \geq 1$	$\frac{\theta - 1}{\theta}$
Clayton		$(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$	$\theta \geq 1, \theta \neq 0$	$\frac{\theta}{\theta + 2}$
Galambos	Extreme Value	$uv \exp \left[\{(-\ln(u))^{-\theta} + (-\ln(v))^{-\theta}\}^{-1/\theta} \right]$	$\theta \geq 0$	n.a.
Hüsler-Reiss*		$\exp \left[\ln(u) \Phi \left(\frac{1}{\theta} + \frac{1}{2} \theta \ln \left(\frac{\ln u}{\ln v} \right) \right) + \ln(v) \Phi \left(\frac{1}{\theta} + \frac{1}{2} \theta \ln \left(\frac{\ln v}{\ln u} \right) \right) \right]$	$\theta \geq 0$	n.a.

* Φ stands for the cumulative distribution function of the standard Normal.

** $D_k(x)$ is the Deybe function.

*** The Gumbel Copula belongs to both archimedean and extreme value families.

As was stated by Chowdhary et al. (2011), the fundamental objective of a copula model is to adequately represent the dependence of the observed data. Another important aspect that a copula model must meet is to ensure suitability in terms of tail dependence characteristics. It is mentioned in their work that certain copulas may exhibit similar overall dependence features while showing different lower and/or upper tail dependence characteristics. The tail dependence of the different analyzed copula models are visually compared with the behavior of the observed data to decide which model is more suitable. Typical shapes of some of the copula models used in this work are presented in the following graphs (See Figures 3.1 and 3.2).

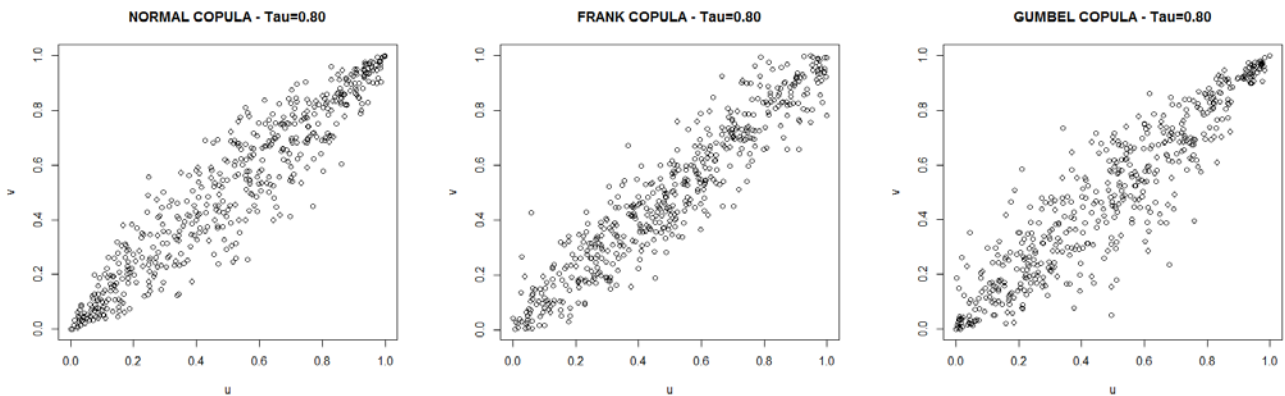


Figure 3.1: Comparison between simulated samples extracted from different 2-dimensional copulas (Normal, Frank and Gumbel), for Kendall's tau of 0.80.

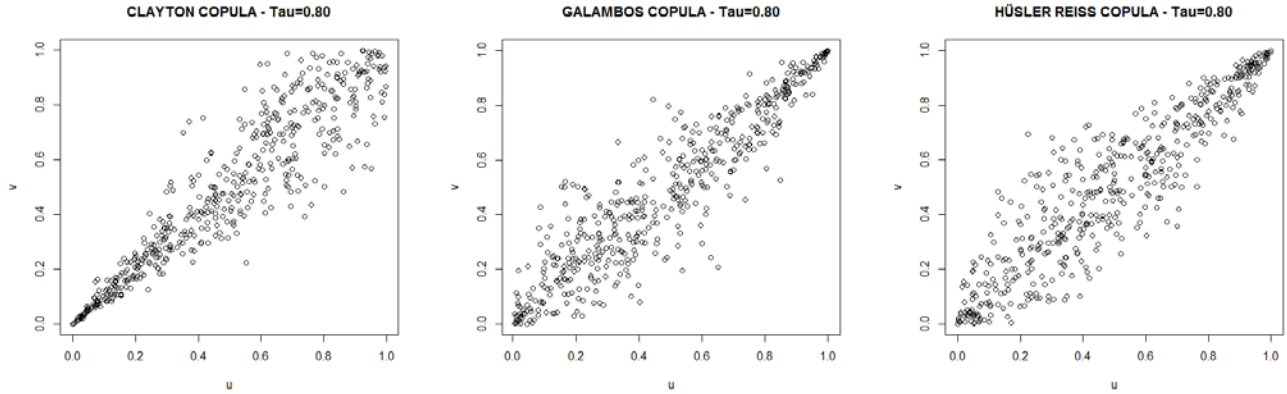


Figure 3.2: Comparison between simulated samples extracted from different 2-dimensional copulas (Clayton, Galambos and Hüsler Reiss), for Kendall's tau of 0.80.

The different behavior among the used copula models, regarding the upper (upper right corner) and lower tails (lower left corner) can be appreciated from the above figures. Note that all the samples are generated using the same criteria of overall dependence, but the different models show a different behavior among the samples in both tails. In the Normal copula the points tend to dispose themselves along the main diagonal, but are more concentrated in both tails. This means that the relationship between low-low and high-high pairs of values is closer to a line than the values that are in the middle range. The Gumbel, Galambos and Hüsler Reiss show a similar behavior, but the relationship between both variables in the lower tail is sparser, compared with the upper tail's pairs. The Frank copula shows a constant dispersion of the pair's respect the main diagonal for all the range of values. The Clayton copula has a different behavior, the lower tail shows concentrated pair of points, whereas in the upper tail the points are sparser.

It was mentioned that a copula model must adequately represent the dependence between variables. The following graphs show different simulations extracted from Clayton copula (See Figure 3.3). The simulations correspond to samples with Kendall's tau of 0.65 (moderate association) and 0.80 (strong association). In both cases the points tend to disperse along the main diagonal, with a stronger concentration in the lower tail. As the Kendall's tau increases, the points get more concentrated along the line.

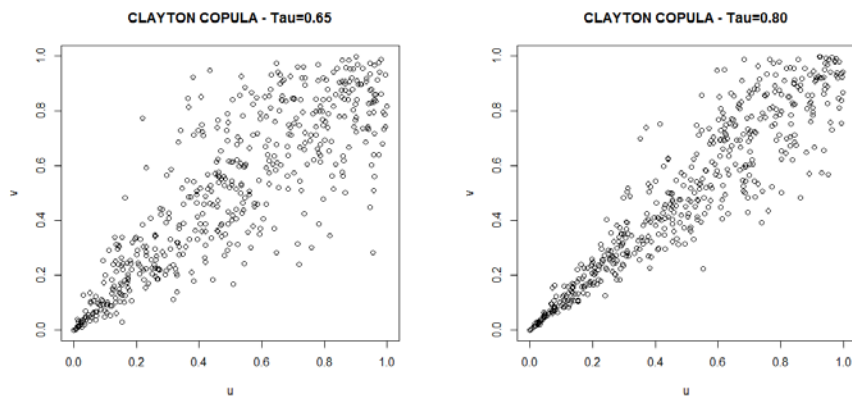


Figure 3.3: Comparison between simulated samples extracted from Clayton's 2-dimensional copulas, for Kendall's tau of 0.65 (left) and 0.80 (right).

The copulas can depend on one or more parameters (θ) that must be estimated. There are different methods of estimation. The most common ones are rank-based estimators that rely completely on the relative ranks of joint variates, determining the dependence structure independently of the choice of marginals. They are called moment-like method, based on the inversion of Spearman's Rho or Kendall's tau, and canonical or maximum pseudo-likelihood method. Other existing methods are: exact maximum likelihood approach, sequential-two step maximum likelihood method, Bayesian approach and inference function for margins method (refer to Shiau et al., 2006; and Wang et al., 2009). The methods used for this work are the moment-like and pseudo-likelihood method. This choice is justified by Genest & Favre (2007a), who argue that as the copula captures the dependence structure independently of the individual behavior of the variables, the parameter estimator should rely only on the ranks of the observations, which are best summary of the joint behavior of the random pairs. These rank-based estimators are explained in the following paragraphs.

The moment-like method based on the inversion of the Kendall's tau, relies on the idea that the bivariate dependence structure is fully defined by the relative ranks of the variables. The non-parametric estimates of θ based on Kendall's tau are obtained from the following relationship (see (18)) reported by Chowdhary et al. (2011).

$$\tau = 4 \iint_{[0,1]} C_{\theta}(u, v) c_{\theta}(u, v) du dv - 1 \quad (18)$$

where $c_{\theta}(u, v)$ is the copula density distribution, which is presented in Klein et al. (2011) as $c(u, v) = d^2 C(u, v) / du dv$. For some copulas, the relationship between the associated parameter θ with the Kendall's tau is available in closed form, as presented in the last column of Table 3.1. For the cases for which closed forms are not forthcoming, the parameter to estimate is related with the Kendall's tau using numerical integration. A limitation of this method is that it can only be applied for single-parameter copula families.

The maximum pseudo-likelihood method estimates the parameter by maximizing a rank-based log-likelihood function in the form presented here (see (19)).

$$l(\theta) = \sum_{i=1}^n \log \left[c_{\theta} \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right] \quad (19)$$

where R_i and S_i are the ranks of x_i and y_i of the sample belonging to (x_1, \dots, x_n) and (y_1, \dots, y_n) respectively, and $c_{\theta}(u, v)$ is the copula density distribution.

These estimation methods have the advantage that they do not depend on the marginal distribution functions, so the estimated parameter are not affected by the choice of the marginal model.

It was mentioned that the application of the copula in the science of hydrology is beginning to make its way. For this reason, there is yet no generic and widely accepted criteria to verify the validity and select an appropriate copula among various, such as the goodness-of-fit test for the univariate distributions. Wang et al. (2009) present 3 methods: Akaike information criterion (AIC), Quadratic distance criterion, and Genest and Rivest method. They state that the AIC is used more often than the other two due to its simplicity. Chowdhary et al. (2011) list other comparison criteria to quantitative asses the performance among various copula models. A comparison the between empirical and computed probabilities can be reflected by comparing error statistics such as: mean

square error, mean absolute error, mean error or maximum absolute error. They apply the Cramer-von Mises type of test statistic, which is also recommended by Genest & Favre (2007a). This method is based on the comparison of empirical and parametric copula probabilities given by the process $\sqrt{n}(C_n - C_\theta)$ and can be estimated using the following formula (see (20)). The Cramer-von Mises criterion is applied in this work.

$$CM_n = n \sum_{i=1}^n \left[c_n \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) - c_\theta \left(\frac{R_i}{n+1}, \frac{S_i}{n+1} \right) \right]^2 \quad (20)$$

Large values of this statistic can lead to the rejection of the copula under consideration. P-values associated to the statistic can be deduced from their limiting distributions, which depend on the asymptotic behavior of the analyzed process. These limiting distributions depend on the family of copulas under consideration and on the unknown parameter values. As a result these asymptotic distributions cannot be tabulated and P-values for the test statistic is approximated either using the parametric bootstrap (see Genest et al.; 2009) or by means of a fast multiplier approach (see Kojadinovic & Yan; 2010). The testing procedure based on multiplier central limit theorems is applied in this work, which is faster, thus requiring less computational time, as explained by Kojadinovic (2009).

A primary and direct application of the copulas is the simulation of dependent variables (Salvadori et al., 2007b). In this work, bivariate dependent variables are simulated using the Copula package. This package provides a random number generator that follows a procedure of iterative conditioning. Fast algorithms exist for some commonly used Archimedean copulas. More details regarding this issue can be found in Yan (2007). The randomly generated numbers are realizations corresponding to a standard uniform random variable (between 0 and 1). These numbers are then transformed into random variables by using the inverse of the marginal distribution corresponding to the variable under consideration.

Like in the univariate case, the methodology presented in this section is followed to analyze the joint frequency of the series of parameters (a and b belonging to the Beta distribution) that characterize the shape of the hydrograph (see sub-chapter 3.4).

3.3.3. Return period

A common criterion employed to design hydraulic structures or water supply systems is the return period of the extreme hydrological events. The notion of a return period to quantify design variables is a standard procedure, because of its simple and efficient description of risks. In a multivariate frequency analysis an event with a given return period is defined by studying the variables jointly. The univariate analysis is no longer satisfactory when dealing with events characterized by multiple variables, because it may lead to an overestimation or underestimation of the risk associated with a given event.

The return period for an event described by a single random variable and based on the annual maximum series, is defined as the average recurrence interval between events equaling or exceeding a specific magnitude, and is given by following formula (see (21)).

$$E(T_X) = \frac{1}{P(X \geq X_T)} = \frac{1}{1 - P(X < X_T)} \quad (21)$$

where $P(X \geq X_T) = 1 - P(X < X_T)$ is the probability of an event with a magnitude equaling or exceeding the specific magnitude X_T . If $F(x)$ is the cumulative distribution of X , then $F(x) = P(X < x)$.

The univariate frequency analysis is useful when only one random variable of the extreme event is significant for the design criterion. It can also be applied to different random variables separately when these variables are not significantly correlated. However if the correlation between the random variables characterizing the extreme event is an important information for the design criterion, the variables should not be analyzed separately.

The practitioners involved in the design of hydraulic structures face several challenges when dealing with multiple variables to define a design hydrograph. Vandenberghe et al. (2012) discuss some of these challenges, especially dealing with how to incorporate the dependence between the variables in the design, and how the joint return period should be defined and applied. They analyze and compare different joint return period methods based on regression analysis, bivariate conditional distributions, bivariate joint distribution, Kendal distribution functions and an ensemble-based method, highlighting the theoretical and practical issues of each method. They conclude that, given a design return period, the method chosen to define the flood event clearly affects the results.

In this section the concept of return period for events defined by the joint behavior of pairs of random variables is discussed. Some of the joint return period methods are explained, with a final remark of the method chosen based on theoretical and practical issues and the recommendations of the consulted bibliography.

Several definitions of joint return periods are possible, each dealing with particular occurrences of the dynamics of the considered event. In case of events characterized by the joint behavior of a pair of random variables, they can be expressed by four different marginal events, which are characterized by either the exceedence or non-exceedence of each of the two random variables. These 4 possible marginal events can then be combined in different ways, using an “OR” or “AND” operator.

In practice, an event is defined as dangerous if either one of the two random variables exceeds a given threshold (OR) or if both variables are larger than some prescribed values (AND). Salvadori & De Michele (2004) mention the example of dangerous events as a storm that has either a large intensity of a large duration (OR), and a flood with high volume and high peak (AND) which may be critical for a dam.

When considering the “primary” return period, at least two different design events can be defined (in the bivariate case). Having a return period value fixed, an event can be considered as critical if either one of the two variables (or both) exceed the given thresholds, this is the “OR” case. But an event can also be considered as critical if both variables are larger than the critical values, this is the “AND” case in which it is necessary that both conditions are fulfilled. The “primary” return periods for an event described by two random variables and based on the annual maximum series, are given by the following formulas (see (22) and (23)).

$$T_{\text{OR}} = E(T_{X_{Y_{\text{or}}}}) = \frac{1}{P(X \geq X_T \text{ or } Y \geq Y_T)} = \frac{1}{1 - F(X_T, Y_T)} \quad (22)$$

$$T_{\text{AND}} = E(T_{X_{Y_{\text{and}}}}) = \frac{1}{P(X \geq X_T \text{ and } Y \geq Y_T)} = \frac{1}{1 - F(X_T) - F(Y_T) + F(X_T, Y_T)} \quad (23)$$

Where X_T and Y_T are specific magnitudes to be equaled or exceeded. $F(x)$ and $F(y)$ are the cumulative distributions of X and Y respectively, and $F(x,y)$ is the joint cumulative distribution represented by a copula model.

Various combinations of the random variables can result on the same joint return period or joint probability, so a simple presentation of the combinations is with contours of equal return period or probability. When plotting the contour lines for different specific return period years, it can be noted that the “AND” and “OR” case exhibit different characteristics. The contour lines present segments in which they are either horizontal or vertical. In these segments the return periods are defined by one of the two variables solely, the variable X for the horizontal case, since Y is constant, and the Y for the vertical. In the following graphs, these two different cases of return period are illustrated (Figure 3.4). They correspond to a Clayton copula, and three selected return periods are plotted (100, 500 and 1000 years). Note that the values of the domains are sub-portions of the total space, in order to provide a clearer picture. The total space is between 0 and 1, and the u axis represents the value of the marginal distribution of X , and v the marginal distribution of the variable Y .

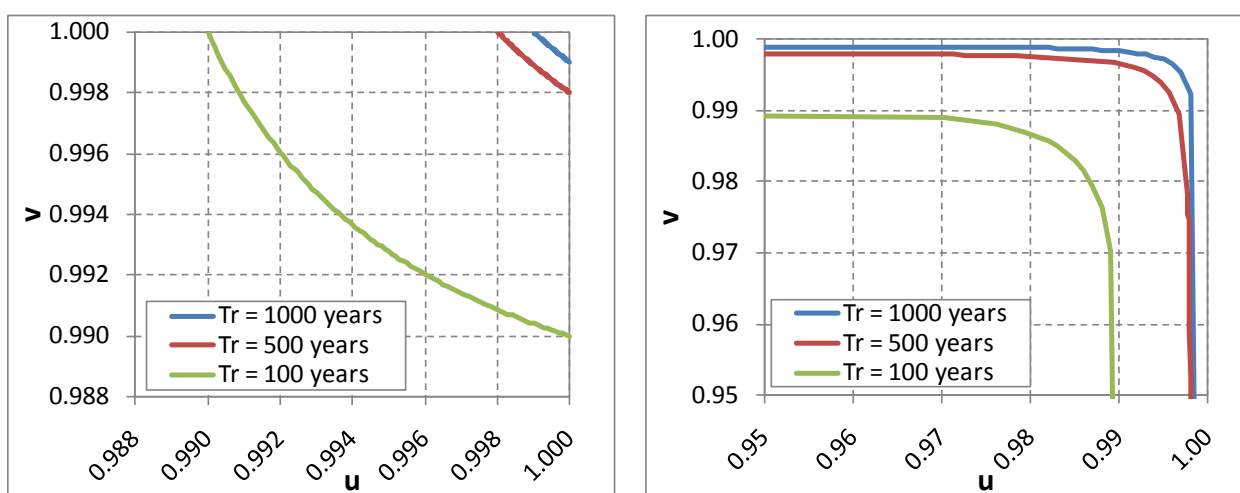


Figure 3.4: Joint return periods for a Clayton's 2-dimensional copula. Left: “OR” case; Right: “AND” case.

If a return period T is fixed, the expected “OR” case appears more frequently and the expected “AND” case appears less frequently compared with the expected value associated with the T , and the following inequality is fulfilled $T_{\text{“OR”}} < T < T_{\text{“AND”}}$ (for further details refer to De Michele et al. (2005), Shiau (2003), Shiau et al. (2006)).

This statement has an important implication for the design of hydraulic structures. For example a pair of values of peak flow and volume is considered for the design. If these values are compared to the “OR” case analysis as critical event, the result would be that the work is at risk because the structure is underdimensioned. If the same values are compared to the “AND” case, then the result would be that the work is overdimensioned. In other words, if a design return period is fixed, the “OR” case would result in paired values of peak flow and volume that are higher than using the “AND” case.

Goodarzi et al. (2011) used the “OR” type of return periods to assess the safety level of an existing dam in Iran, as they argue that the dam can be at risk either if the peak discharge is too high or the flood volume is too large. Klein et al. (2008) argue that the “OR” or “AND” joint return period is relevant depending on the hydrological design problem. Shiau (2003) states that the design criterion should be chosen according to the situations that would destroy the structure. The “OR” case should

be used in cases in which either the flow peak or volume will cause a damage if they exceed a certain magnitude. The “AND” case is used when both flood volume and peak must exceed certain magnitudes in order to cause damage.

Another concept of joint return periods, when considering events characterized by two random variables, is defined by considering the conditional return period of one of the variables with the other one given. In this case, the return period is defined as a conditional return period for one variable given a certain threshold value of the second variable. The definition of these return periods for an event described by two random variables and based on the annual maximum series, are given by the following formulas (see (24) and (25)). Further details can be found in Shiau (2003).

$$E(T_{X/Y \geq Y_T}) = \frac{1}{P(X \geq X_T, Y \geq Y_T)} = \frac{1}{[1 - F(Y_T)][1 - F(X_T) - F(Y_T) + F(X_T, Y_T)]} \quad (24)$$

$$E(T_{Y/X \geq X_T}) = \frac{1}{P(X \geq X_T, Y \geq Y_T)} = \frac{1}{[1 - F(X_T)][1 - F(X_T) - F(Y_T) + F(X_T, Y_T)]} \quad (25)$$

There is an additional case to be considered that is derived from a bivariate conditional distribution. In this case one of the design variables is defined by the univariate frequency analysis and this design value is then conditioning the bivariate distribution function of the second variable. The resulting conditional distribution function is an univariate one and is used to define the design value of the second variable. Following this criteria, the resulting return periods for an event described by two random variables and based on the annual maximum series are given by following formulas (see (26) and (27)).

$$E(T_{X=X_T}) = \frac{1}{P(X \geq X_T)} = \frac{1}{1 - F(X_T)} \quad (26)$$

$$E(T_{Y/X=X_T}) = \frac{1}{1 - F_{Y/X}(Y / X = X_T)} \quad (27)$$

It should be noted that this analysis should not be considered as a real joint design criteria, because the resulting design event is not a real bivariate case with an associated return period. The bivariate distribution is conditioned to a design value that resulted from an univariate analysis. Using the bivariate conditioned function for obtaining the second variable is again based on the principle of univariate return period. For further details refer to Vandenberghe et al. (2012).

To analyze the flood frequency in two rivers in North America, Zhang & Singh (2006) use bivariate distributions to define conditional joint return periods. Since the flood variables are positively correlated, their results show that a high specified value, for example peak flow, under a low conditioning value, for example volume, is less likely to occur than a high specified value under a high conditioning volume. They also found out that the calculated return periods vary depending on the criteria of the conditioning term. The return period of a specified value (e.g. flow peak) is lower if the conditioning is a fixed term (e.g. flood volume) than if the conditioning is in terms of a value less than or equal to the fixed value.

Another concept of bivariate joint return period called Secondary or Kendall's return period is presented by Salvadori (2004), Salvadori & De Michelle (2007a) and Vandenberghe et al. (2012). In this case the return period corresponds to the mean interarrival time of events called “super-critical” or “dangerous” events, which are more critical than the design event. The probability distribution function is partitioned into a super-critical and non-critical region. This partitioning is based on the

two-dimensional Kendall distribution (K_C), which represents multivariate information. Following this approach, the definition of a return period and the associated critical probability level, divides the events into super-critical and non-critical. The return period for an event described by two random variables and based on the annual maximum series, is given by following formula (see (28)).

$$T = \frac{1}{1 - K_C(t)} \quad (28)$$

where t is the probability level corresponding to the return period, which is the isoline of the two-dimensional analysis. For further details refer to Salvadori et al. (2007b).

According to Salvadori (2004), the “secondary” return period may have more practical significance, since it is defined such that the events can be partitioned into two groups whose boundary defines events corresponding to a certain severity. When a flood protection structure (e.g., a dam) is designed a critical threshold associated to a return period is defined and the secondary return period is the mean interarrival time of a critical event for the structure under consideration.

The concepts presented in this section show that considering two random characteristics of a flood jointly leads to more complex forms of return periods. In all the above equations, the magnitudes of the two random variables are considered simultaneously. Various combinations of pair values of the random variables can result from one same return period value.

In this work, the “OR” case is considered as design criteria, implying that the assumed dam can be under risk by an extreme flood characterized by either a high value of peak flow or a big volume. Any of the two extreme cases of flood can be considered as damaging for the dam and its structures in terms of maximum reservoir levels and spilled flows.

3.4. Synthetic hydrograph generation

The observed flood runoff hydrographs provide integrated information regarding the response of the basin and river to extreme precipitation. They capture characteristics of extreme storms including the duration, time-intensity pattern, amount and spatial distribution, as well as the antecedent moisture conditions in the watershed.

As mentioned in Yue et al. (2002) the severity of a flood depends not only on the flood peak, volume and duration, but also on the shape of the hydrograph. Two flood events may have the same peak flow and volume, but the event that has the peak delayed will require more storage volume to get the same degree of protection. As a result, the costs or flood-control policies will differ according to the adopted shape of hydrograph. An effective management of the water resources therefore requires a good estimation of the shape of the design-flood hydrograph.

In this work a flood event is represented by the following two characteristics: annual flood peak and flood volume. Given the peak flow and volume, in order to define the flood hydrograph the shape must be known. There are different methods to construct a flood hydrograph. According to Yue et al. (2002) they are classified into four groups: traditional unit-hydrograph methods, synthetic unit-hydrograph methods, typical hydrograph methods and the statistical methods. A description of each group can be found in the cited work.

Bell & Om Kar (1969) found that the parameters describing the shape of a single peaked hydrograph are: peak discharge, volume of runoff, time to peak and duration. Yue et al. (2002) state that a real flood hydrograph has three characteristic values: peak, volume and duration.

A triparametric Hermitian hydrograph is proposed by Aldama & Ramirez (1999). This hydrograph is a function of the peak discharge, time to peak and runoff volume, described by one equation for the first part up to the time to peak, and another one from the time to peak to the total duration of the event (base time). This type of hydrograph has the property that the runoff volume is related by the peak discharge and base time with the following relation: $V=Q_{\text{peak}}*t_{\text{base}}/2$.

Bergmann et al. (2000) use an analytical two-parametric function as standard hydrograph. The parameters are estimated by standardizing the direct runoff, using the standard volume and standard time.

The shape of the hydrograph is a random phenomena, because many of the factor governing this shape are random, for example: rainfall intensity, rainfall pattern, rainfall amount, rainfall center, snow-depth spatial distribution, temperature, etc (Yue et al., 2002). They represent the randomness of the flood hydrograph by two statistical properties: the mean and the variance of the shape. In this work, the shape of the flood hydrograph is described using a Beta probability density function by estimating the parameters with the shape variables. A design flood hydrograph is obtained by defining a return period and its corresponding flood peak, volume, duration, and Beta shape. This method is only applicable for basins whose flood hydrograph has one dominant peak, because the Beta function has one peak.

Bhunya et al. (2007) explore the potential of four probability distributions to develop the synthetic unit hydrographs. The distributions analyzed are: two-parameter Gamma, three-parameter Beta, two-parameter Weibull, and one-parameter Chi-square. To derive to the synthetic unit hydrograph they use two characteristics of the unit hydrograph: the peak flow and the time to peak. They conclude that the Beta and Weibull are more flexible in the description of the hydrograph shape, and the Beta should be a preferred method.

Mediero et al. (2010) propose a methodology to obtain the design flood hydrographs for designing dams in Spanish catchments. They generate synthetic Peak flow-Volume pairs, and then use the ratio between the two synthetic values to estimate the shape of the hydrograph. They state that the selection of one method to construct a hydrograph restricts its shape and homogenizes the results, and argue that in order to relax this restriction random shape hydrographs must be used. When the set of observed hydrographs is large and sufficiently varied, it can be utilized as a random sample. They worked with a set of 919 hydrographs and used the standardized time to peak and location of the hydrograph centroid to test variability of the hydrograph shapes. The variability resulted wide enough to consider the observed hydrographs as random variables. Then they calculated the ratio Peak Flow-Volume for each hydrograph, and finally standardized each graph by the peak flow. In order to assign a hydrograph shape to the synthetic Peak flow-Volume pairs, they compare this ratio and select the observed hydrograph shape with the most similar ratio, and then resize it by the synthetic peak discharge.

The Kozeny-function and the Hyperbolic functions were presented by Bender et al. (2011) to generate synthetic hydrographs. These functions depend on the peak flow and time to peak values. They point out the limitation of these functions to single-peak shaped hydrographs, they are not applicable for catchments with multi-peak or slowly rising hydrographs and long (in duration) peak flows.

Serinaldi & Grimaldi (2011) state that, along with the statistical information of the event rarity, a design hydrograph should synthesize and preserve some physical properties, like peak discharge,

volume, duration and shape. They use and compare the following distribution functions to represent and synthesize direct runoff hydrographs: the two-parameter Beta family and the three-parameter generalized standard two-sided power distribution. They conclude that the generalized standard two-sided power distribution is a good alternative to the Beta family when the hydrographs claim for a two-sided J-shaped pattern. The approach on the basis of this distribution functions allows for two degrees of freedom, so two main attributes of the hydrograph must be chosen. They emphasize that the peak flow, volume, duration and shape of the hydrograph (parameters summarizing the shape) are linked to each other and could be studied by a joint frequency analysis.

Bergmann et al. (2000) developed a methodology to study the overtopping probability of a dam. One part of the methodology involves the generation of synthetic hydrographs. They state that there is always a certain event range in terms of the peak flow-volume pairs, delimited by the extreme peak runoff hydrograph and the extreme runoff volume hydrograph. For example there are no floods with an extreme peak runoff and a volume equal to zero. According to their work, the boundaries of the event range can be estimated in terms of probable minimal and maximal values for the time to peak of a certain watershed, and then relate these values with volume/peak ratios to delimit the possible pairs of peak flow-volume.

Bell & Om Kar (1969) studied different characteristic response times of small catchments in the United States in order to enable the synthesis of a design hydrograph. The characteristic times studied are: rise time (time to peak), time of concentration (average rise time), lag, time to equilibrium and volume/peak ratio. The results of the study demonstrated that the rise times for a particular catchment are not constant, with values ranging between 40% and 200% of the median values. They also concluded that larger floods do not have smaller rise times, and that the median rise time is consequently an appropriate value to use for the design floods. They state that the rise times are not completely dependent on the physical features of the catchment because they vary to some extent with the effective storm duration. They also found a systematic variation of the volume/peak ratios suggesting that these two characteristic times depend on the physical nature of the catchment may be dependent on the storm type as well.

From a review of a large number of unit hydrographs, the Soil Conservation Service suggests a triangular approximation of the hydrograph, for which the time of recession may be approximated as $5/3$ times the time to peak, i.e. a total duration of $8/3$ times the time to peak (see. Chow et al., 1988; Maidment, 1993). In another study, the Soil Conservation Service based on a large number of unit hydrographs from many large and small rural watersheds, indicates that the hydrograph can be approximated to a curvilinear shape. The total duration is approximately 5 times the time to peak, and the basin lag time, which is the time between the centroid of the rainfall excess and the time to peak, can be approximated to $0.6T_c$, where T_c is the time of concentration of the watershed. They observed that approximately $(3/8)$ of the total volume occurs before the time to peak, but this value can be greater in mountainous watersheds and lower in flat, swampy areas (see. McCuen, 1998).

Viglione et al. (2010) studied the dependence of the flood response on the space-time interactions between rainfall, runoff generation and routing mechanisms by analyzing four different events and the resulting hydrograph characteristics. The hydrographs are described by three different characteristics: the catchment rainfall excess rate, and the first and the second temporal moments of the flood response. The first temporal moment is defined as the mean runoff time of the catchment (or time to peak) and it is the time to the centre of mass of the runoff hydrograph. The

second moment is the variance of the timing of the runoff, i.e. the temporal dispersion of the runoff hydrograph. They conclude that the first and second temporal moments of the hydrograph, hence the hydrograph shape, are affected by the space-time rainfall patterns.

Gaál et al. (2012) studied the duration of floods and the relationship with the climatic controls (storm type) and catchment controls (soil, soil moisture, geology, land form), by comparing the hydrographs of maximum annual flood events registered in different catchments in Austria. They state that the duration of a flood is related to the duration of the meteorological input and to the delay of the input in the catchment. Storms with durations similar to the response time-scale of the catchment typically produce the maximum annual flood, they lead to larger floods than shorter and longer storms. They mention that the flood duration is the result of a complex interplay of numerous factors, involving the routing of surface runoff, which is a mix of soil moisture and groundwater response, and the temporal behavior of the precipitation. The results of the study indicate that the duration of the floods do not depend much on catchment area, and rather other controls are much more important. According to their results the durations of the snowmelt events are more than twice the duration of other event types.

The duration of the direct runoff hydrograph is generally uncertain and is a function of the base flow separation technique (England, 2003). Zhang & Singh (2006) analyzed the correlation between the following flood characteristics: flood peak, volume and duration for two rivers in North America. The results show that in both cases the linear correlation between the flood peak and duration is small and negative. Therefore they assumed these two variables as independent from each other.

As mentioned previously, in this work the flood event is represented by the two characteristics: peak flow and volume. After reviewing and comparing the different approaches to generate synthetic hydrographs and the considerations regarding the estimation of duration of flood events, it was considered appropriate to use the two-parameter Beta distribution to model the shape of the hydrograph, and the total duration of the event is estimated so that the selected shape accurately reproduces the peak flow and volume.

The procedure to construct the synthetic hydrographs using the Beta distribution is done on the basis of the method proposed by Yue et al. (2002), which is summarized in the following paragraphs.

The first step is the separation of the flood events from the total serie, i.e. defining a start and end date for the flood. Then the base flow has to be subtracted from the total hydrograph resulting in the direct flow hydrograph. This direct flow will be modeled by the two-parameter Beta distribution. The probability density function of the Beta distribution is given as follows (see (29)).

$$f(t, a, b) = \frac{1}{B(a, b)} * t^{a-1} * (1 - t)^{b-1} \quad (29)$$

where a and b are the parameters of the function, and $B(a, b)$ is the complete Beta function. The parameters can be estimated using the method of moments, or in terms main characteristics of the hydrographs: the peak position (mode) and the value of the density function in the peak position (maximum value of the density). The second approach is chosen, and the equations used for the estimation, as presented by Serinaldi & Grimaldi (2011), are the described here (see (30) and (31)).

$$t_{peak} = \frac{a - 1}{a + b - 2} \quad (30)$$

$$f(t_{peak}) = \frac{(a-1)^{(a-1)} * (b-1)^{(b-1)}}{(a+b-2)^{(a+b-2)} * B(a,b)} \quad (31)$$

where t_{peak} is the time to the peak. It is important to point out that both the base of the Beta distribution and the area surrounded by it are equal to one, so in order to estimate the parameters a and b from the observed direct hydrographs, they must be modified to meet this conditions. This is done by dividing the base of the flood hydrograph by its duration and by multiplying each ordinate of the flood hydrograph by the ratio duration/volume. This way the modified flood hydrograph has a total duration and a total volume equal to one. For each observed flood hydrograph the parameters a and b are estimated, resulting in a serie of pairs of parameters for the whole data. For further details regarding this methodology refer to Yue et al. (2002), Bhunya et al. (2007) and Serinaldi & Grimaldi (2011).

The observed pairs of shape parameter (a and b) are used to generate possible shapes of hydrographs. This direct hydrographs modeled with the Beta distribution must be again modified to represent the characteristics of the flood: direct flow peak and volume. This is done by multiplying the base of the Beta distribution function by the total duration of the event and by multiplying each ordinate by the ratio volume/duration. This procedure is the opposite as the one described in the previous paragraph. At this point the duration of the event is estimated, so that the pair of characteristics of the flood matches the selected shape. This is achieved by equalizing the direct flow value, which is information given by the pair of flood characteristics, with the value resulting by multiplying the magnitude of the Beta function in the time to peak and the ratio of Direct Volume/Duration. In this equation the only unknown is the duration, which is calculated so that the selected shape of hydrograph exactly reproduces the flood characteristics: direct peak flow and volume.

To generate synthetic hydrographs, pairs of shape parameters must be generated from the observed ones. This procedure is similar to the one followed to generate the pairs of flood characteristic values, i.e. the values are either randomly taken from each marginal distribution function separately or from a Copula model considering the correlation between the two shape parameters.

Once the synthetic hydrograph is obtained from the Beta distribution shape and the modified according to the pair of values direct peak flow and direct volume, the direct flow hydrograph is ready. The next step is to add up the base flow.

The value of the base flow corresponding to the time to peak (Peak Base Flow) is estimated using a linear regression model between the total peak flow and the direct peak flow. The base volume is also estimated using a linear regression model between the direct volume and total volume of the flood. It is important to mention that the regression model of the flows is more accurate than the one corresponding to the volumes, as will be presented in the results section. For this reason, in every case of synthesis, the Peak Base Flow is exactly reproduced, and the base volume is adjusted, and can either be exactly reproduced, overestimated or underestimated. The adjustment of the base volume is done by estimating a Base Flow*, corresponding to the beginning of the event which is considered as equal to the one at the end of the event, so that the total base volume equals the one estimated with the regression model.

Following this procedure, there are three possible results which must be analyzed different. The first case is that the Base Flow* is lower than the one in the time to peak. This is the behavior of the

baseflow according to the technique of separation used in this work. In this case the base volume is equal to the one estimated by the regression model. In a second case the Base Flow* is higher than the one in the time to peak, this is a typical result for floods with low durations. This type of behavior indicates that during a flood the base flow is decreasing, which differs from the behavior of the base flow separation with the applied technique. In this case a constant base flow value, equal to the value in the time to peak is adopted, underestimating the base volume. The last case is a typical result of floods with long durations, in which the resulting Base Flow* is a negative number, which is not physically possible, so a value of Base Flow* equal to zero is adopted, overdimensioning the base volume. The different possible cases previously explained are presented in the following table (see Table 3.2).

Table 3.2: Possible cases in base flow estimation approach.

N°	CASE	Adopted Base Flow*	Graphical Explanation	Base Flow Volume
I	Base Flow* < Peak Base Flow	Estimated Base Flow*		Estimated Value from the Regression
II	Base Flow* > Peak Base Flow	Peak Base Flow		Underdimensioned
III	Base Flow* < 0	0		Overdimensioned

For every synthesis of hydrograph, the peak flow and direct volume are used to estimate the peak base flow and base volume. The shape of the event in combination with the pair of values of flow and volume is used to estimate the duration. This duration along with the peak base flow and base volume will define to which of the three cases of base flow estimation the hydrograph belongs, and the base flow is added up to the direct hydrograph previously estimated with the Beta distribution.

3.5. Reservoir routing

The risk of a flood to a dam can be assessed by routing the hydrograph through the reservoir and defining the maximum water level reached or/and the maximum flow released. The hydrological design of dams is based either on the Probable Maximum Flood or on a given return period. The second design criterion requires the identification of flood events associated with high return periods. In order to adequately assess the safety of a dam in the presence of any flood or to define flood events associated with high return periods a large number of routed hydrographs is required. As a result the observed set of hydrographs must be extended by generating synthetic hydrographs.

In this work two different groups of hydrographs are routed. The first group is a set of 1000 hydrographs that are randomly generated in order to assess the risk of overtopping the dam. The synthesis of these hydrographs is done by matching random shapes with random flood characteristics, i.e. pairs of peak flow-direct volume. The second group corresponds to a set of critical hydrographs that are routed through the reservoir, which is presented as a possible procedure for designing a dam and its hydraulic structures. The design flood events are high flood hydrographs that have a low probability of being exceeded and ensure the dam's safety to a given level. The synthesis of this last group of hydrographs is done by selecting critical shapes of hydrographs and matching them with pairs of characteristic of floods that are associated with a defined design return period. The critical shapes are selected in terms of producing maximum reservoir levels and spilled flows. All the synthetic hydrographs preserve the statistical characteristics of the observed flood events.

In order to perform the routing of a flood through a reservoir the following variables and conditions must be defined: the inflow hydrograph, the water level in the reservoir at the beginning of the flood, the reservoir storage capacity as a function of the height and the discharge capacity of the spillway. The spillway operation policy should also be defined in case of spillways with gates. In this section the different considerations involved in the process of routing the hydrographs is presented.

The sets of hydrographs are created by combining pairs of peak flow and volume with different shapes of hydrographs. The hydrological safety of the dam in terms of maximum water levels and adequacy of the spillway is assessed by routing through the reservoir the synthetic hydrograph sets. The generation of the synthetic hydrographs is explained in more detail in the previous section (see sub-chapter 3.4).

The discharge equation for an overflow spillway without gates, according to Novak et al. (2007), is given by following formula (see (32)).

$$Q = \frac{2}{3} * \sqrt{2g} * b * C_d * H^{3/2} \quad (32)$$

where, g is the value of the gravity acceleration, b is the width of the spillway, C_d is the discharge coefficient and H is the total head over the spillway crest. Note that the discharge coefficient does not have units, and the value is a function of the ratio between the design head used for the derivation of the spillway shape and the maximum actual head going through the spillway. Recommended values can be found in Novak et al. (2007).

The initial reservoir level at the beginning of the flood event can be chosen randomly. Note that the reservoir levels are influenced not only by the precipitation, evaporation, inflows but also by the operating policy of the dam manager. Salvadori et al. (2011) analyzed the correlation between the

flood peak, volume and the initial water level of the reservoir, observed in a dam in Italy. They state that the water level is arbitrarily fixed by the dam manager. The correlation tests indicate that the initial water level is significantly independent from both flood characteristics.

Goodarzi et al. (2011) conducted a risk analysis to assess the safety level of an existing dam in Iran. They considered the following uncertain variables: quantile initial depth of water in the reservoir and discharge coefficient of the spillway, because they cannot be quantified exactly. They used sampling techniques in order to quantify the uncertainty of these random variables.

The total reservoir elevation is the result of the initial water level plus the additional elevation height due to the random flood event. The synthetic hydrographs are continuous equations that are routed through the reservoir with a time step calculation of 1% of the total duration of the event. Three different water elevations are calculated in each time step (i): an initial, a medium and a final (see (33) to (35)).

$$L_{INI(i)} = L_{FIN(i-1)} \quad (33)$$

$$L_{MED(i)} = L_{INI(i)} + \Delta L_{INFLOW} \quad (34)$$

$$L_{FIN(i)} = L_{MED(i)} - \Delta L_{OUTFLOW} \quad (35)$$

where $L_{X(i)}$ are the levels in the step of calculation i , ΔL_{INFLOW} is the variation of the reservoir level due to the incoming flood hydrograph and $\Delta L_{OUTFLOW}$ the variation due to the flows spilled. These variations are estimated as the total incoming or overflowing volume in the time step divided by the reservoir area. The volume is the flow (inflow or spilled flow) times the time step of calculation, and the reservoir area is a function of the water elevation (see sub-chapter 4.3).

The routing of the floods is done for their total duration plus an additional time of 30% of the duration in which a constant base flow of the hydrograph is routed. The routing is performed with this additional time in order to detect maximum water elevations in the case of floods with times to peak close to the end of the event. The response of the reservoir elevations is delayed respect to the time of the peak flow, and in some cases the maximum reservoir elevation can occur when the flood is over. As an example, the elevation of the water in the reservoir for a flood with a total duration of 10 days will be computed every 0.1 days, and the routing will be performed for 13 days in order to identify the maximum water elevation reached for this event. In the following figure the components of the routing of one hydrograph are presented (See Figure 3.5). Note that the base of the graph shows 130% of the duration of the flood event.

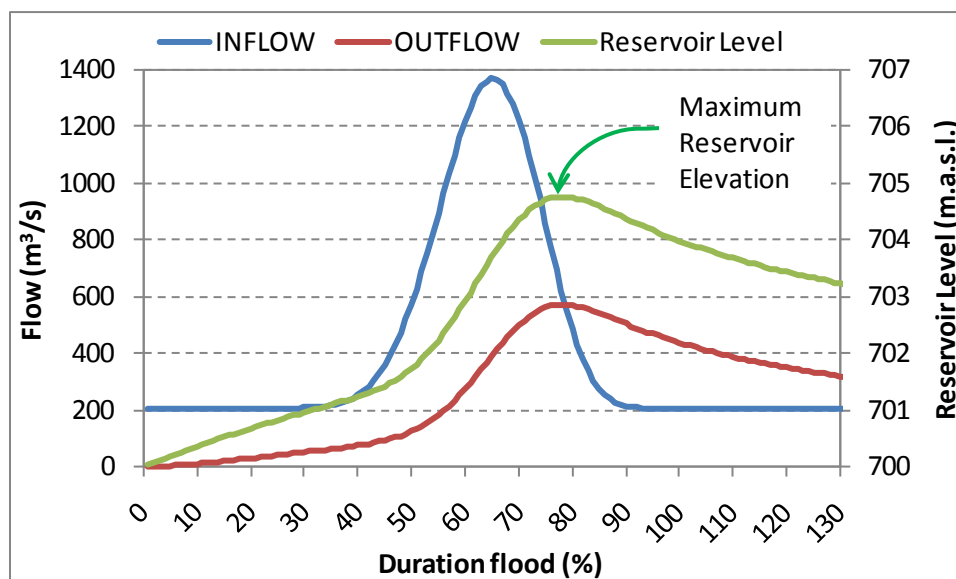


Figure 3.5: Reservoir routing components: Inflow hydrograph, spilled flows and reservoir levels.

In this work, the overtopping potential of the dam is assessed by routing different hydrographs through a reservoir and considering an initial water elevation equal to the spillway crest level. The spillway is assumed to operate with a constant discharge coefficient, in all the considered cases and time steps. The maximum reservoir elevation reached by each of the routed hydrographs is computed along with their empirical distribution. The resulting reservoir elevations are analyzed to determine the percentage of time that the reservoir exceeds a certain level. The same procedure is done with the maximum outflow or spilled flow values.

It is important to mention that the design of a dam involves the definition of the total height of the dam. In order to define this height a freeboard should be considered. A minimum freeboard is defined by the Bureau of Reclamation (1987) as the difference in elevation between the crest of the dam and the maximum reservoir water surface that would result if the design flood occurs and the outlet works and spillway function as planned. Freeboard requirements depend on maximum wind velocity, fetch, reservoir operating conditions, spillway capacity, whether coping or parapet walls are used, and also on the type of dam. The estimation of the freeboard is not considered in the present work.

4. STUDY AREA AND DATA

4.1. Description of the catchment

The river Agrio basin is located in the Province of Neuquén in mid-west Argentina (see Figure 4.1). The river is originating at the Volcano Copahue, feeds lake Caviahue in the upper most catchment and continues flowing downstreams after the lake. Along the way the river Agrio receives flows from many small streams covering a total drainage area of approximately 9700 km². The Agrio is the principal tributary to the Neuquén river. The Neuquén catchment has poor vegetal land cover, steep mountains and very few natural regulators. These natural characteristics cause important runoff events in the upper catchment and floods that often threaten the inhabitants in the lower catchment.

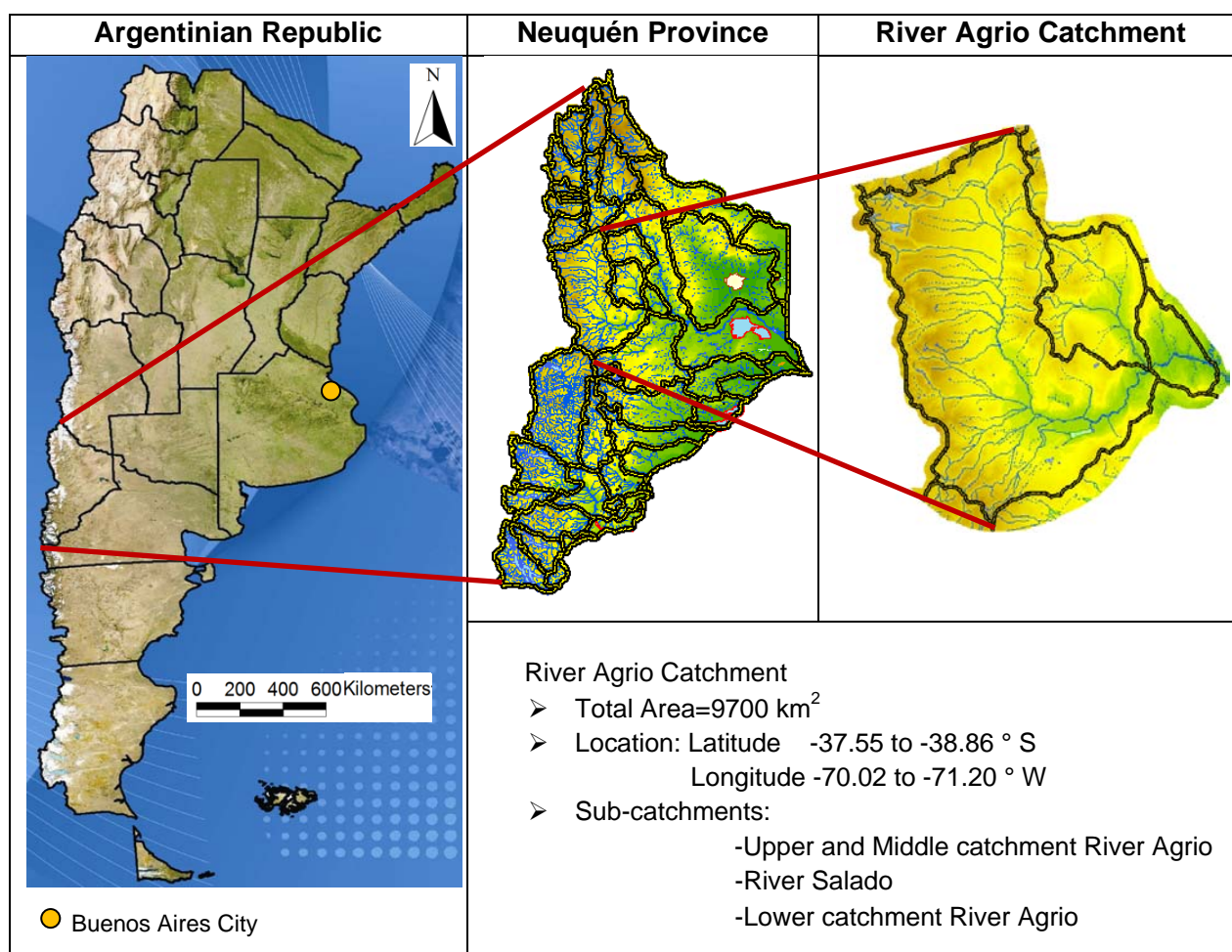


Figure 4.1: Map with general location of the study area.

The study area can be characterized as highly heterogeneous due to the spatial variation of precipitation and temperature. The precipitation varies with the meridians, with high amount of rainfall on the border to Chile, followed by a very rapid decrease towards the eastern region. This phenomena is caused by the incoming masses of humid air from the Pacific Ocean that are forced to climb the mountain chain of the Andean Mountains, cooling down due to the decrease of the atmospheric pressure, condensating and precipitating as rain or snow. The temperature variation within the region is associated with the altimetry having low temperatures in the highest peaks. (See Gobierno de la Provincia de Neuquén (2008), Valicenti (2001)).

In the river Agrio basin the region with the highest precipitation shows a magnitude higher than 1000 mm/year on the border to Chile and the lowest values are around 200 mm/year in the eastern region, according to the isohyets provided by the Subsecretaría de Recursos Hídricos (SSRH, 2002). The altimetric extreme values range from 2800 m.a.s.l. in the highest peak to around 550 m.a.s.l. in the discharge to the Nequén river, according to the elevation information provided by SIG250 (Benedetti, 2000). Moyano & Diaz (2006) classified different climate zones according to the hidric-regimes, taking into account the Aridity Index (UNEP 1997). From west to east, the climate zones found in the catchment are: hyper-humid, humid, sub-humid and semiarid.

As was mentioned previously, the region has poor vegetal land cover. The areas with natural forests are not relevant in terms of extension. Small wild bushes are more frequent, which results in a low natural capacity of regulating the floods. This natural phenomenon is aggravated by the grazing activity in extended areas, especially cattle.

The different seasons in the region are divided along the year in the following way: the Spring starts at the end of September and goes until the end of December, Summer takes place from January to March, Autumn from April to June and Winter from July to the end of September.

The river Agrio's natural hydrological regime shows two annual flood events that can be clearly characterized: The highest winter floods are caused by rainfall-excess, whereas the floods in spring are generally generated by the combination of snowmelt and rainstorms. During the winter, especially from May to August, most of the precipitation falls and is partly stored as snow in the highest mountains of the catchment. The mean temperature values and total monthly precipitation registered in the upper catchment of the river Agrio are shown in the following figure (see Figure 4.2).

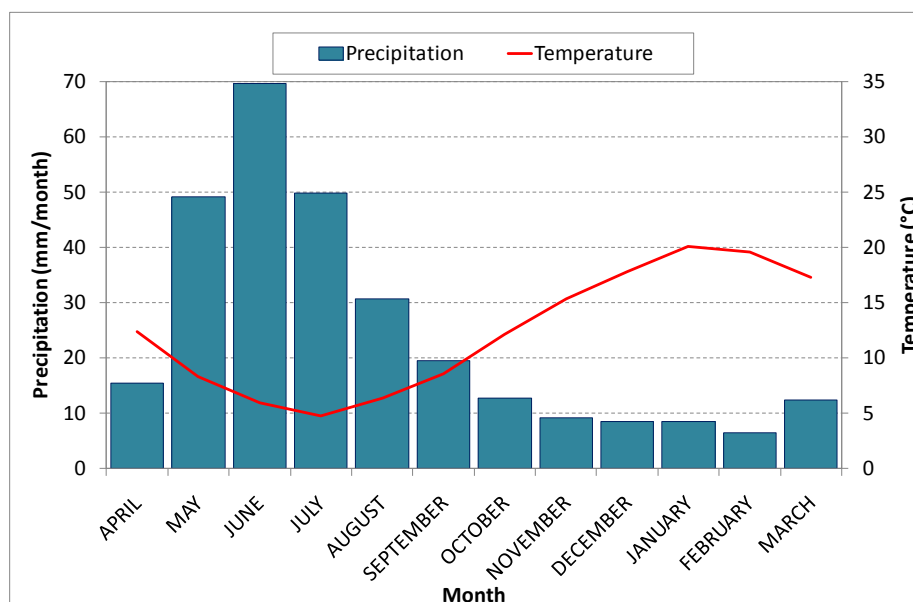


Figure 4.2: Monthly values of precipitation and temperature in the upper catchment.

The winter flood events are caused by the portion of rainfall that is not stored as snow and are characterized by high values of peak flows in relation to the associated volumes. The second peak occurs in spring, generated by the fusion of the accumulated snow and in some cases in combination with rain events. The associated peak flows are lower compared with the winter events. The low streamflow period is observed on the end of the summer, resulting in a hydrological year

from April to March. The monthly flow values (mean, maximum and minimum) of the river Agrio are presented in the following figure, showing the two peaks regime (see Figure 4.3).

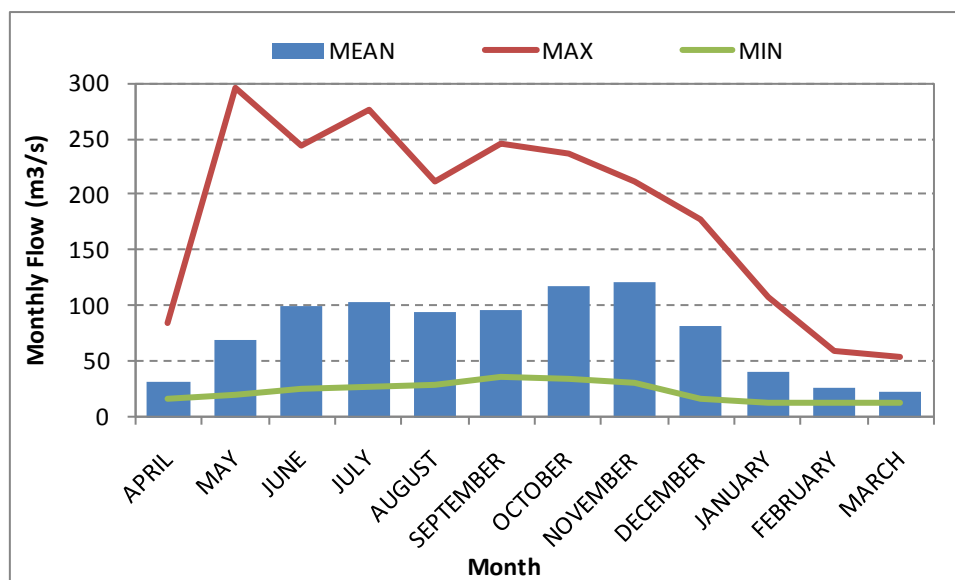


Figure 4.3: Monthly flows river Agrio.

For further information regarding the characteristics of the catchment and region refer to Gobierno de la Provincia de Neuquén (2008) and Valicenti (2001).

4.2. Hydrological data

This study is carried out based on the daily flows provided by the Subsecretaría de Recursos Hídricos de Argentina. The available data was registered at the station called Bajada de Agrio, located at latitude -38.37°S and longitude -70.03°W and 660.00 m.a.s.l., covering a total surface area of 7300 km^2 . In the following figure this area is shown, corresponding to the Upper and Middle catchment of river Agrio, as well as the location of the available gauging station (see Figure 4.4).

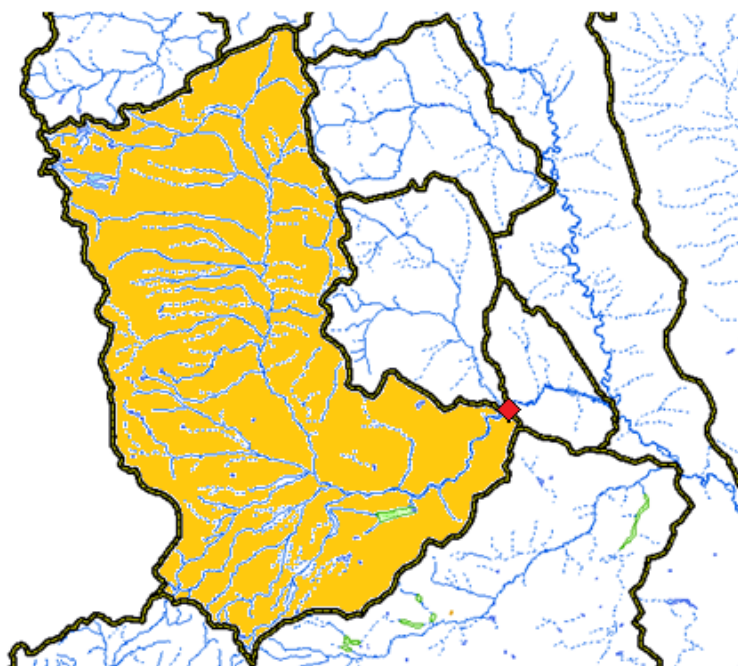


Figure 4.4: Map of the catchment area and location of the gauging station (red mark).

The data base includes daily flow data registered from 1958 to 2011, with missing data in three years (1967, 1977 and 1989), providing a total of 50 complete years to study the maximum flows. Two annual peak flows were identified for every year, the winter peak was identified in the period from April to September, and the spring peak from October to March. The characteristics of the peak flows data are described in the following table (see Table 4.1).

Table 4.1: Basic statistics of observed flood peaks series.

Type of Event	Winter: Rain Events	Spring: Snow-melt Events
Length [years]	50	49
Flood Peak [m³/s]		
Mean	541.8	232.4
Maximum	1249.5	692
Minimum	48.4	42.6
Standard Deviation	377.7	134.4

As mentioned in the previous section, the magnitude of the winter peaks is greater than the ones occurring in spring. Note that the spring flood events contain one less value than the winter events, this is caused by a particular case (year 1990-91) in which the spring event cannot be identified as independent from the winter flood event, and as the spring event occurs right after the winter flooding this flood event was discarded from the analysis.

4.3. Hypothetical reservoir

There are no dams existing along the river Agrio. In order to compare the univariate and multivariate frequency analysis, a hypothetical reservoir must be implemented. An elevation model with terrain curves every 100 meters, provided by SIG250 (Benedetti, 2000), was available. In the previous section the elevation of the gauging station was mentioned to be 660 m.a.s.l.. The dam is assumed to be at the location of the gauging station and perpendicular to the river flow. The spillway crest is considered to be at the elevation of 700 m.a.s.l.. The resulting total reservoir surface at the spillway crest elevation is 8025 Ha. This surface was compared with values of existing projects in the region, to decide whether it is viable. According to the information provided by the Subsecretaría de Recursos Hídricos (SSRH, 2002), the following projects have similar reservoir surfaces: Florentino Ameghino, Futaelufú, Alicurá and El Nihuil.

The characteristic curves of the reservoir, i.e. the curve of the total reservoir surface and the one representing the volume above the spillway crest, both as functions of the water level are presented in the following graphs (see Figure 4.5).

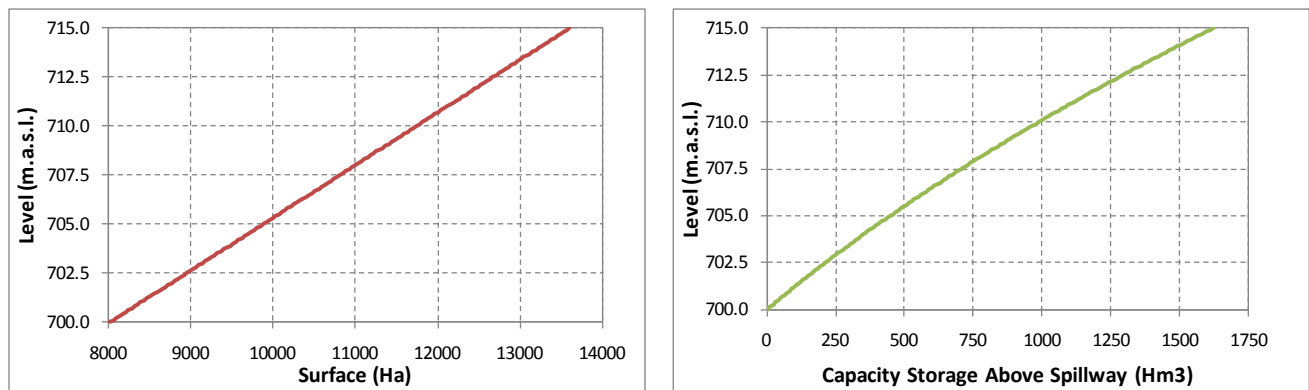


Figure 4.5: Characteristic curves of the reservoir. Left: Surface-height curve; Right: Volume above spillway crest-height curve.

Note that the relationship between the reservoir surface and water level is linear. This curve was estimated using the available elevation information, and is the result of the interpolation between the reservoir surfaces corresponding to the elevations 700 and 800 m.a.s.l..

The spillway considered for the reservoir routing is an overfall (ogee) spillway without gates. The total width is 25 meters. The spillway is assumed to operate, in all the considered cases, with a water head equal to the design head, resulting in a constant discharge coefficient. This discharge coefficient was chosen as 0.745, as it is recommended by Novak et al. (2007).

5. RESULTS AND DISCUSSION

In the following sections the results obtained in the different steps of calculation are resumed. Some graphical examples are included for a better understanding of the process followed. The general calculation includes: Frequency Analysis of the flood variables (i.e. peak flow and direct volume), Return Period Analysis of the variables, Hydrograph Analysis, Synthesis of the Variables for the risk assessment case and for the design case, Synthesis of the Hydrographs, and finally the Reservoir Routing.

In the appendixes A to D, flow calculation charts are presented describing the steps followed in each of the general calculations. The Appendix A describes the different steps followed in the Frequency and Return Period Analysis of the flood variables. In the Appendix B the steps followed for the Hydrograph Analysis are described. Appendix C presents the steps followed for the Synthesis of the flood variables and hydrographs and the reservoir routing for the risk assessment case. The last Appendix (D) is similar to Appendix C, but for the design case.

5.1. Hydrograph separation

To perform the work it is necessary to check the independency of the flood events. When the flood event is identified peak flow, total volume, duration and time to the peak of the event are defined. In order to generate synthetic hydrographs, the direct runoff hydrograph must be described. The base flow has to be separated from the total flow, defining a base volume and base peak flow, and the direct volume and peak flow. The different components describing the flood are presented in the following figure (see Figure 5.1).

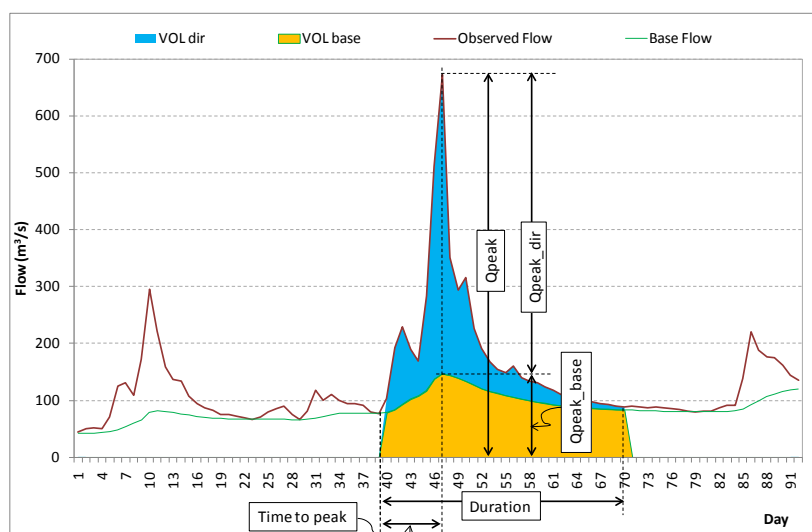


Figure 5.1: Components of the hydrograph.

5.1.1. Base flow separation

The base flow program was run for the three complete daily streamflow series separately. The results indicate that the baseflow contribution is around 75% of the total flow of the river.

An example of the base flow separation resulting from a particular year is shown in the following figure (see Figure 5.2). From this figure the difference between the winter and spring events can be identified. Note that the winter event presents a higher peak, but the duration is lower resulting in a lower contribution from the base flow.

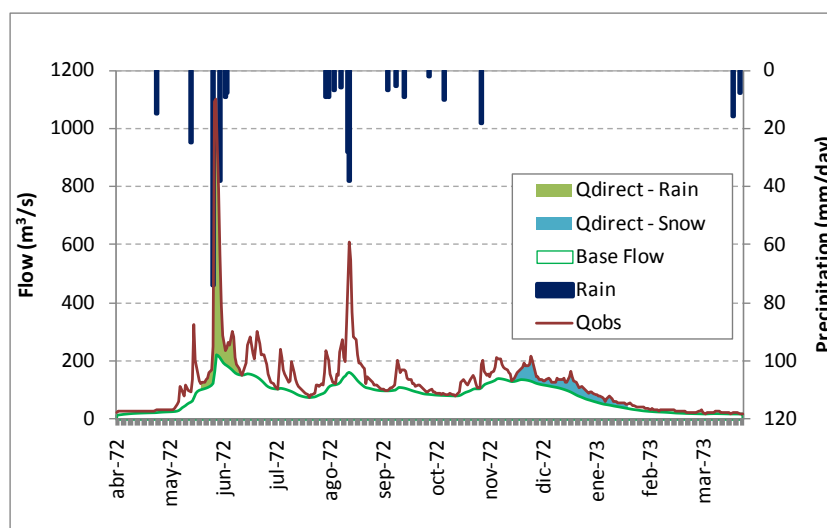


Figure 5.2: Example of the resulting base flow separation for one year.

5.1.2. Identification of independent events

The two highest peak flows for every year, corresponding to the rain and snow melt case, were identified from the available daily streamflow data. As was mentioned in the Section (see sub-chapter 3.1), once the direct flow is separated from the total flow, the flood event is defined as independent from other events according to the criteria proposed by Cunnane (1979), which states that the time between flood peaks has to be three times the average time to peak and the low flow between the peaks has to be lower than two thirds of the first peak. The average time to peak is defined as the mean value of the times to peak of several clean typical flood hydrographs, and was found to be 3.5 days for the winter events and 13.5 days for the spring events. A case example showing the identification of independent flood events is shown in the following figure (see Fig 5.3).

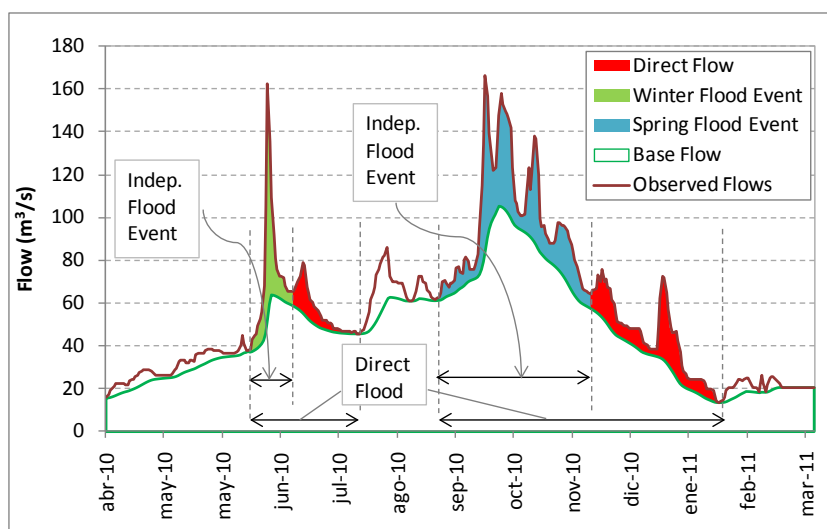


Figure 5.3: Example of the resulting independent flood events for one year.

5.1.3. Results of the hydrograph separation

Once the independent flood events are identified, the different characteristics of the floods are analyzed. These characteristics include: peak flow (total, direct and base), volume (total, direct and

base), duration and time to peak. The statistics of the observed peak flows and their corresponding volumes, for both seasons, are summarized in the following table (See Table 5.1).

Table 5.1: Basic information of the characteristics of the winter and spring floods.

Type of Event		Winter: Rain Events		Spring: Snow-melt Events	
Characteristic		Mean	Standard Deviation	Mean	Standard Deviation
Length (years)		50		49	
Peak Flow (m ³ /s)	Total	541.8	337.7	232.4	134.5
	Direct	432.7	298.4	121.0	106.4
	Base	109.1	46.2	111.4	38.1
Volume (m ³ /s*day)	Total	4111.5	1811.1	7072.1	2655.0
	Direct	1889.0	1160.0	1739.2	724.4
	Base	2222.5	828.7	5333.0	2060.4
Duration (days)		29.1	10.8	73.1	24.0
Time to peak (days)		11.0	9.6	16.0	9.0

In a further analysis, both the time to peak and duration values showed a weak dependence with the rest of the flood characteristics, indicating that these variables are not clearly related to the peak flow or volume magnitudes. These results coincide with the ones published in the reviewed and cited papers (See sub-chapter 3.4).

The aim of this work is to characterize a flood by the magnitude of the flow at the peak and the volume involved. The cross-correlation between the peak flow and the volume series (both total and direct) was estimated in order to decide which pair of flood characteristics should be used. The resulting cross-correlation coefficients, following the standard formulas, are presented in the following table (See Table 5.2).

Table 5.2: Cross-correlation coefficients for the peak flow and volume series.

Type of Event	Winter: Rain Events		Spring: Snow-melt Events	
Volume (m³/s*day) Peak (m³/s)	Direct	Total	Direct	Total
	Cross-Correlation Coefficients			
Direct	0.930	0.836	0.586	0.306
Total	0.942	0.865	0.679	0.439

These coefficients indicate that the series of total peak flow and direct volume show the highest correlation, both for the winter and spring cases, meaning that these two variables have the

strongest dependence. For this reason, these two characteristics are chosen to be used for the rest of the work, to analyze the bivariate frequency and to generate synthetic floods.

The following values, graphs and results are done on the basis of the series of total peak flows and direct volumes corresponding to both the winter and spring flood events. Some sample statistics are summarized in the following table (See Table 5.3).

Table 5.3: Sample statistics of the flood characteristic series: peak flow and direct volume.

Type of Event	Winter: Rain Events		Spring: Snow-melt Events	
	Total Peak Flow (m ³ /s)	Direct Volume (m ³ /s*day)	Total Peak Flow (m ³ /s)	Direct Volume (m ³ /s*day)
Mean	541.8	1889.0	232.4	1739.2
Standard Deviation	337.7	1160.0	134.5	724.4
Coefficient of Variation	0.62	0.61	0.58	0.42
Coefficient of Skewness	0.40	0.26	1.73	0.25
Coefficient of Kurtosis	-1.03	-1.18	3.74	-0.53
Maximum	1249.5	4001.9	692.0	3299.5
Minimum	48.4	111.1	42.6	285.9
Median	468.8	1663.9	196.4	1785.4
Interquartile Range	615.6	1812.4	147.8	1086.2

The peak flows and their corresponding direct volumes are shown in the following graphs (see Figure 5.4). The two graphs are plotted in the same scale in order to compare the behavior of the winter and spring floods.

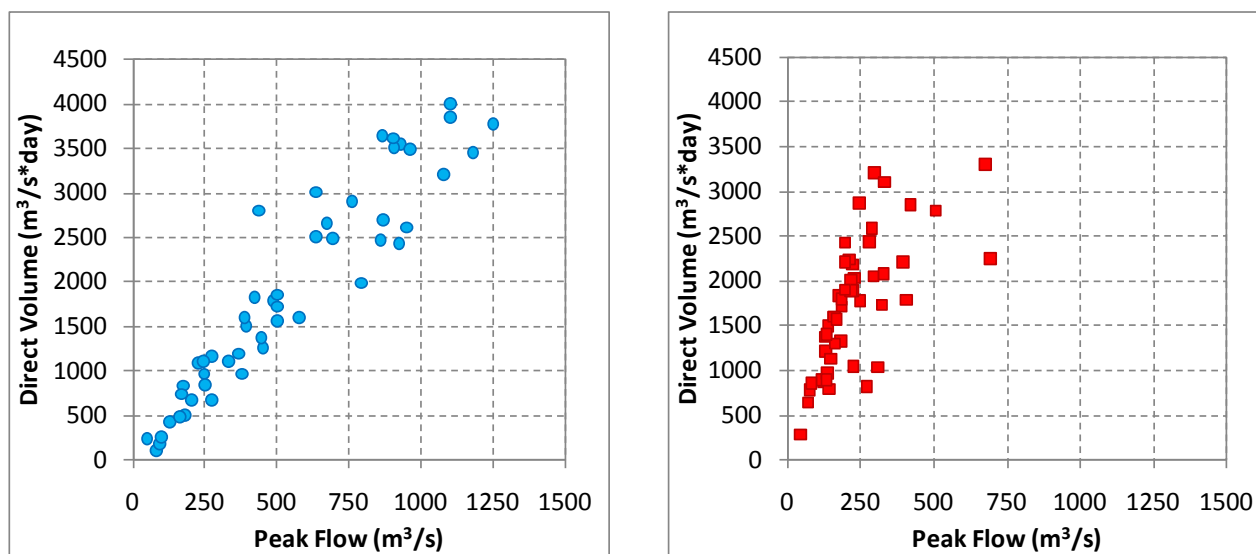


Figure 5.4: Observed pairs of peak flow and direct volume series. Left: Winter events; Right: Spring events.

The winter events show a clear linear relationship between the two series, while the spring events indicate a more disperse relationship. These results agree with the coefficients of the cross-correlation presented in Table 5.2, in which the values for the winter events are higher than the ones corresponding to the series of the spring floods. Another important difference is the behavior of the two types of events in terms of magnitudes. The clear linear relationship characterizing the winter events indicates that low values of peak flows have associated low direct volumes and the same is observed for the high magnitudes, whereas in the case of the spring events, a low value of peak flow (e.g. $250 \text{ m}^3/\text{s}$) shows wider range of associated direct volumes.

The Box & Whisker Plot corresponding to the four studied series are presented in the following figures (see Figure 5.5). The series were plot by grouping the different flood characteristics in order to compare the different behavior of the floods occurring during winter and spring.

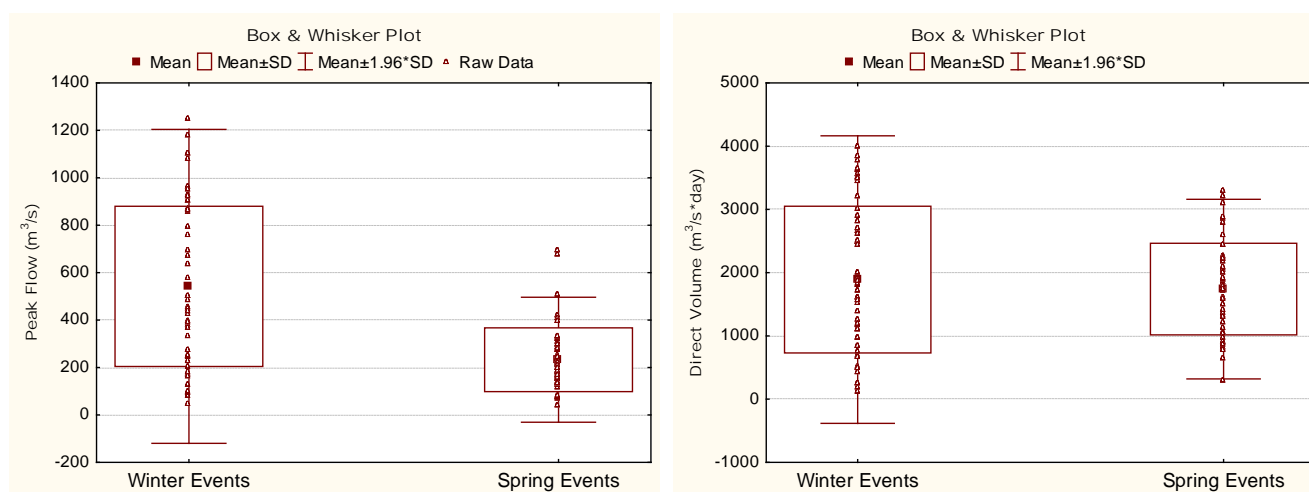


Figure 5.5: Box Plots of total peak flow and direct volume series for both the winter and spring flood events.

Note that the observed characteristics of the winter events (both peak flow and direct volume) are more disperse compared with the spring events. The magnitude of the winter flow peaks is higher than the ones occurring during spring. The direct volumes in both seasons have a similar mean value, but the winter events can have either higher or lower direct volumes compared with the spring flood events.

Both winter and spring events show some high values of peak flow that lay outside of the upper whisker end. These highest values were checked to decide whether they are outliers or just extreme cases. The Grubbs and Beck test results indicate these values are not outliers.

The Double Sum graphics were used to assess the possibility of inhomogeneities in the analyzed series. The graphs do not show any significant change in the relationship between the test and reference series, indicating that all series are homogeneous. The graphics are presented in the following figures (see Figures 5.6 and 5.7).

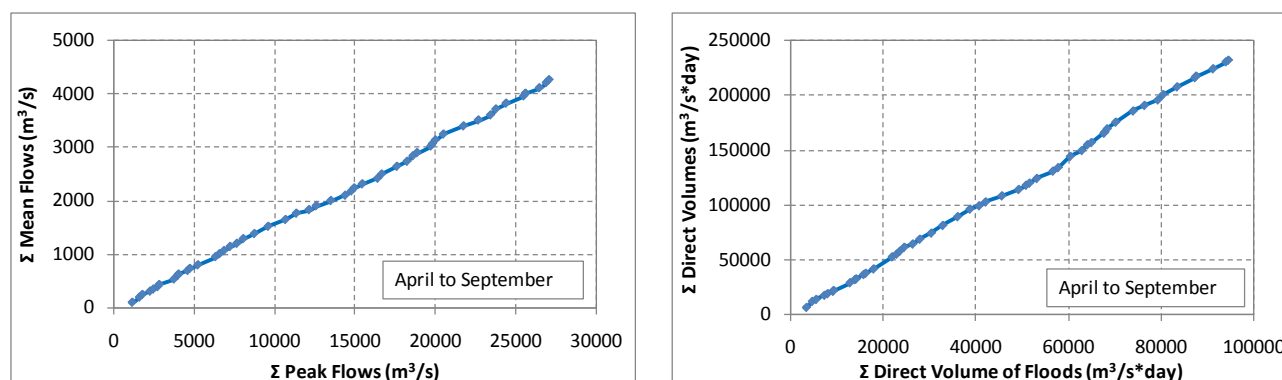


Figure 5.6: Double sum analysis of series observed during winter. Left: Peak flow; Right: Direct volume.

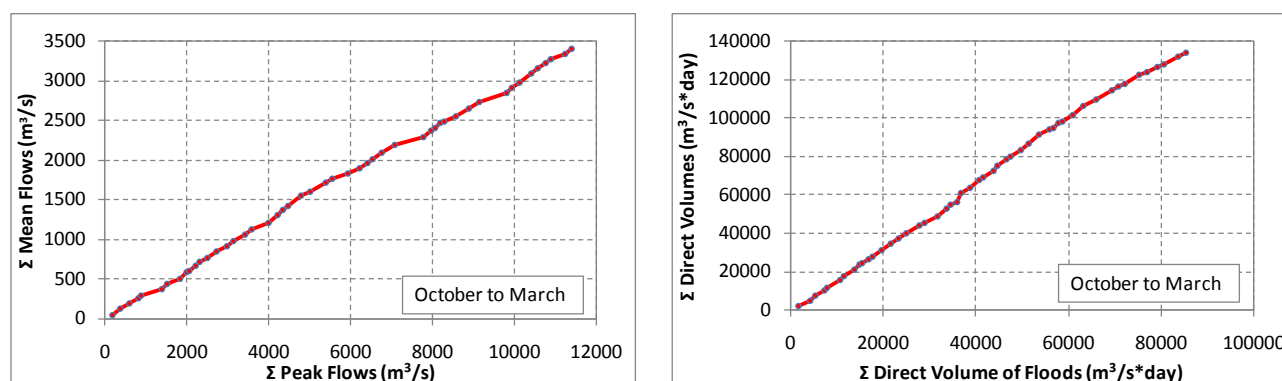


Figure 5.7: Double sum analysis of series observed during spring. Left: Peak flow; Right: Direct volume.

The results of the Mann-Kendall and Pettitt tests applied to the four analyzed series are presented in the following table (See Table 5.4). As was mentioned in Sections 3.2.3 and 3.2.4, the significance level used in this work is of 5%. The Mann-Kendall results with p-values higher than 0.05 indicate that the null hypothesis cannot be rejected, meaning that the tendency is not significant. The Pettitt results indicate the significance probability of a change point, and values lower than 95% indicate that the change point is not significant for a 5% of significance level. The obtained results indicate that the possible trends and jumps are not significant for the adopted significance level, so the series can be considered as stationary.

Table 5.4: Results on the Mann-Kendall (p-values) and Pettitt Tests (probabilities of a change point) applied to the peak flow and volume series.

Type of Event		Winter: Rain Events		Spring: Snow-melt Events	
Test	Serie	Peak Flow	Direct Volume	Peak Flow	Direct Volume
	Mann-Kendall	0.33	0.12	0.35	0.16
	Pettitt	80 %	87 %	55 %	85 %

All the results indicate that the observed series of peak flow and direct volume associated to the winter and spring flood events, meet the statistical criteria of consistency, homogeneity and stationary, and therefore can be used for a frequency analysis.

5.1.4. Relationship between the resulting series

The total flow at the time to peak (Peak Flow) and the direct volume of the floods, are used to characterize them. Therefore these two characteristics are analyzed to study the relationship between them and the rest of the flood component, in order to derive them from this pair of values.

A simple linear regression model showed a good representation of the direct flows using the total peak flows, both for the winter and spring events, with associated values (R^2) higher than 0.95 for both cases as shown in the following left images (see Figures 5.8 and 5.9). The total volume of the floods are also derived from linear regression models, but in these cases the associated values of R^2 are lower than 0.9 for the winter floods, and lower than 0.75 for the spring events, indicating a weaker linear dependence as shown in the following right images (see Figures 5.8 and 5.9).

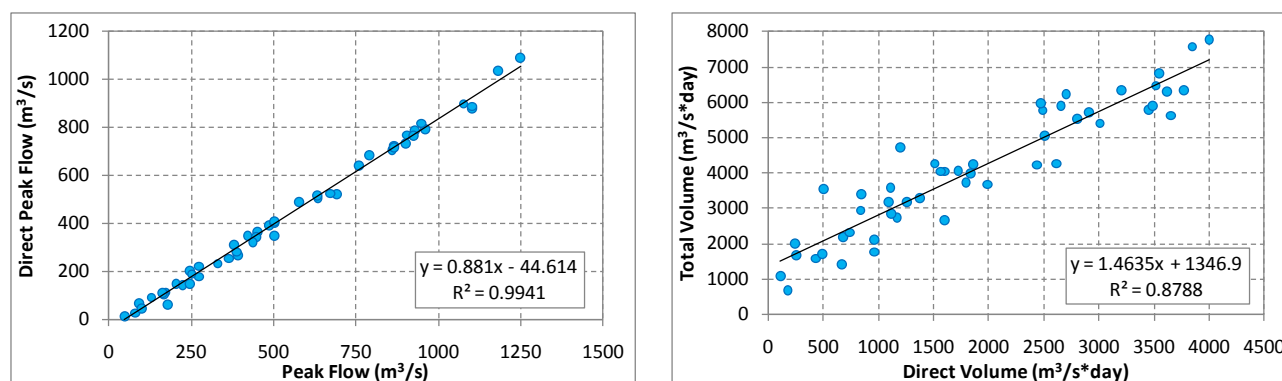


Figure 5.8: Relationship between flow and volume series for the winter flood events. Left: Peak flow; Right: Volume.

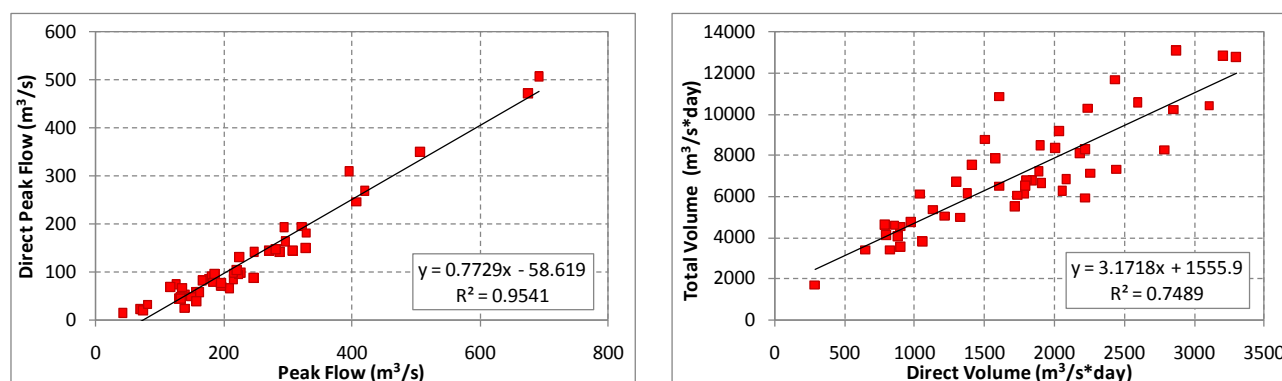


Figure 5.9: Relationship between flow and volume series for the spring flood events. Left: Peak flow; Right: Volume.

The base components of both flow and volume are estimated by subtracting the direct magnitude from the total. The previous applied linear regression models indicate that the representation of the peak flows is better than the one of the volumes. For this reason, as mentioned previously (see sub-chapter 3.4), the synthetic hydrographs will reproduce the exact values of the peak flows. The total volume will be represented in an approximate way by adjusting the base volume according to the methodology applied. It is important to make clear that the characteristics used in this work to represent the flood, i.e. direct volume and total peak flow, are exactly reproduced in all the synthetic hydrographs.

As mentioned in the previous section, both the time to peak and duration series showed a weak dependency with the rest of the flood characteristics. Multiple-linear regression models were studied in search of a satisfactory way of relating these time scale characteristics with flow and volume, but a representative model could not be found. For this reason, the duration is estimated by fitting the characteristic values of the flood (peak flow and direct volume) to a selected shape, and the time to peak is defined as a random variable respect to the flood characteristics. These issues are discussed in the sub-chapter 5.3 along with some results.

5.2. Extreme value statistics

5.2.1. Univariate frequency analysis

As was mentioned in sub-chapter 3.3.1, a variety of distribution functions and parameter estimation methods were applied to decide which model best represents the peak flow and volume series. The results of the Cramer-von Mises (CvM) statistics are presented in the following table (see Table 5.5). The results indicated with bold-black letter are the ones that are not rejected with a 5% of confidence level, and the selected models are marked with a green background.

Table 5.5: Goodness of fit of univariate models applied to the series of peak flow and direct volume. Cramer-von Mises statistics.

Type of Event and Serie		Winter: Rain Events		Winter: Snow-melt Events	
Model					
Distribution Function	Method of Estimation	Total Peak Flow	Direct Volume	Total Peak Flow	Direct Volume
Normal	Maximum Likelihood	0.168	0.166	0.316	0.042
Normal	L-Moments	0.150	0.144	0.292	0.038
Exponential (1 parameter)	Moments	0.419	0.401	1.125	1.310
Exponential (2 parameters)	L-Moments	0.203	0.214	0.092	0.313
Gamma	Maximum Likelihood	0.078	0.160	0.075	0.060
Gamma	L-Moments	0.083	0.106	0.081	0.061
Weibull (2 parameters)	Maximum Likelihood	0.077	0.089	0.160	0.030
Weibull (3 parameters)	L-Moments	0.068	0.084	0.081	0.027
Log-Normal (2 parameters)	Maximum Likelihood	0.128	0.174	0.043	0.113
Log-Normal (3 parameters)	L-Moments	0.089	0.106	0.039	0.033
Generalized Extreme Value	L-Moments	0.086	0.099	0.030	0.032
Pearson 3	L-Moments	0.083	0.102	0.072	0.032
Gumbel	Maximum Likelihood	0.103	0.112	0.051	0.061
Gumbel	L-Moments	0.089	0.109	0.077	0.082
Generalized Pareto	L-Moments	0.030	0.034	0.076	0.049

The Kolmogorov Smirnov (KS) statistics resulted in values that indicate that most of the models are not rejected. These statistics are not presented because they do not give additional information for the selection of the appropriate model. The models rejected by the CvM statistic are discarded, and the rest of the models are compared using the QQ-plots.

The final selected models were inspected visually, with a special emphasis in the representation of the highest values. As an illustrative example the two best models representing the peak flow series (winter floods) are presented in the following figure (see Figure 5.10). These models show the lowest CvM statistics.

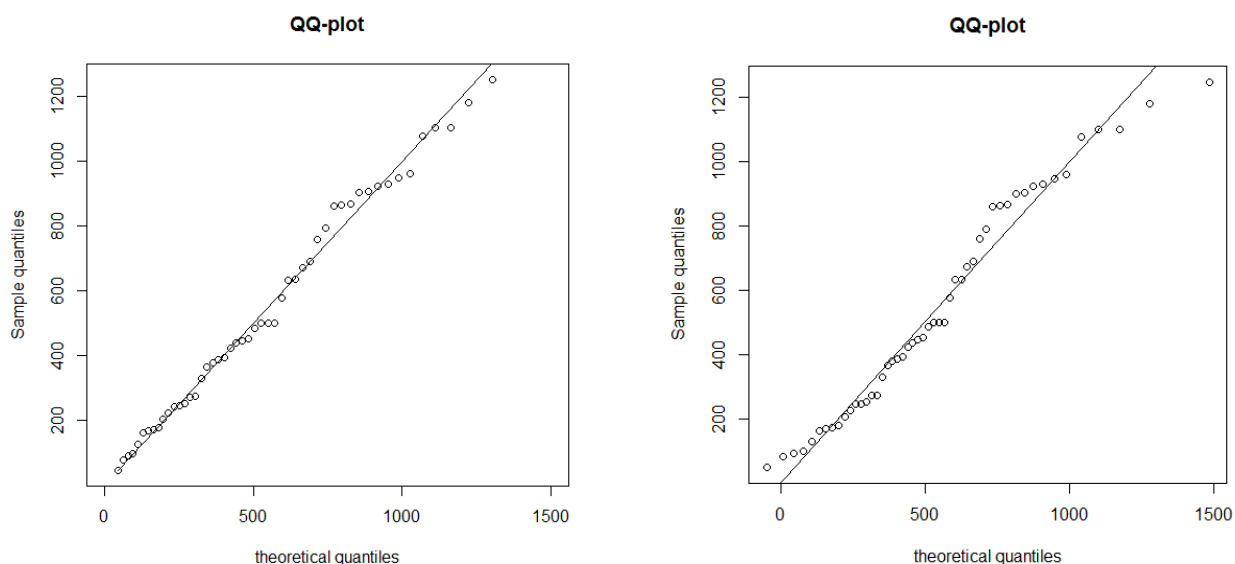


Figure 5.10: Comparison of two univariate models fitted to the peak flow series observed during the winter events. Left: Generalized Pareto model; Right: Weibull (3 parameters) model.

The two plots show the theoretical and the corresponding sample quantiles for each of the models. The Generalized Pareto model (see left image in Figure 5.10) provides a better representation of the observed values, especially the highest flows, for this reason this model was chosen to represent the population. A similar comparison was done for the rest of the models and series analyzed to select the best representations.

As indicated in Table 5.5, the selected models are the Generalized Pareto, for both series of flood characteristics observed during winter, the Generalized Extreme Value (GEV) for the peak flow series corresponding to the spring events, and the Weibull (3 parameters) for the direct volume series (spring events). The QQ-plots and the estimated parameters are presented in the following figures and table (see Figures 5.11 and 5.12, and Table 5.6).

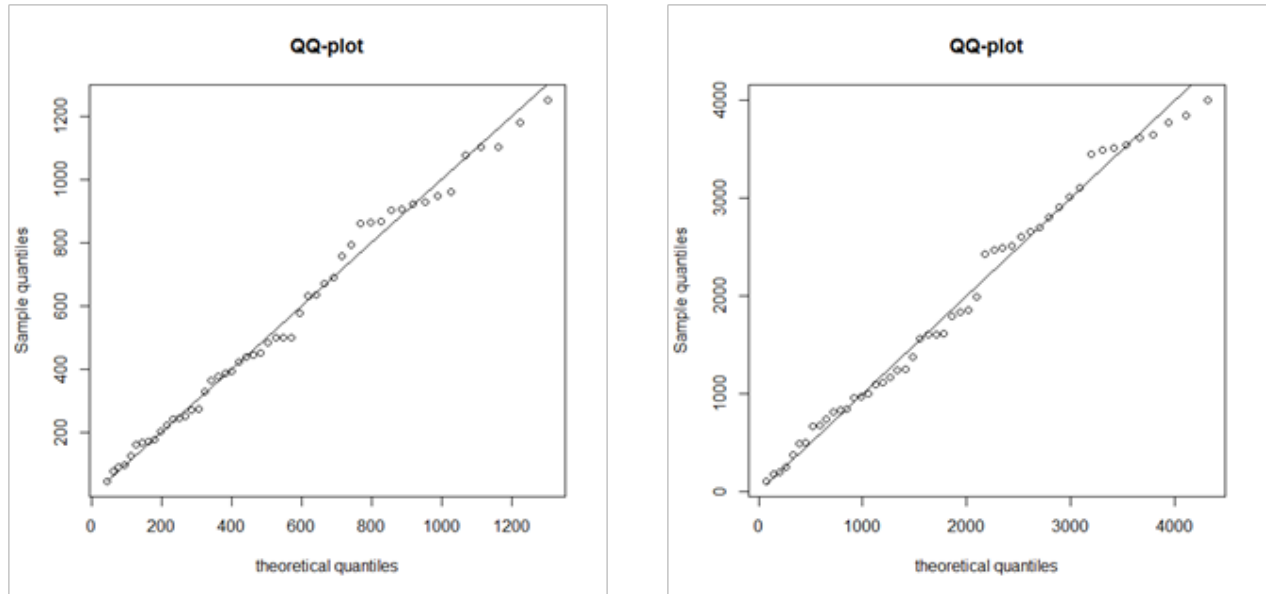


Figure 5.11: Univariate models selected to represent the flood characteristics of the winter events. Left: Generalized Pareto model fitted to the peak flows; Right: Generalized Pareto model fitted to the direct volumes.

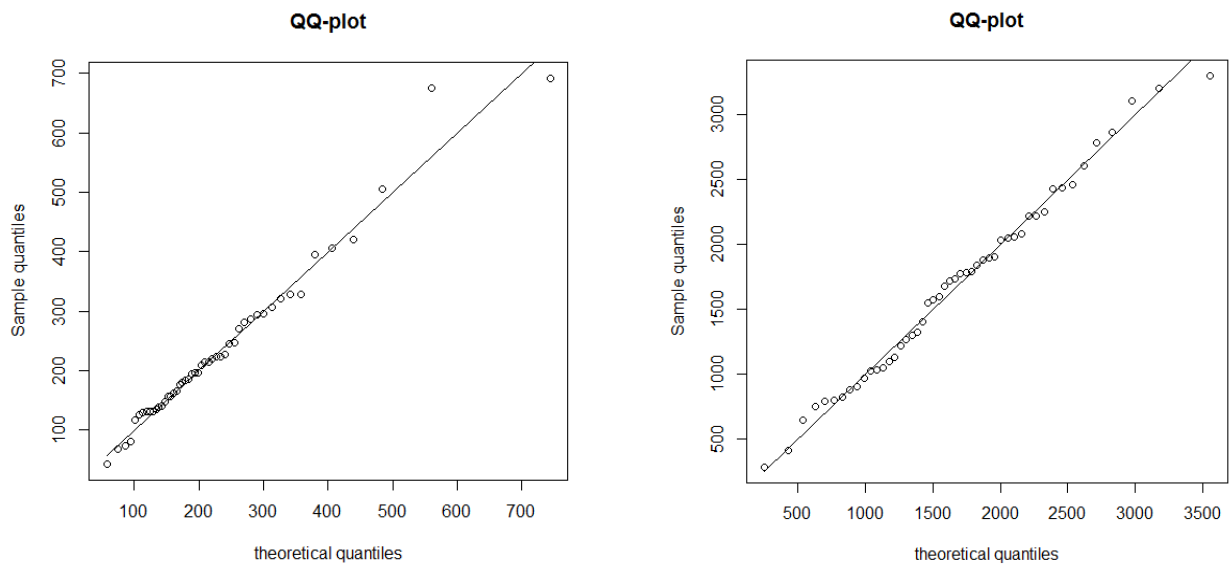


Figure 5.12: Univariate models selected to represent the flood characteristics of the spring events. Left: GEV model fitted to the peak flows; Right: Weibull (3 parameters) model fitted to the direct volumes.

Table 5.6: Estimated parameters of the selected univariate models representing the flood characteristic series: peak flow and direct volume.

Type of Event		Winter: Rain Events		Spring: Snow-melt Events	
Serie		Total Peak Flow	Direct Volume	Total Peak	Direct Volume
Model	Distribution	Generalized Pareto	Generalized Pareto	GEV	Weibull (3 parameters)
	Estimation Method	L-Moments	L-Moments	L-Moments	L-Moments
Parameters	Location	37.85	46.30	168.06	49.64
	Scale	802.96	3081.65	82.29	1942.13
	Shape	0.59	0.69	-0.17	2.46

These four models are used in two different ways: as four independent models for the univariate frequency analysis and as the marginal distributions for the bivariate frequency analysis. In the first case the models are used to randomly generate values of the four variables independently. In this case these models are also applied to estimate the magnitude of our variables for a given design value of return period. In the second case, the marginal distributions are used along with the Copula model to generate random pairs of values and possible combination of flood characteristics associated to the selected design value. The details and results of the second case are presented in the following section (see sub-chapter 5.2.2). In this section the results of the first case are presented, i.e. the application of the models for the univariate frequency analysis. It is important to mention that this analysis is done in order to compare the results with the bivariate approach.

The following graphs show the results of the synthetic peak flow and direct volume series, randomly generated from the univariate distribution functions (see Figure 5.13). A total number of 1000 values for each variable were generated. Note that even though the values are graphed as pairs of peak flow and volume, the generation of each of the series was performed independently of each other.

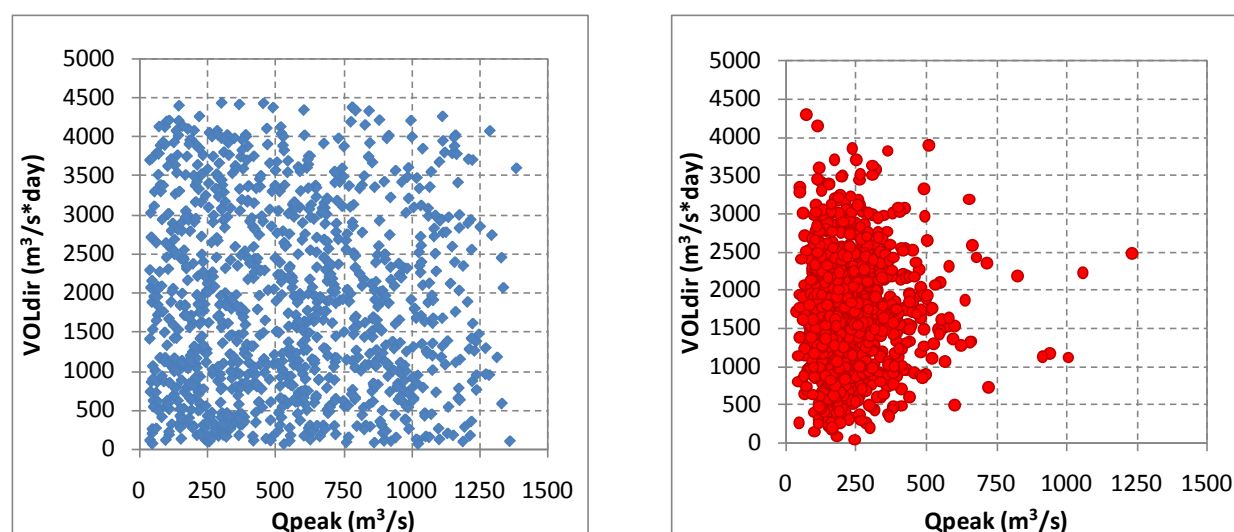


Figure 5.13: Synthetic series of peak flow and direct volume, random generation using the univariate distribution functions. Left: Winter events; Right: Spring events.

The randomly generated values show the difference in the behavior of the models. The Generalized Pareto models result in a more uniform distribution of the random generation, whereas the GEV model results in some values of spring peak flows that are very high compared with the maximum observation ($692 \text{ m}^3/\text{s}$). These synthetic series of peak flow and direct volume, are later combined with synthetic shapes. The combination results in synthetic hydrographs that are routed through the reservoir.

The last application of the univariate frequency analysis focuses on the estimation of the variables associated to a certain design criteria. The criteria used in the present work is the return period of 1000 years, and the values corresponding to the different models are presented in the section with the results of Return Period (see sub-chapter 5.2.3).

5.2.2. Bivariate frequency analysis using copula

In order to have an initial idea of the dependence, the scatter plots of the scaled ranks are plotted. They are presented in the following figure (see Figure 5.14).

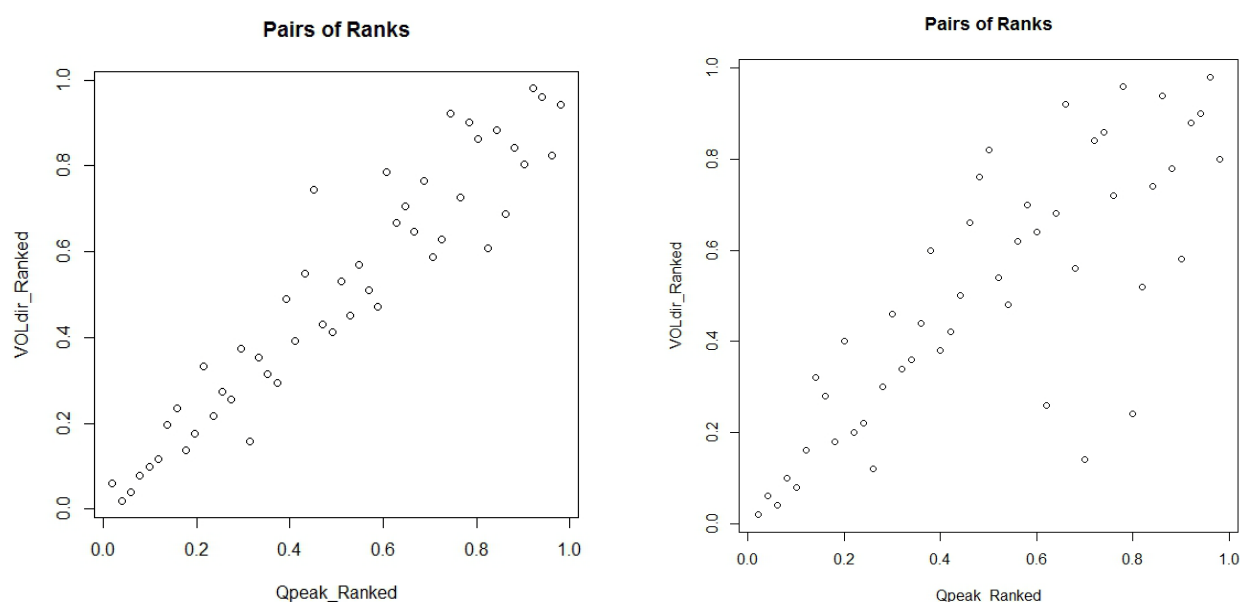


Figure 5.14: Scatter plots of scaled ranks of the pairs of observed values (peak flow, direct volume). Left: Winter events; Right: Spring events.

The graphs show the different behavior of the two analyzed series of pairs. The winter events show a conical behavior, with a linear relationship between low values of ranks and a more disperse relationship for the high pairs of ranks. The spring events show a more disperse behavior overall, that could also be characterized as conical, but with a wider opening, especially for high rank values of peak flow. The strength of the dependence is quantified using the Spearman's rho and Kendall's tau, the corresponding values are presented in the following table (see Table 5.7). The p-values associated to the Kendall's tau coefficient are included in the table. If the p-values are small then the alternative hypothesis is accepted, that is the true tau is not equal to 0, this coefficient of dependence is significantly different to zero. This is the result for both series, indicating a significant grade of dependence, stronger for the winter events characteristics.

Table 5.7: Measures of the degree of dependence corresponding to the analyzed series.

Type of Event	Winter: Rain Events		Spring: Snow-melt Events	
	Coefficient	p-value	Coefficient	p-value
Spearman's rho	0.945		0.796	
Kendall's tau	0.803	2.22e-16	0.650	4.918e-13

As mentioned in Section 3.3.2, six different copula models and two parameter estimation methods were applied to decide which model represents the characteristics of the analyzed floods in the best way. The results of the Cramer-von Mises (CvM) statistics and the associated p-values are presented in the table (see Table 5.8). The p-value represents the level at which the Copula is not rejected, meaning that models with higher p-values are better in terms of not rejecting them. The results indicated with bold-black letter are the selected models.

Table 5.8: Goodness of fit of bivariate copula models applied to the series of peak flow and direct volume. Cramer-von Mises statistics and associated p-value.

Type of Event		Winter: Rain Events		Spring: Snow-melt Events	
Model					
Copula	Method of Estimation	CvM	p-value	CvM	p-value
Normal	Moment-like: rho	0.015	0.223	0.022	0.109
Normal	MaxPLikelihood	0.020	0.163	0.028	0.083
Frank	Moment-like: rho	0.014	0.250	0.024	0.077
Frank	MaxPLikelihood	0.016	0.292	0.028	0.069
Gumbel	Moment-like: rho	0.020	0.098	0.028	0.023
Gumbel	MaxPLikelihood	0.042	0.021	0.049	0.007
Clayton	Moment-like: rho	0.012	0.367	0.025	0.101
Clayton	MaxPLikelihood	0.025	0.148	0.030	0.096
Galambos	Moment-like: rho	0.020	0.081	0.028	0.026
Galambos	MaxPLikelihood	0.043	0.017	0.052	0.005
Husler Reiss	Moment-like: rho	0.020	0.087	0.028	0.026
Husler Reiss	MaxPLikelihood	0.054	0.010	0.052	0.005

The models are compared by the CvM statistic and its associated p-value. All the models are used to generate random pairs of values which are graphed along with the observed data. The generated series and observed data are compared, with a special look into the models with highest p-values. This comparison is performed in order to visually decide whether the random pairs represent the observed ones, or in other words to decide if the model could reproduce the observed values in a satisfactory way.

The following graphs are presented in order to compare the two copula models that best represent the correlation of the spring events peak flow-volume series (see Figure 5.15). These models are the Normal and Clayton copula, with parameters estimated using the moment-like method.

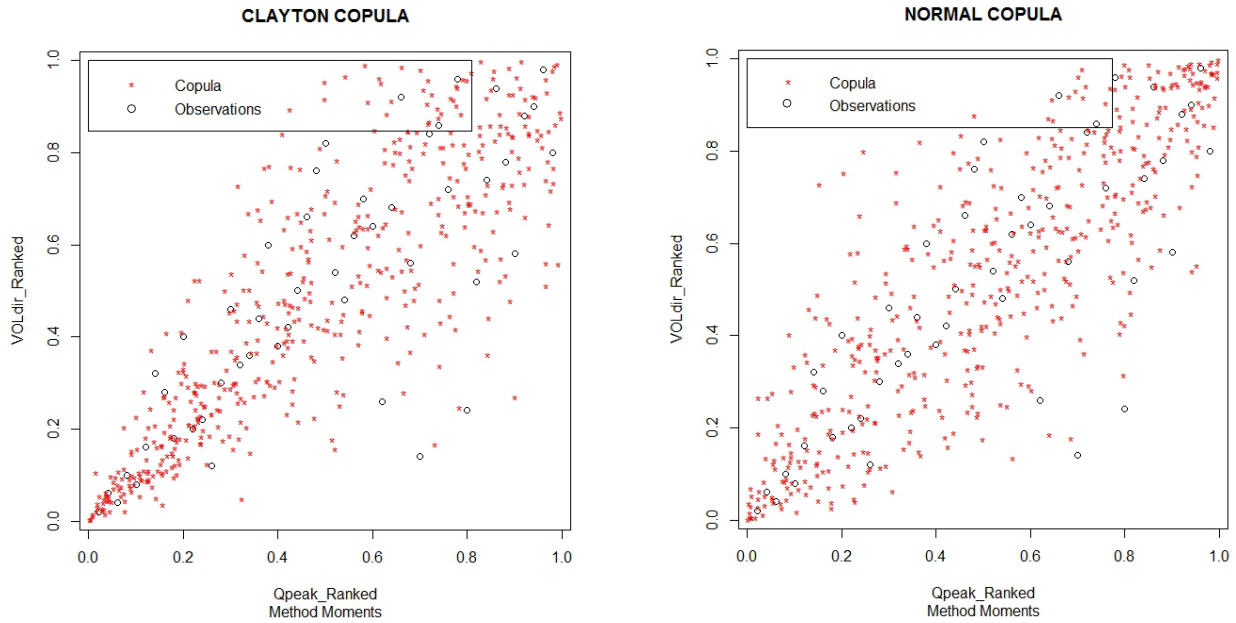


Figure 5.15: Ranked-pairs: observed during the spring events and generated using the two best copula models. Left: Clayton copula - Moments-like estimation. Right: Normal copula - Moments-like estimation

By comparing both figures, the Clayton copula reproduces better the observations. The Normal copula is discarded because it generates random variables that are more disperse respect the region of the observed pairs, especially in the upper left region, in which some random pairs are generated but there are no observations. Similar graphs are reproduced and analyzed for all the studied models in order to compare them and select the best one.

In order to compare between the two methods of parameter estimation, the following graphs are presented (see Figure 5.16). The graphs show the Clayton copula, representing the correlation of the winter events. The parameters estimated using the pseudo-likelihood method is 5.931, whereas the parameter estimated using the moment-like method is 8.124.

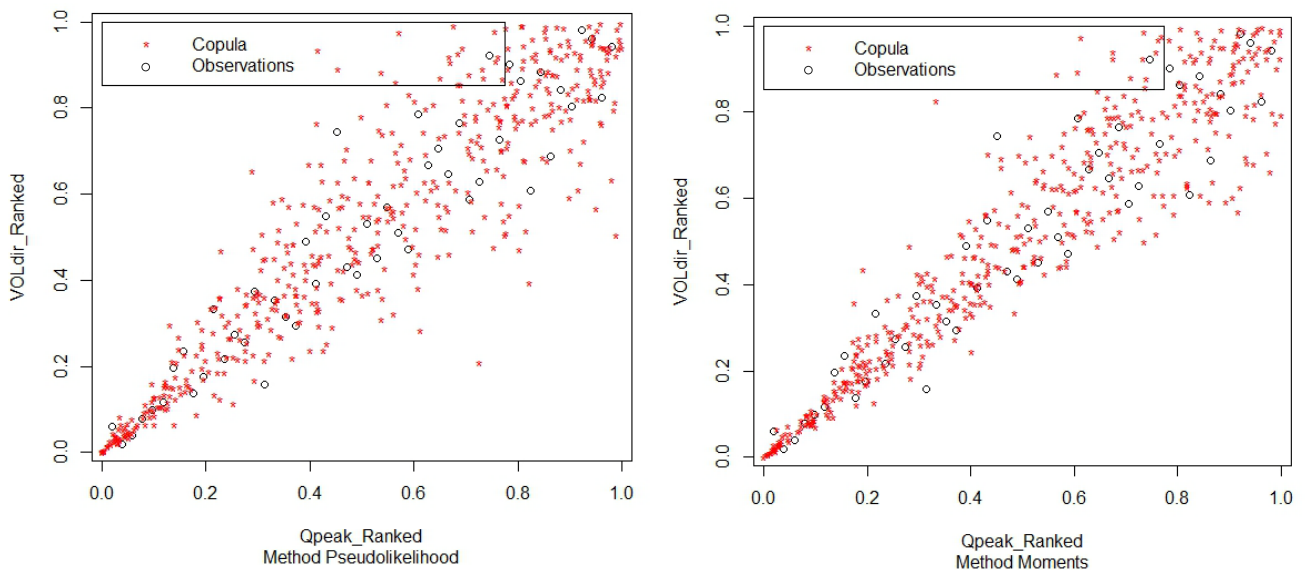


Figure 5.16: Ranked-pairs: observed during the winter events and generated using the same copula and different estimation methods. Left: Clayton copula – Pseudolikelihood estimation. Right: Clayton copula - Moments-like estimation

The comparison of both figures shows that one of the two models, the one that was estimated using the pseudo-likelihood method (left figure), generates random variables that are more disperse respect the region of the observed pairs. The Clayton copula with the parameter estimated using the moment-like method best represents the correlation of the observed pairs.

The models that best fit the two analyzed flood series are the Clayton with a parameter of 8.124 for the winter flood events, and the Clayton with a parameter of 3.709 for the spring events. Both parameters are estimated using the moment-like method. The following graphs show the synthetic series of the peak flow and direct volume, randomly generated using the best copula models (see Figure 5.17). A total number of 1000 pairs of variables are generated for each flood type (winter and spring).

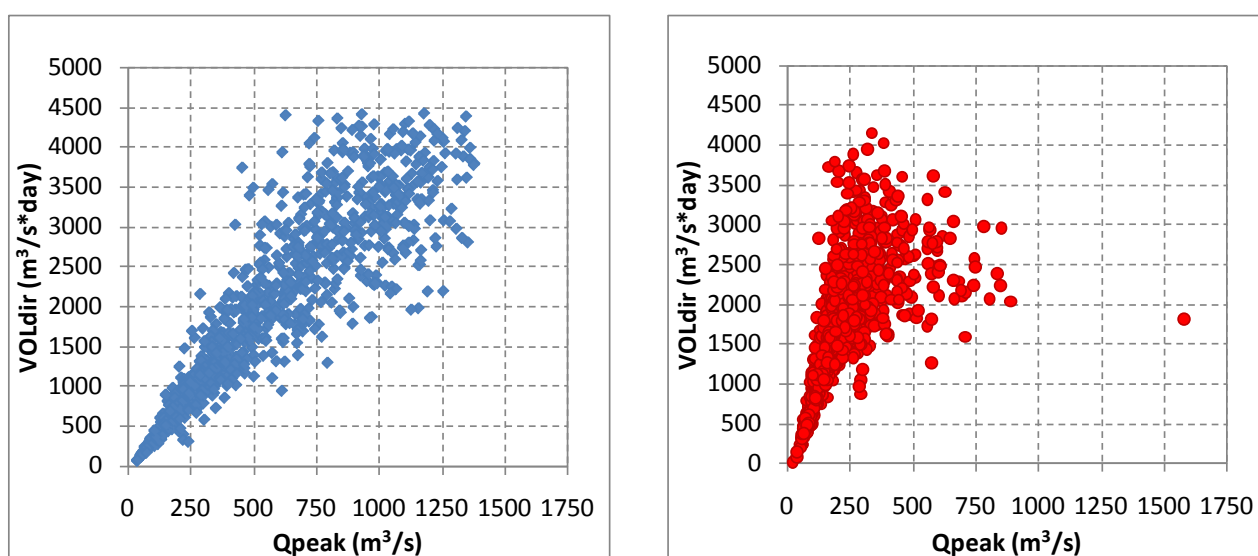


Figure 5.17: Synthetic series of peak flow and direct volume, random generation of pairs of values using the copula models. Left: Winter events; Right: Spring events.

The figures show that the model representing the winter events is more concentrated in the main diagonal compared with the model representing the spring events, which is more disperse. This difference of behavior between the two models coincides with the values of the correlations between the two analyzed variables, which was higher for the winter events than for the spring events (see Table 5.2). It is important to mention that the behavior of the marginal distributions corresponding to each of the variables is also shown in these graphs. The random pairs of values randomly generated by the copula models was then transformed to pairs of peak flow and direct volume values using each of the marginal distributions. As in the results obtained by the univariate analysis, the GEV model fitted to the peak flows of the spring events results in some values that are very high compared with the maximum observation ($692 \text{ m}^3/\text{s}$). These synthetic series of peak flow and direct volume, are latter combined with synthetic shapes to generate hydrographs to be routed through the reservoir.

The last application of the bivariate frequency analysis is the estimation of the variable magnitudes associated to a certain design criteria. The criterion used in the present work is the return period of 1000 years, and the values corresponding to the different models are presented in the section with the results of Return Period (see sub-chapter 5.2.3).

5.2.3. Return period

The last application of the univariate and bivariate frequency analysis is the estimation of the variables magnitudes associated to a certain design criteria. The criterion used in the present work is the return period of 1000 years. The associated values estimated using the univariate models are presented in the following table (see Table 5.9).

Table 5.9: Estimated variables associated to a design criteria of 1000 years return period.

Type of Event	Winter: Rain Events		Spring: Snow-melt Events	
	Total Peak Flow (m ³ /s)	Direct Volume (m ³ /s*day)	Total Peak Flow (m ³ /s)	Direct Volume (m ³ /s*day)
Estimation	1369	4464	1266	4210
% above maximum observation	10%	12%	83%	28%

The estimated values indicate the behavior of the different models. Note that the estimated magnitudes for the winter events (both Generalized Pareto models) are only 10 or 12% higher than the maximum observations, whereas the GEV gives a value 83% higher than the maximum peak flow. The pairs of univariate estimation of peak flow and direct volume associated to a given design return period, are latter combined with synthetic shapes to generate hydrographs to be routed through the reservoir.

A similar analysis is performed for the bivariate models, but as was explained previously, in this case the solution is not unique, due to the fact that different realizations of the variables may yield the same return period. In the following graphs (see Figure 5.18), the different combinations of peak flow and volume, associated with a return period of 1000 years are presented. These values are estimated considering the “OR” case joint return period. The pairs of values obtained by the univariate frequency analysis (see Table 5.9) are also plotted in the graphs. Note that these values are the limiting values of the asymptotes, i.e. very large values of peak flow have the same associated direct volume, corresponding to the univariate estimation, and the same for very large values of direct volume with the peak flow estimated using the univariate analysis.

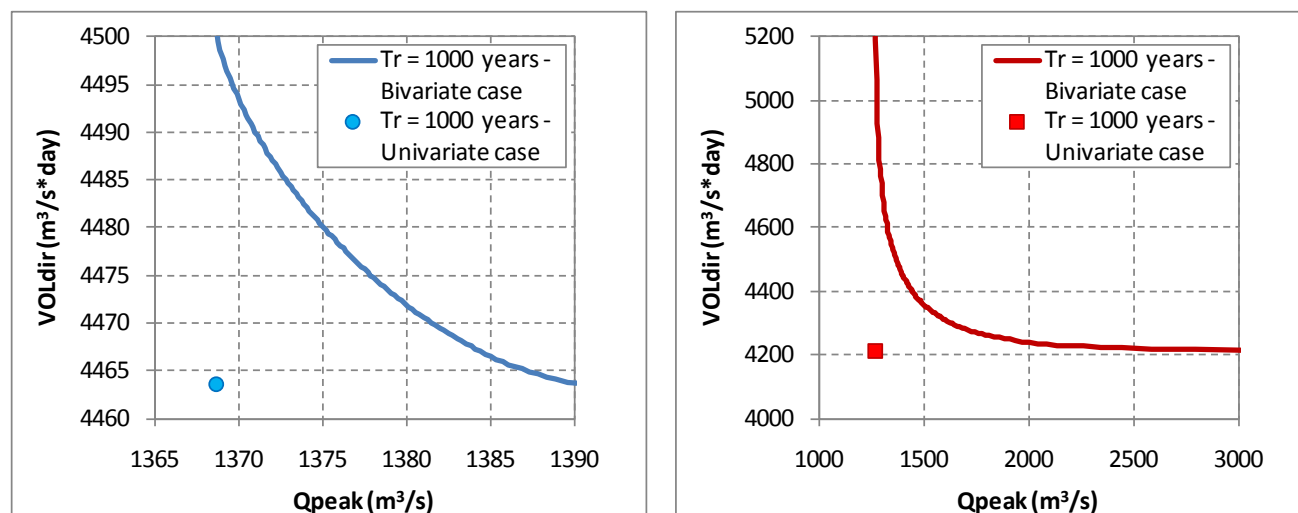


Figure 5.18: Pairs of peak flow and direct volume values associated to the design return period of 1000 years, estimated by univariate and bivariate analysis. Left: Winter events; Right: Spring events.

All the estimated pairs of values associated to the design return period, are latter combined with the selected critical shapes to generate synthetic hydrographs that are routed through the reservoir.

5.3. Synthetic hydrograph

5.3.1. Analysis of the observed hydrographs

The independent flood events are identified from the observed series of daily flows, to obtain the different flood characteristics: peak flow (total, direct and base), volume (total, direct and base), duration and time to peak. The part of the flood corresponding to the direct flows is studied to characterize the shape of the event. The Beta distribution, which is defined by two parameters, is used to represent the shape of the direct hydrographs. In order to represent the statistical properties of the shape variables, it is necessary to determine the series of the shape parameters from the available observed floods. The steps and results obtained are presented in this section.

The Beta distribution function has a base and total area equal to one, so the observed flood hydrographs must be modified to meet these conditions. This is done by the base of each flood hydrograph by its total duration and multiplying each ordinate of the observed flow by the ratio between the duration and the total direct volume. After this transformation, each of the new direct hydrographs has a unitary duration and a total direct volume of one.

The transformation of the direct hydrographs results in series of parameters characterizing the shape of the different hydrographs. These parameters are the adimensional time to peak (t_{peak}^*), the adimensional peak flow (Q_{peak}^*), and the parameters of the Beta distribution function (a and b). The time to peak is the ration of the occurrence of the peak flow respect the total duration of the flood. This value gives an idea of the shape of the hydrograph regarding the position of the peak. A value lower than 0.5 refers to a hydrograph with a prior-peak shape (or positively skewed Beta distribution), a value equals to 0.5 is a symmetrical or mid-peak shape, and a value higher refers to a posterior-peak shape hydrograph. The adimensional peak flow gives an idea of the distribution of the volume around the peak. As the total volume is equal to one for all the hydrographs, one with a high value of adimensional peak flow indicates that it has a shape in which the volume of the flood is more concentrated towards the peak, compared with another hydrograph with a lower

adimensional peak flow. This concept can be better understood with a numerical example extracted from the observed hydrographs (See Figure 5.19). The parameters of the Beta distribution (a and b) depend on the values of the time to peak and adimensional peak flow. Some sample statistics of the obtained series are summarized in the following table (See Table 5.10). The statistics of the Beta parameters are also included because they are later used for the synthesis of hydrograph shapes.

Table 5.10: Sample statistics of the shape parameter series: Adimensional time to peak (t_{peak}^*), adimensional peak flow (Q_{peak}^*) and the Beta parameters (a and b).

Type of Event	Winter: Rain Events				Spring: Snow-melt Events			
	t_{peak}^*	Q_{peak}^*	a	b	t_{peak}	Q_{peak}	a	b
Mean	0.388	6.11	24.16	25.99	0.252	4.27	7.35	14.39
Standard Deviation	0.195	2.16	24.96	13.27	0.149	1.85	8.73	12.83
Coefficient of Variation	0.503	0.35	1.03	0.51	0.593	0.43	1.19	0.89
Coefficient of Skewness	0.215	0.85	1.72	0.74	1.136	1.20	2.16	2.24
Coefficient of Kurtosis	-0.607	1.08	3.29	0.26	1.371	1.51	4.55	6.06
Maximum	0.826	13.32	119.10	60.66	0.708	10.30	38.51	67.49
Minimum	0.065	2.89	1.32	4.49	0.033	1.45	1.11	1.61
Median	0.389	6.04	15.02	24.46	0.237	3.83	3.71	10.16
Interquartile Range	0.266	2.92	31.28	16.18	0.186	2.20	6.30	13.42

These values indicate that overall the spring flood events are characterized with earlier peaks compared with the winter events, resulting in lower values of adimensional time to peak. The winter events have flood shapes that are characterized mainly with more concentrated volumes towards the peak flow (higher values of adimensional peak flow) compared with the spring flood events.

In order to show the effect of the magnitude of the adimensional peak flow, the shapes with the three mentioned magnitudes (minimum, median and maximum) are presented in the following figure (see Figure 5.19). These adimensional hydrographs are derived from the winter flood events observed in the years: 1963, 1984 and 2001, and their adimensional peak flows are approximately 3, 6 and higher than 13 respectively. As can be seen from the figure, the magnitude of the adimensional peak flow is related with the shape of the hydrograph in terms of the distribution of the volume around the peak flow. The observed flood in the winter of 2001 shows an important concentration of the volume around the peak flow, resulting in a high value of adimensional peak flow, whereas the flood observed in 1963, has a higher distribution of the volume along the time of the occurrence of the flood, resulting in a four times smaller value of adimensional peak flow.

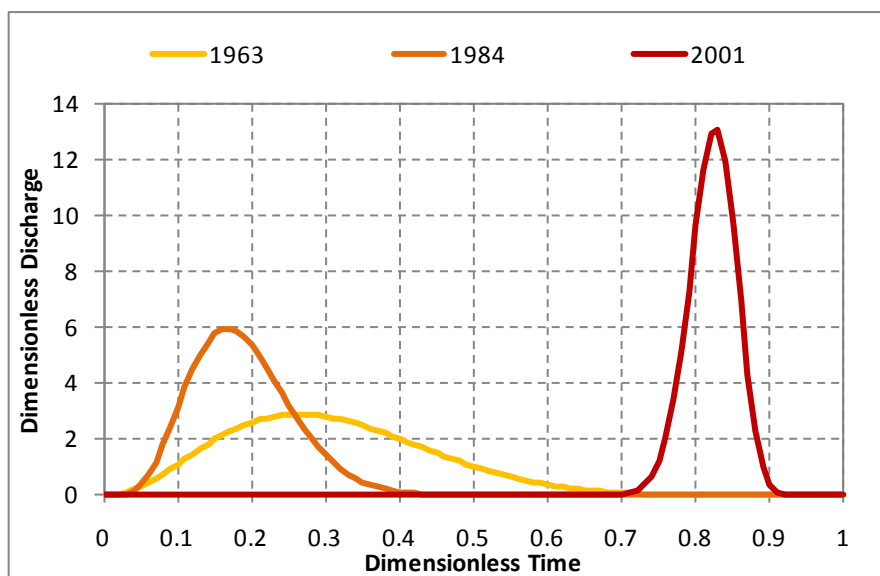


Figure 5.19: Beta distributions representing the transformed hydrographs of the three extreme winter events in terms of the adimensional peak flows with values: ≈ 3 (1963), ≈ 6 (1984) and > 13 (2001)

The Beta distribution is a good representation of the direct flood hydrographs, especially when the floods are characterized with one dominant peak, due to the fact that it is a single peak probability distribution function. This is shown in the following figures, in which the direct hydrographs of the extreme flood events are presented along with their Beta representation. The extreme floods are chosen in terms of maximum peak flow and maximum direct volume (see Figures 5.20 and 5.21). Note that to plot this graphs, the Beta distributions are modified so that the base of the hydrograph represents the total duration of the flood and the area is equal to the observed direct volume. This modification is done by multiplying the base of the Beta times the duration of the flood, and each ordinate by the ratio of direct volume over duration. The difference between the Beta distribution and its modification can be seen by comparing the figures 5.19 and 5.20, in which the representation of the event observed in the winter of 2001 are plotted both as the Beta distribution and its modified version.

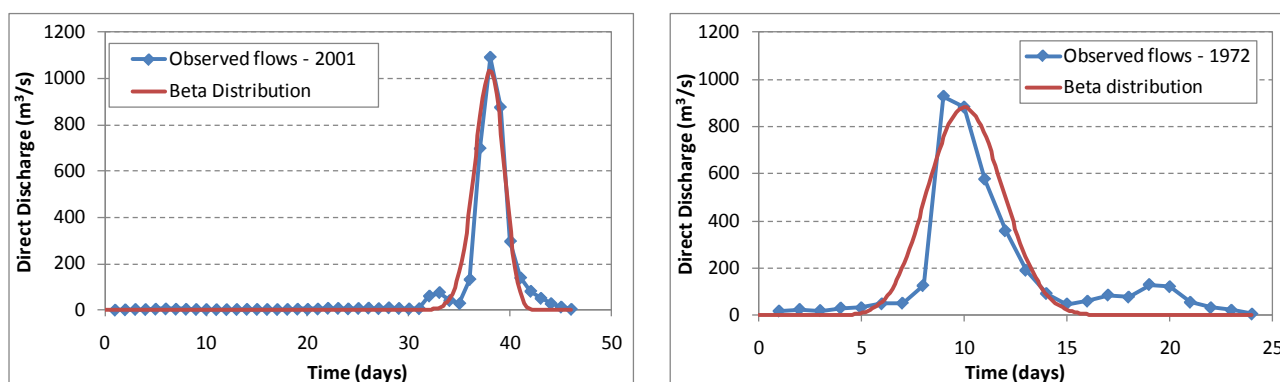


Figure 5.20: Application of the Beta distribution to reproduce the maximum winter flood events. Left: Maximum peak flow, 2001; Right: Maximum direct volume, 1972.

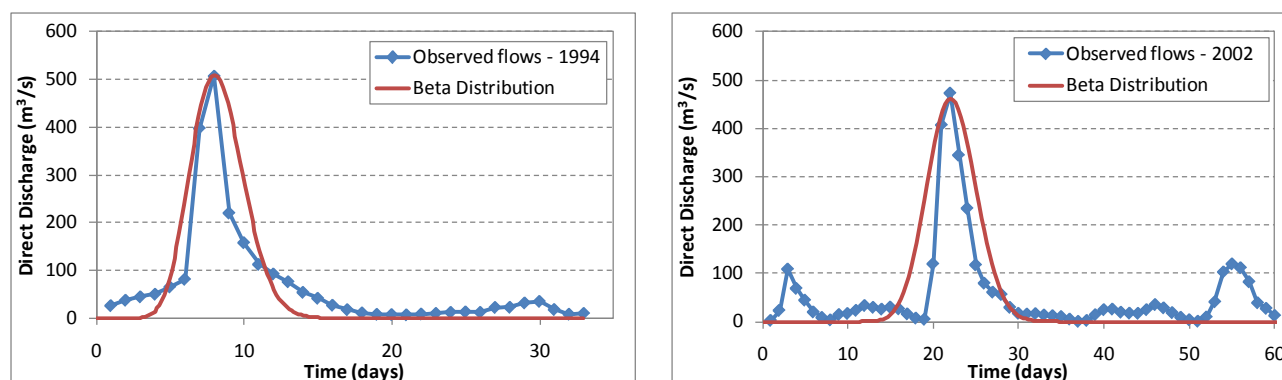


Figure 5.21: Application of the Beta distribution to reproduce the maximum spring flood events. Left: Maximum peak flow, 1994; Right: Maximum direct volume, 2002.

These graphs show that the Beta distribution reproduces the observed extreme events in a satisfactory way. The series of parameters characterizing the shapes of the observed floods are then used to derive models from which random shapes of hydrographs can be generated. This is done by statistically analyzing the four series of the Beta parameters (a and b), finding the best univariate distribution to represent each of the series, and finally finding the best Copula model to represent the correlation between the pairs of values for the winter and spring events. The univariate and copula functions and methods of estimating the parameters are the same as the ones used for the frequency analysis of the series of peak flows and direct volumes. The results obtained and models selected are summarized in the following table (see Table 5.11).

Table 5.11: Selected univariate and copula models representing the shape parameters: a and b (Beta distribution parameters).

Type of Event		Winter: Rain Events		Spring: Snow-melt Events	
Serie		a	b	a	b
Model	Univariate Distribution	Pearson (3 parameters)	GEV	Weibull (3 parameters)	Log-Normal (3 parameters)
	Copula model	Normal		Gumbel	

The selected copula models and their capacity to represent the observed parameters are shown in the following figures (see Figure 5.22). The graphs show that the relationship between the shape parameters of the winter events (left graph) is weaker, compared to the one of the spring events, which is closer to the main diagonal. This is in agreement with the correlation coefficient which is lower for the winter events with Kendall's tau of 0.45 against 0.63 for spring events.

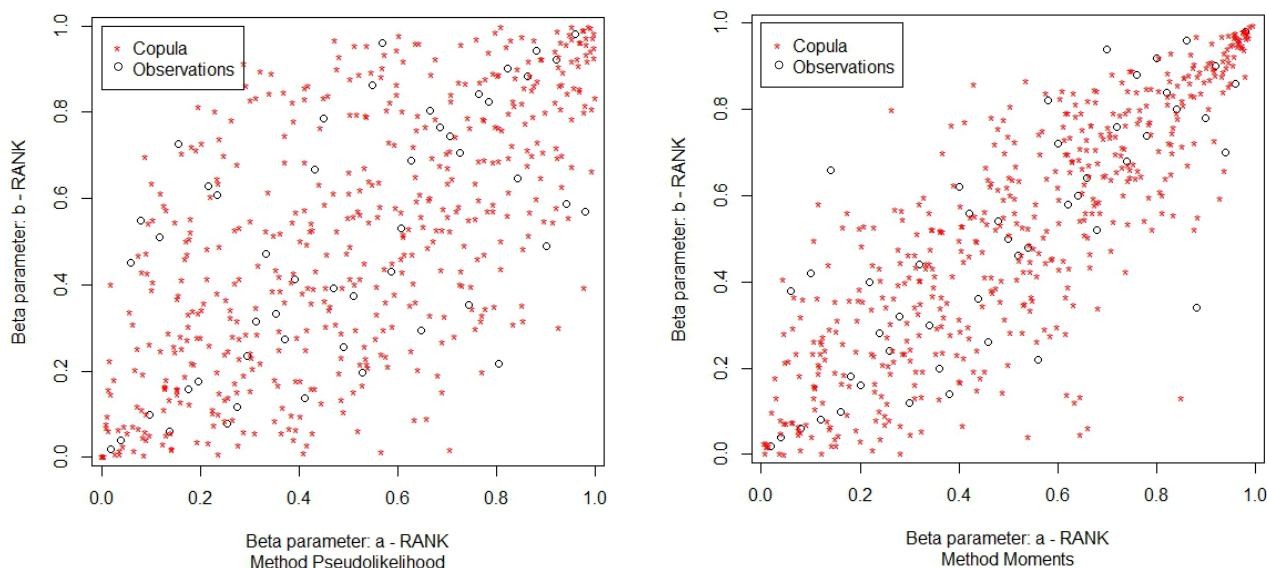


Figure 5.22: Ranked-pairs: observed and generated using the best copula models. Left: Normal copula, winter events. Right: Gumbel copula, spring events.

The models selected to represent the statistical properties of the series are used to generate random values of the parameters and then synthetic series of hydrograph shape. The following graphs show the synthetic series of Beta parameters, along with the ones characterizing the observed floods (see Figures 5.23 and 5.24). The values are randomly generated using the univariate models independently (left images) and considering the correlation between the two variables (right images). More than 1000 pairs of values are generated, and only the ones resulting in logic values of adimensional time to peak, i.e. values between 0 and 1, are selected. A total number of 1000 pairs of variables are selected for each flood type (winter and spring), shown in the following graphs.

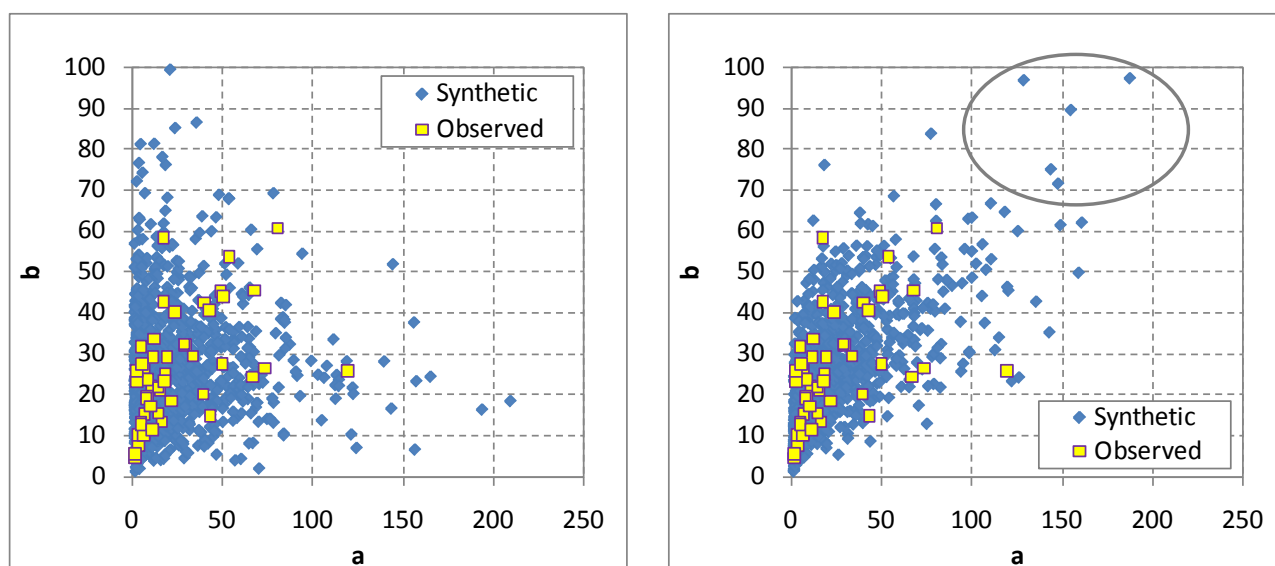


Figure 5.23: Observed and synthetic series of Beta parameters (a and b) corresponding to the winter floods. Left: Univariate analysis; Right: Bivariate analysis.

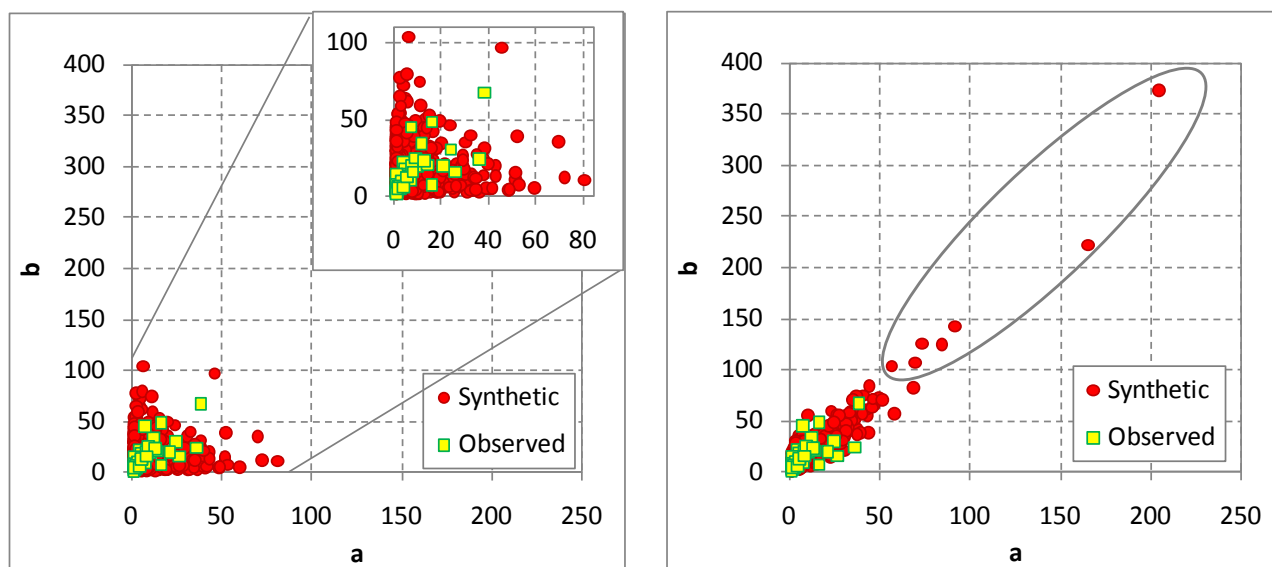


Figure 5.24: Observed and synthetic series of Beta parameters (a and b) corresponding to the spring floods. Left: Univariate analysis; Right: Bivariate analysis.

The random samples generated using the bivariate models show a concentration of points towards the main diagonal, while the univariate models result in a disperse distribution of synthetic pairs with respect to the diagonal, and more concentrated towards the lower left corner. As a consequence of this behavior, the univariate models are not able to reproduce higher values of both variables, especially in the case of the spring events. In comparison the bivariate models are able to generate this kind of pairs, which are highlighted with circles in the previous corresponding figures (see Figures 5.23 and 5.24).

Investigation of the generated pairs of high values (circled in the previous figures) was conducted to decide whether they represent possible shapes. The five pairs of points identified for the winter events have an adimensional time to peak of around 0.4. The adimensional time to peak was found to be in the order of the highest value of the observed events. The seven pairs belonging to the group of the spring events have an adimensional time to peak of around 0.6, and five of the points result in peak flows that are around the value of the highest observed one. The other two pairs show adimensional peak flows of around 1.5 and 2 times bigger than the highest observed. These results indicate that the synthetic floods would have shapes in which the volume is more concentrated towards the peak compared with the observed floods. This concentration could be the result of a natural phenomenon, like a very short and intense rain event, or the simultaneous occurrence of a heavy storm with high amounts of melted snow. As the available information has 50 observed floods, and these particular cases are 2 out of 1000, they are accepted as possible.

To visualize the physical meaning of all the synthesized pairs of parameters in terms of shape, the time to peak and peak flow of the adimensional hydrographs are calculated. The results obtained, along with the ones corresponding to the observed floods, are presented in the following graphs (see Figures 5.25 and 5.26).

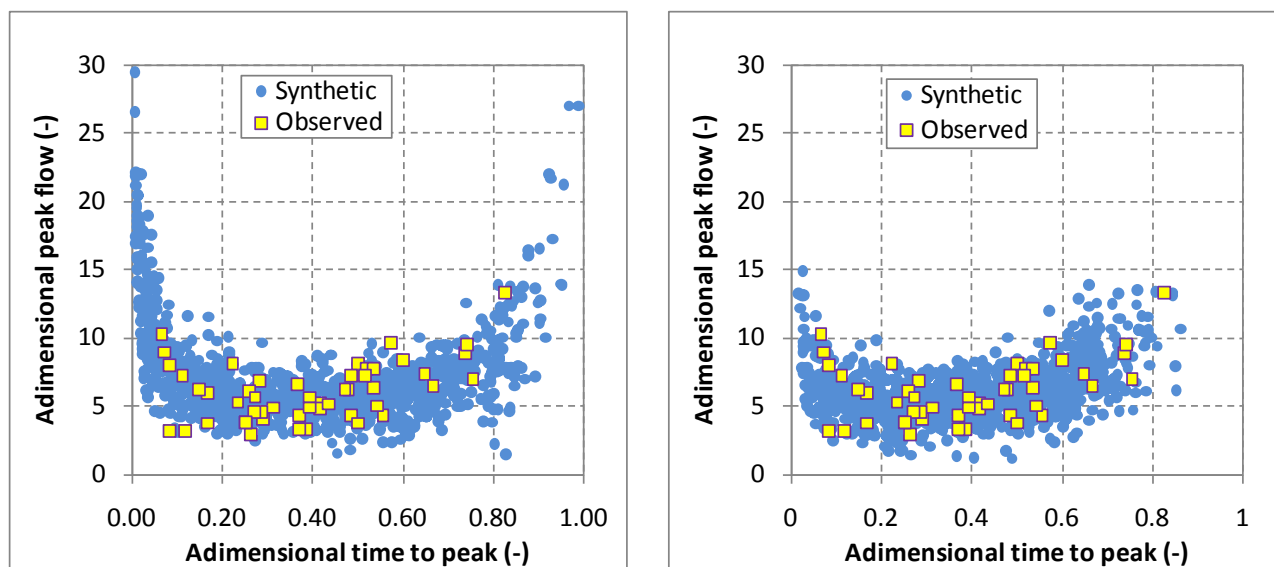


Figure 5.25: Observed and synthetic series of adimensional time to peak and peak flow corresponding to the winter floods. Left: Univariate analysis; Right: Bivariate analysis.

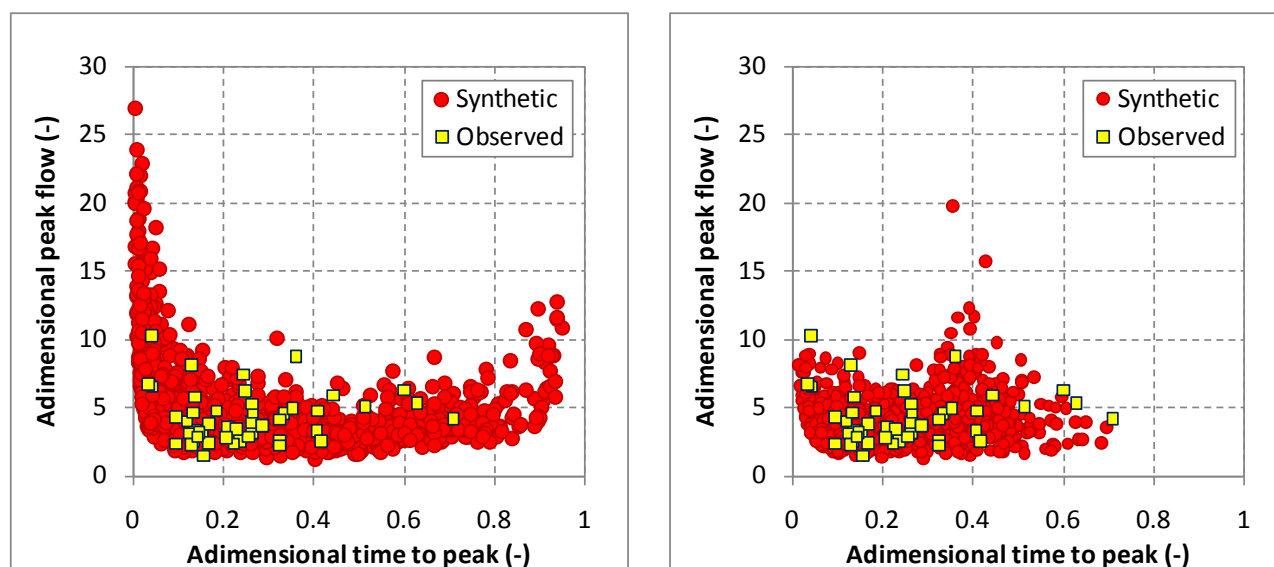


Figure 5.26: Observed and synthetic series of adimensional time to peak and peak flow corresponding to the spring floods. Left: Univariate analysis; Right: Bivariate analysis.

The figures show that the observed pairs of peak flow and time to peak are better reproduced by the bivariate models. The univariate models generate values of adimensional peak flows that are different than the observed range, especially for low or high values of times to peak.

The synthetic shapes derived from any of the models presented in this section are used to generate synthetic hydrographs. In order to do this, the magnitudes of peak flow and direct volume are necessary. The generation of synthetic hydrographs combining random pairs of peak flow and volume with random shapes is presented in the following section.

5.3.2. Synthesis of hydrographs

As explained previously, in order to generate synthetic floods the random values characterizing the floods, i.e. peak flow and direct volume, must be combined with the random shapes of hydrographs.

It was mentioned (see sub-chapter 3.4) that the combination of the flood characteristics with the shape of the hydrograph requires an adjustment of the duration of the event in order to reproduce the exact values of peak flow and volume. Two different cases are considered for a better understanding of this procedure. The first case is the combination of one pair of flood characteristics with three different shapes. The second case is the combination of three different flood characteristics with one shape. The total hydrographs of all these cases are presented in the following figures (see Figures 5.27 and 5.28). The dashed lines represent the discharges belonging to the base flow.

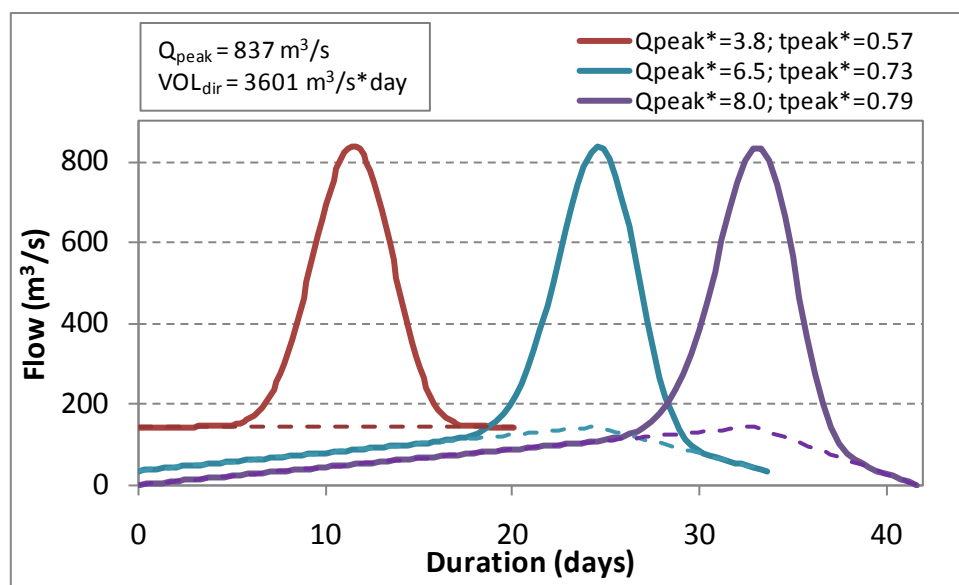


Figure 5.27: Synthetic hydrographs generated with one pair of flood characteristics and different hydrograph shapes.

The three hydrographs have the same magnitude of peak flow and direct volume. The durations for each of the cases were estimated in order to reproduce these exact values with the different shapes, resulting in: 20.0, 33.6 and 41.6 days respectively. The base flow is adjusted according to these durations, as was explained previously (see sub-chapter 3.4). In these examples, the base flow volume estimated using the direct volume is exactly reproduced by the two longest events and is underestimated in the shorter event, but the base volume of this event is only 3% lower than the other two cases.

In the following hydrographs, one shape of flood was adjusted to three different pairs of flood characteristics. Once again, the durations were estimated for each of the cases to reproduce the exact values of the flood characteristics. In this case the resulting durations are 38.1, 30.7 and 27.8 days respectively.

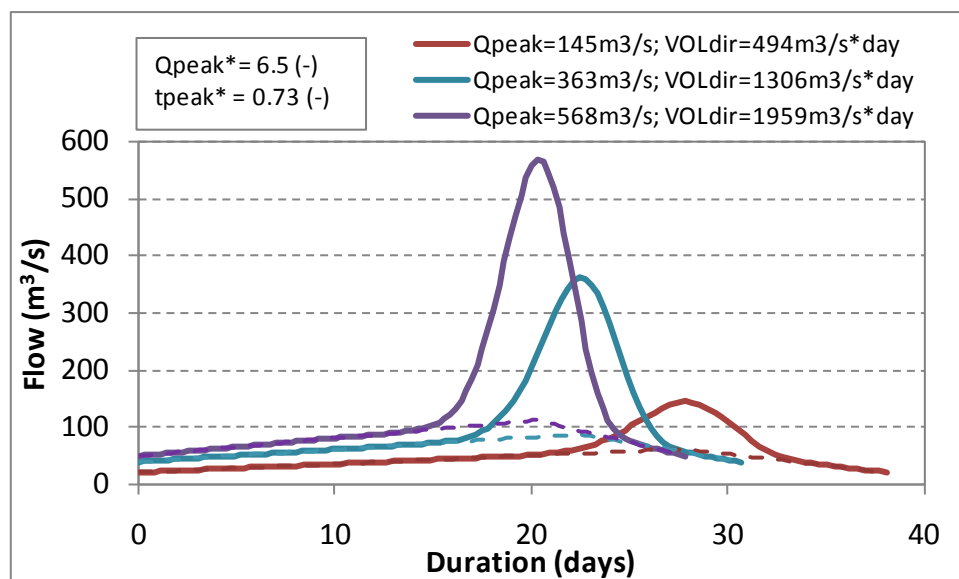


Figure 5.28: Synthetic hydrographs generated with one hydrograph shape and different combinations of flood characteristics.

The figures show how the duration of the flood depends on both the magnitude of the flood characteristics and the shape of the hydrograph.

In a similar way, all the random pairs of flood characteristics are combined with all the random shapes of hydrographs, and the synthetic hydrographs are obtained. It is important to make clear that the peak flow and direct volume, magnitudes characterizing the flood event, are in every case exactly reproduced. The base flow is then adjusted according to the duration of the event, and could be under/overestimated or exactly reproduced.

It was mentioned previously that the design procedure proposed in this work requires the identification of “critical” shapes. These shapes are defined as critical in terms of the purpose of the design. In this work a reservoir is under analysis, so the design would involve defining the height of the dam and the size and shape of the spillway, among others. The elevation of the water in the reservoir, which is associated with the magnitude of the spilled flows, is necessary for the design of these structures. The “critical” shapes are here defined as the shapes that result in the highest water elevations when they are routed through the reservoir. Yue et al. (2002) state that in order to get the same degree of protection, two flood events with the same peak and volume but different shapes, require different reservoir volumes. A flood event with a posterior peak shape requires much more storage volume of the reservoir, compared with a prior-peak shape event.

Following the argument discussed in the previous paragraph, the shapes with highest times to peak are considered to be critical in terms of maximum water elevations. To verify this statement all the random shapes are combined with one pair of flood characteristics, resulting in a set of synthetic hydrographs that are routed through the reservoir. All these synthetic hydrographs represent the same flood characteristics, which are distributed along the time of the event in different ways according to the random shapes. The maximum levels reached by each of the routings are identified, and then plotted along with the adimensional times to peak characterizing each shape. This procedure is done for the winter and spring events separately. The results indicate that the floods with shapes characterized by a delayed time to peak reach higher levels when they are routed through the reservoir. These results are presented in the following graphs (see Figure 5.29).

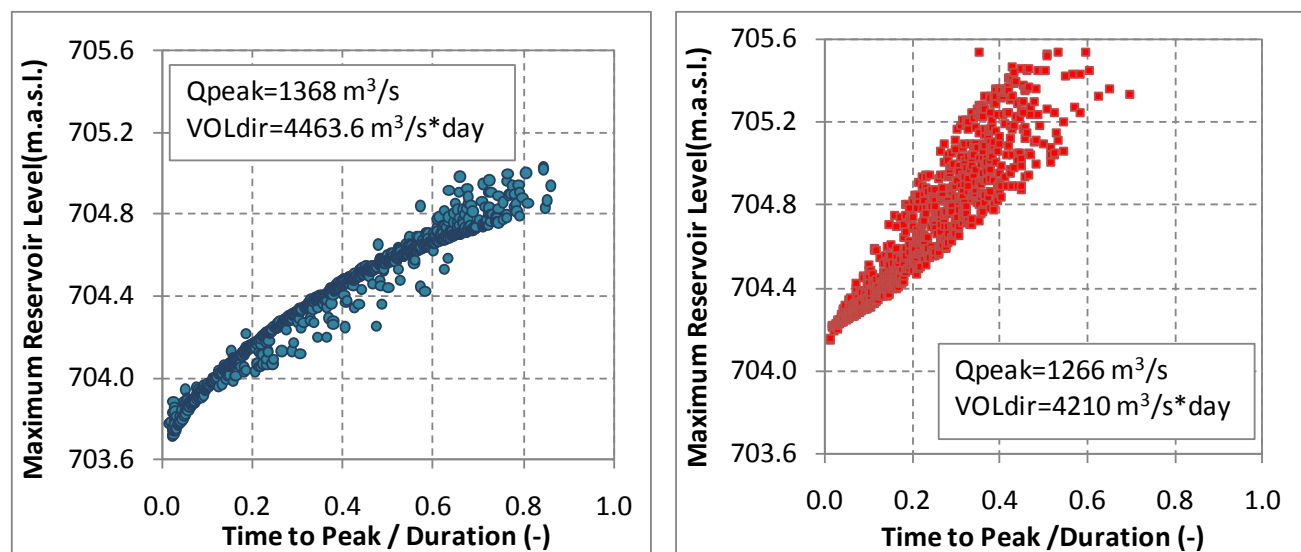


Figure 5.29: Relationship of the adimensional time to peak and the maximum water elevations reached in the reservoir. Left: Winter events; Right: Spring events.

In this work, the “critical” shapes used in the design procedure are chosen as the ones that have the highest adimensional times to peak. The 1000 random shapes are compared, and a total of 50 shapes with the highest times to peak are extracted to be used for the generation of the synthetic hydrographs in the design procedure. This is done for the winter and spring events separately.

5.4. Reservoir routing

In this section the results of the reservoir routing are presented. As mentioned in the previous chapters, the different reservoir simulations are run considering a constant discharge coefficient and in all the cases the same initial reservoir level equal to the spillway crest. The simulations are performed for a total time of 1.3 times the duration of the flood, and the time step for the simulations is of 1% of the total duration.

Each synthetic flood is routed, resulting in a series of water levels in the reservoir and outflows. The maximum values of water level and outflow are identified for each of the routed hydrographs. The results are presented as empirical distributions of the maximum reservoir level reached during the synthetic flood event. The distributions corresponding to the univariate and bivariate analysis are plotted together in order to compare the results obtained by the two approaches. The empirical frequencies of the maximum outflows are also presented.

This section is divided into two subsections in order to present the results obtained for the risk assessment for dam safety and the proposed design procedure separately.

5.4.1. Risk assessment for dam safety

The results obtained regarding the study of the adequacy of the spillway performance, in terms of maximum water levels reached with the 1000 synthetic flood events, are presented in the following graphs (see Figures 5.30 and 5.31). The synthetic flood events are the result of combining the 1000 random pairs of flood characteristics, obtained considering the univariate and bivariate models, with the 1000 random shapes, also obtained with the univariate and bivariate approaches respectively. The graphs include a zoom to the region of 95% of frequency for a better visualization of the differences.

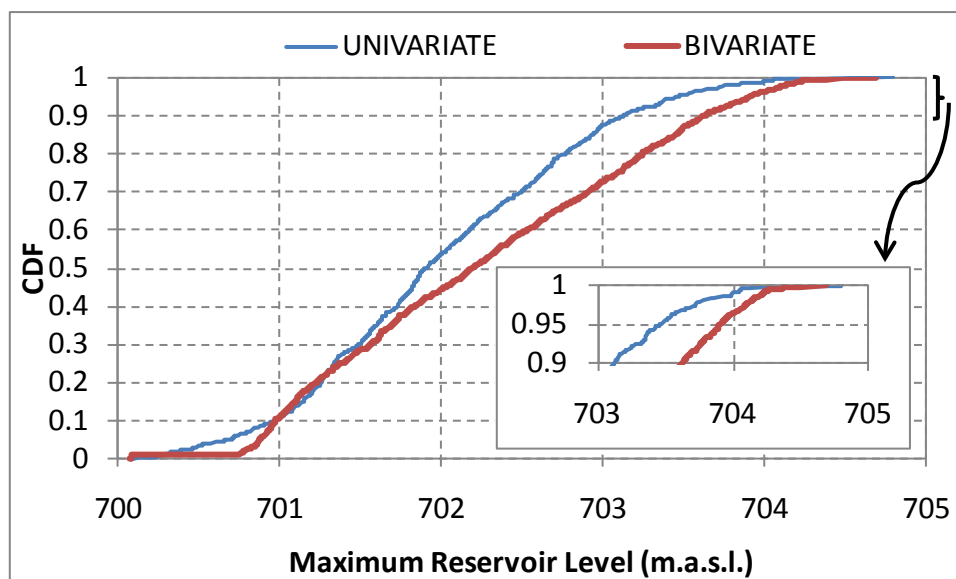


Figure 5.30: Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic winter floods.

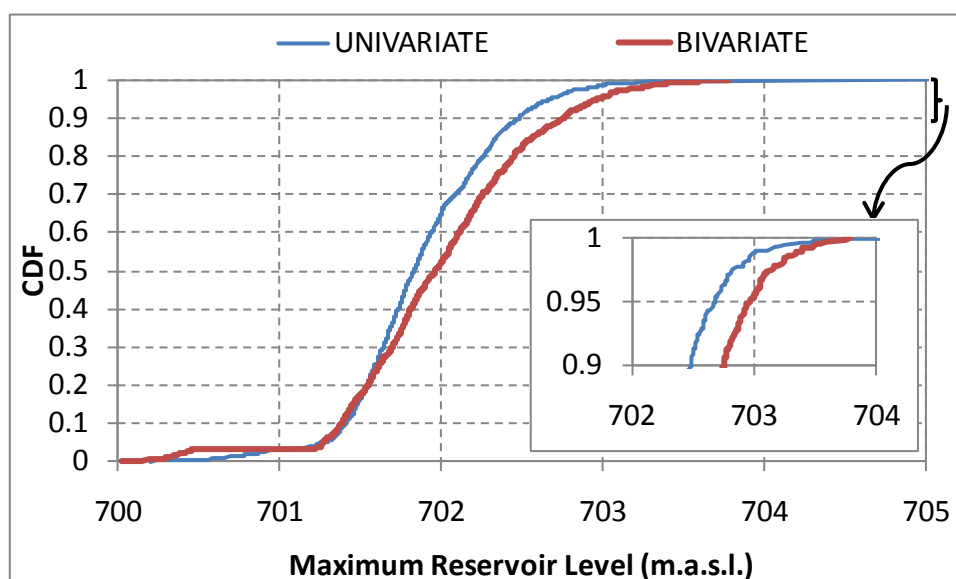


Figure 5.31: Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic spring floods.

The figures show a clear difference between the maximum levels reached with the synthetic floods derived from the univariate and bivariate models. The events generated with the bivariate approach result in levels of reservoir that are overall higher than the levels obtained using the univariate approach. If the level with a frequency of 95% is taken, the results show a difference of 45 cm. (703.45 against 703.90 m.a.s.l.) for the winter events, and 29 cm. (702.67 against 702.96 m.a.s.l.) for the spring events.

Similar results concerning maximum flows spilled through the spillway resulting from the routing of the synthetic hydrographs are presented in the following graphs (see Figures 5.32 and 5.33).

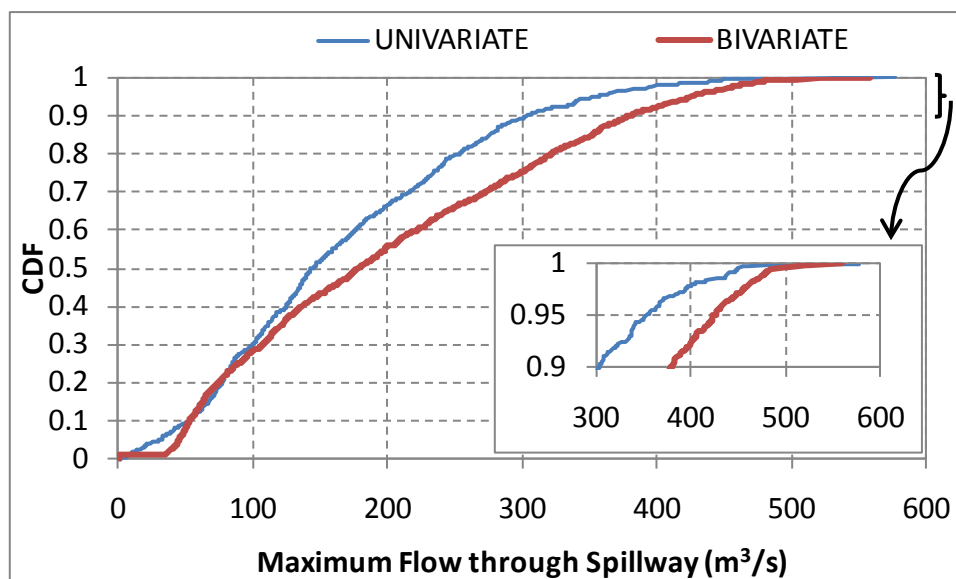


Figure 5.32: Empirical distributions of maximum spilled flows resulting from the routing of 1000 synthetic winter floods.

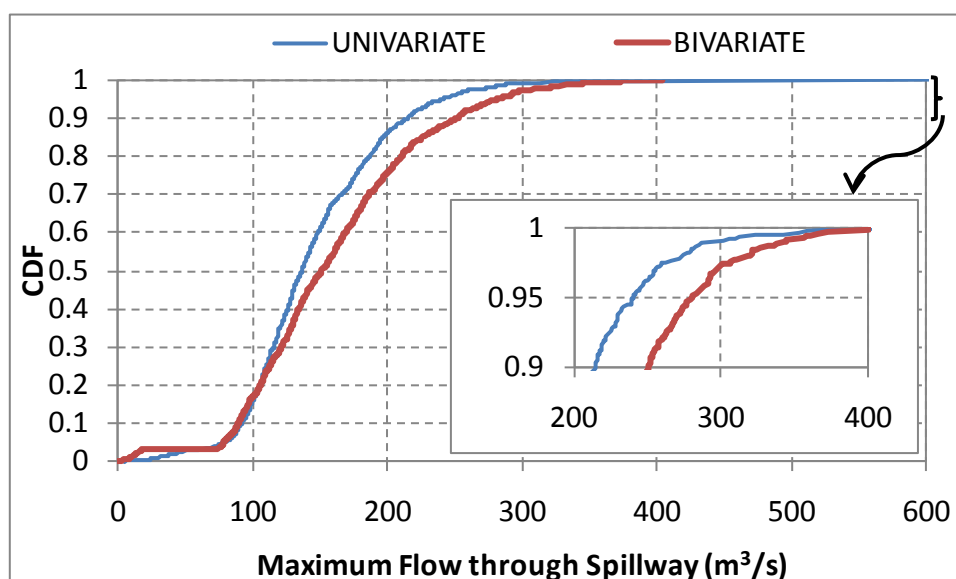


Figure 5.33: Empirical distributions of maximum spilled flows resulting from the routing of 1000 synthetic spring floods.

The figures show that the results obtained by both approaches are different, and the outflows are lower for the univariate approach compared with the bivariate. If the spilled flow with a frequency of 95% is taken, then the results show a difference of 70 m³/s (356 against 423 m³/s) for the winter events, and 40 m³/s (240 against 280 m³/s) for the spring events.

An additional case was considered to decide how sensitive the results are to the synthesis of the shape of the hydrographs. In order to analyze this, the random pairs of characteristics were combined with one shape of hydrograph. This shape was defined using the series of parameters a of the Beta distribution generated by the bivariate approach. The most frequent parameter a and its corresponding pair b are used to generate one random shape. This random shape is combined with the two series of flood characteristics generated by the univariate and bivariate models. The

synthetic hydrographs are then routed through the reservoir resulting in the water elevations presented in the following graphs (see Figure 5.34).

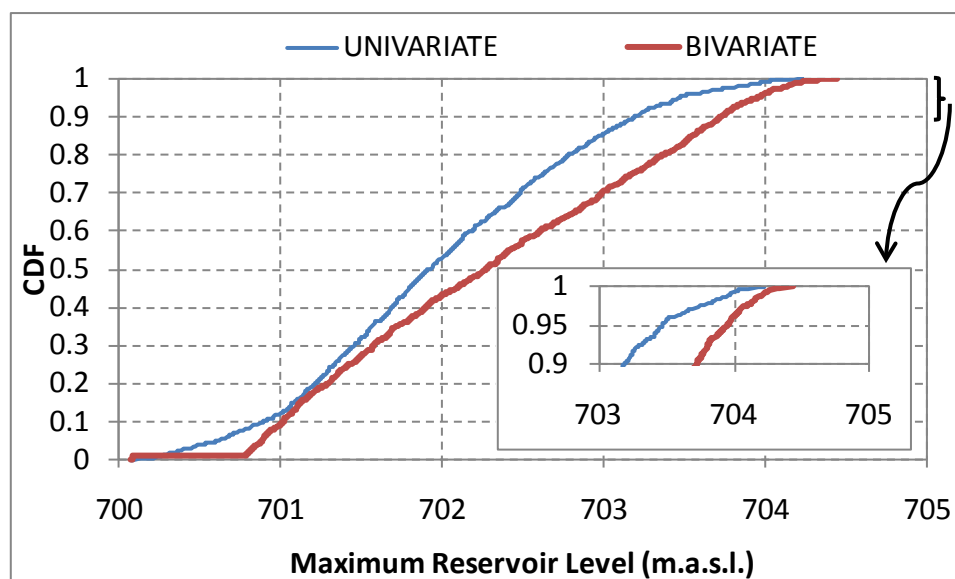


Figure 5.34: Analysis of the sensitivity of the shape parameters. Empirical distributions of maximum reservoir level resulting from the routing of 1000 synthetic winter floods, considering only one shape of hydrograph for all the cases.

The figures show that the results obtained by both approaches are different even if we consider the same shape for all the hydrographs. This means that the univariate or bivariate approach used for randomly generating flood characteristics give different results, independently of the criteria of defining the shape of the flood.

Note: In order to analyze the sensitivity of the reservoir characteristics in the results, the different reservoir routings were repeated changing the total surface of the reservoir and total length of the spillway. The behavior of the reservoir levels showed similar results as the ones presented in this section, i.e. the univariate analysis resulted in lower reservoir levels compared to the bivariate case. These results are not included in this report.

5.4.2. Dam design procedure

The design criteria proposed in this work involves different steps of calculation that are explained here. First is the definition of a design criteria in terms of a return period. In the second step the variables characterizing the flood associated to the design criteria have to be estimated. When 2 variables are considered, a set of pairs associated to the joint return period should be selected. In this work a set of 20 pairs is taken, shown in the following figure (see Figure 5.35). The next step involves the selection of a group of “critical” shapes of hydrographs which produce the maximum reservoir levels. A group of 50 “critical” shapes are used here. The estimated pairs of values are combined with the different “critical” shapes to generate 1000 synthetic hydrographs associated to the design return period. These hydrographs are routed through the reservoir, resulting in a series of water levels in the reservoir and outflows. The maximum values of water level and outflow are identified for each of the routed hydrographs.

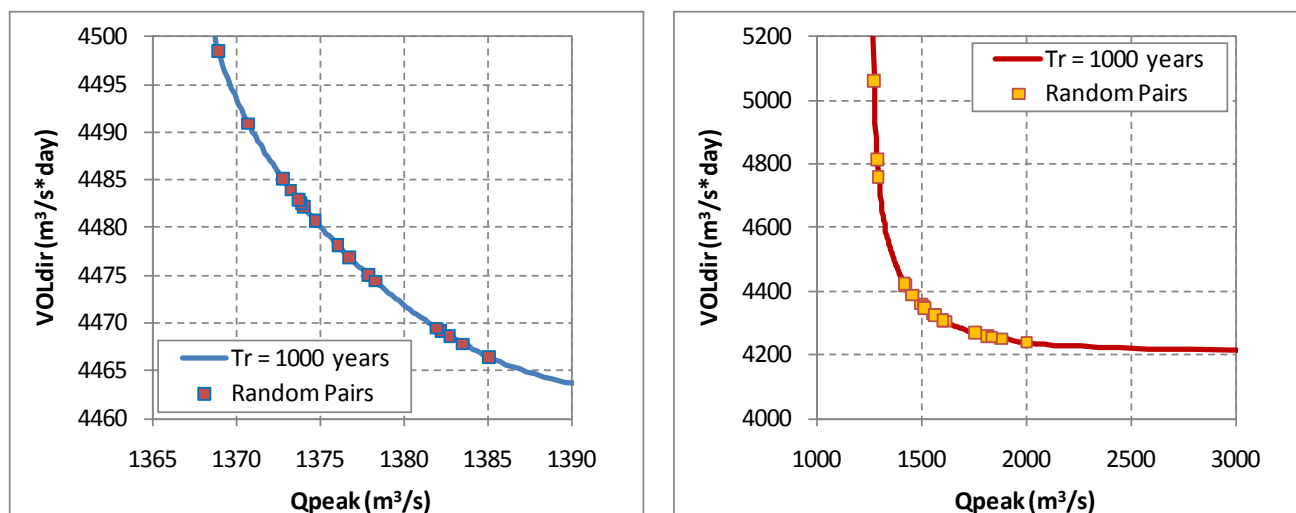


Figure 5.35: Random set of flood characteristics associated to a joint return period of 1000 years. Left: Winter flood events; Right: Spring flood events.

The univariate and bivariate approaches are compared by using the different pairs of values associated to the design return period, i.e. one pair of values for the univariate approach and the set of 20 pairs of values for the bivariate approach. There is a total number of 50 “critical” shapes, which are used both for the univariate and bivariate case. This results in a total number of 50 synthetic hydrographs for the univariate approach and 1000 for the bivariate, which are routed through the reservoir. The results are presented as empirical distributions of the maximum reservoir level and outflows reached during the synthetic flood events, in the following figures (see Figures 5.36 to 5.39). The graphs include a zoom to the region of 95% of frequency for a better visualization of the differences.

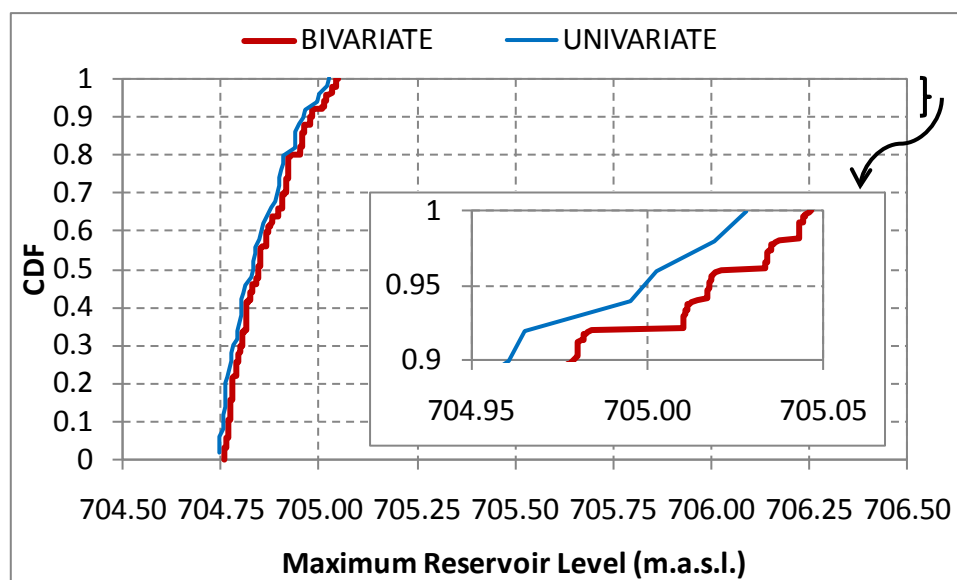


Figure 5.36: Empirical distributions of maximum reservoir level resulting from the routing of synthetic winter floods associated to a return period of 1000 years.

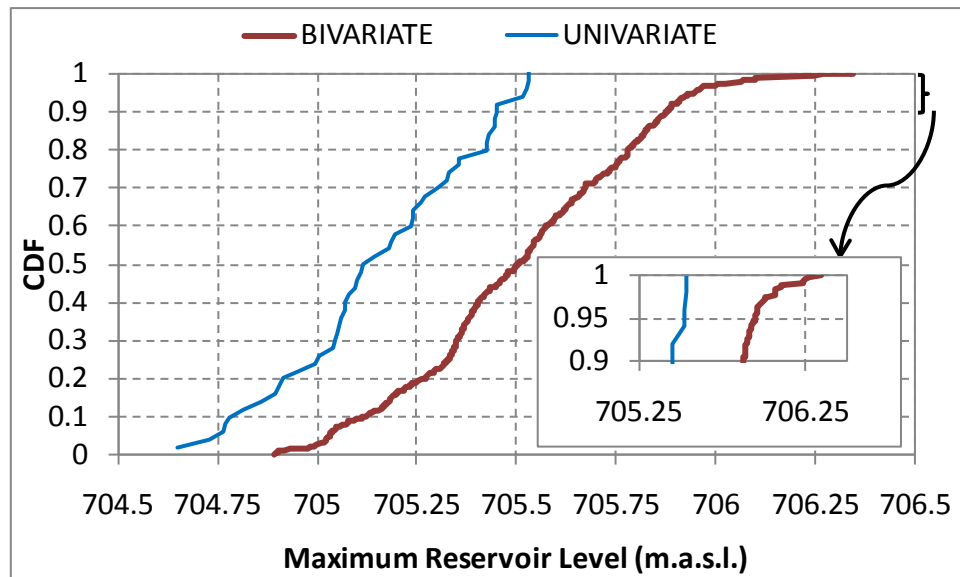


Figure 5.37: Empirical distributions of maximum reservoir level resulting from the routing of synthetic spring floods associated to a return period of 1000 years.

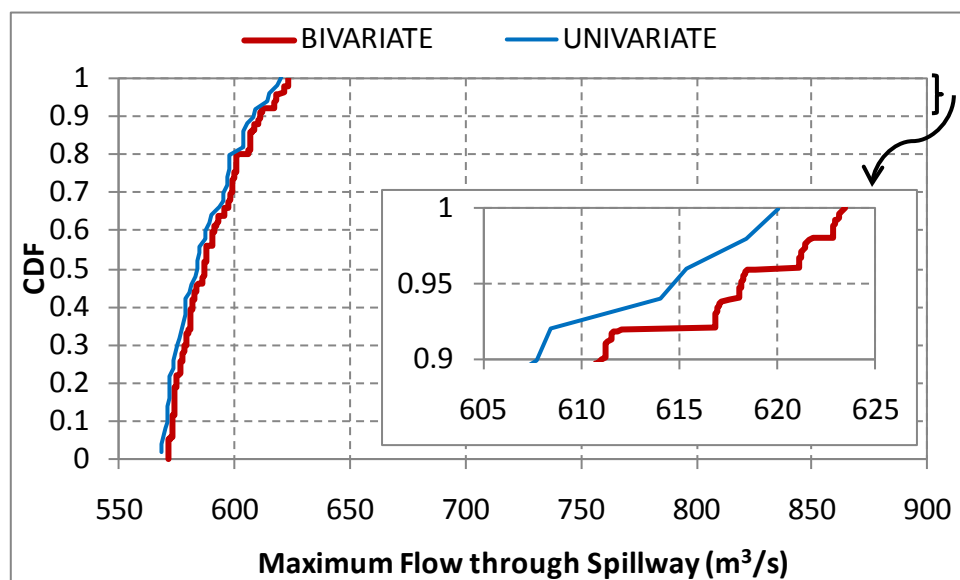


Figure 5.38: Empirical distributions of maximum spilled flows resulting from the routing of synthetic winter floods associated to a return period of 1000 years.

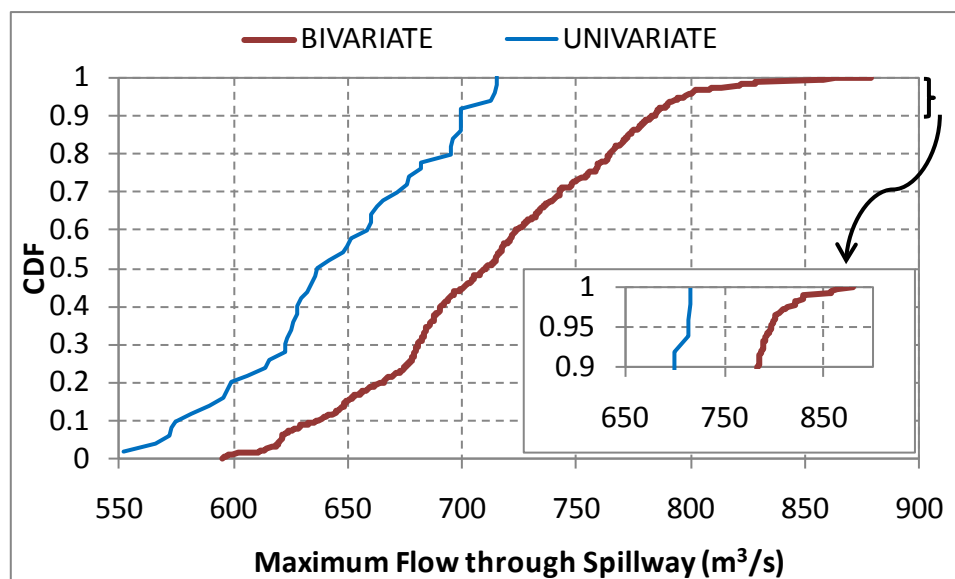


Figure 5.39: Empirical distributions of maximum spilled flows resulting from the routing of synthetic spring floods associated to a return period of 1000 years.

The figures are plotted in the same scale in order to compare the different behavior of the winter and spring events. In both cases the bivariate approach results in values of maximum reservoir levels and outflow that are higher compared with the univariate approach. The winter events do not show a relevant difference between the curves obtained using the two approaches, and this difference is much more important for the spring events. This difference of behavior between the univariate and bivariate approach in the design criteria is a result of the marginal distribution functions used to estimate the pairs of values associated to a return period.

It was shown that the correlation of the flood characteristics of winter events is higher than the one belonging to the spring events. This phenomenon has an effect on the copula model estimation. The estimation of the contour line representing the joint probability associated to a return period of 1000 years does not differ significantly when the two models are compared. The shape and position of both curves are almost the same, indicating similar combinations of marginal distributions. The difference in the results becomes visible when transforming the marginal distribution values into the variables under study, i.e. converting the values between 0 and 1 to peak flow and direct volume values.

If a value with 95% of frequency is chosen for designing a dam, the bivariate approach resulting from the spring events gives a maximum water elevation of 705.95 m.a.s.l. and a maximum outflow of 797.11 m³/s. These values are higher compared with the univariate approach, for which the resulting values are 705.52 m.a.s.l. and 713.5 m³/s respectively. If the univariate case is taken for defining the height of a dam and spillway capacity, both structures would be underdimensioned compared to the bivariate approach. The effect of including the information of correlation between the two flood characteristics in the design, results in higher design values associated to a given return period.

6. SUMMARY AND CONCLUSIONS

In this work a methodology of assessing the risk of an existing dam, as well as a procedure for a risk-oriented design of a dam, both based on a bivariate frequency analysis are presented. The bivariate model is a copula representing the joint behavior of two random variables describing a flood, peak flow and volume, which are of primary interest in hydrological practice.

A multivariate analysis allows the inclusion of several variables characterizing a flood, as well as taking into account the dependence structure linking them. This type of analysis uses more information given by the observed hydrographs, compared with the traditional hydrological frequency analysis, which focus on each of the variables separately in a univariate context. The advantage of using copulas to model the relationship between the variables is the possibility to apply multivariate distribution functions with different marginal functions. Different hydrological scenarios associated to a given joint return period can be derived easily from these models. The information derived from this analysis is more adequate for flood control structures than the traditional return period analysis.

Different steps of calculation are followed in the proposed methodology and procedure. The first step of calculation involves the identification of the peak flows for the different flooding seasons of the analyzed river, which are winter and spring. Then the flood event associated to each peak flow is defined, i.e. the volume and duration of the event. This step involves an independence analysis of peak flows and the base flow separation, resulting in the direct flow hydrograph belonging to the flood event. The resulting hydrographs are analyzed, and different series of variables are extracted describing the magnitude of the flood, characterized by peak flow and direct volume, and the shape of the flood, described by two shape parameters. The shape of the hydrograph is represented using two-parameter Beta distribution function. The next step involves univariate frequency analysis of the different series of variables, the traditional univariate models applied in hydrology are analyzed and compared. Then the pairs of values describing the flood magnitude and flood shape are analyzed and different copula models are compared to identify the ones that best represent the joint behavior of the variables. A copula model is defined for the flood magnitude pairs and another for the flood shape pairs. The events belonging to the different flooding seasons are analyzed separately, due to the fact that the climatic conditions generating them differ resulting in different flood magnitudes and shapes. As the river used for this study has two flooding season, two different groups of variables are analyzed separately.

The following step involves the generation of synthetic hydrographs. The direct flow hydrographs are synthesized by combining random pairs of flood characteristics, i.e. peak flow and volume, and random shapes of hydrographs. These random values are obtained using each of the copula models described in the previous paragraph. The base flow is added to each hydrograph to obtain the total synthetic hydrographs. Each synthetic hydrograph is used to simulate the routing of the flood through the reservoir resulting in a maximum water level and outflow reached during each event. The obtained values are presented as empirical distributions of maximum reservoir levels and maximum spilled flows resulting from the routing of the synthetic flood events through the reservoir.

Using the copula models to generate synthetic floods and simulating their pass through a reservoir gives the possibility to identify critical events for the hydraulic structures (dam or spillway) and to evaluate the effect of the flood control of the reservoir. These results can provide valuable

information to water resources managers and hydraulic engineers involved with this kind of structures.

In order to evaluate the results obtained using the copula models, a similar procedure is followed based on the traditional univariate frequency analysis. The results are compared with the univariate approach, so the main conclusions are derived from the comparison of both approaches. The main results obtained for the case study analyzed in this work are the following:

- The use of additional information regarding the shape of the hydrograph provides a better understanding and representation of the whole process generating the floods, compared to only considering either the peak flow or volume.
- Using the copula models to generate random pairs of values of peak flow and volume results in pairs that represent more adequately the observed pairs compared to the univariate analysis, due to the fact that the copula models include the information of dependence structure linking the two observed sets of variables. The copula models provide pairs of values that are more logical from the physical point of view, because they are in the region of the observed floods, whereas the univariate approach could result in a pair of values with a very high peak flow and a very low volume, which would differ from the natural behavior of the river.
- The copula models provide a set of pairs of values associated to a joint return period, which are either equal or higher than the values estimated using the univariate frequency analysis. This allows the incorporation of uncertainty in the design of hydraulic structures, rather than designing them with a one-design-event method.
- The synthetic flood events generated from the copula models result in higher values of reservoir levels and outflows, compared to the ones generated from the univariate analysis.
- If the univariate analysis is used for the design of hydraulic structures, the resulting values would be underestimated compared to a bivariate approach.
- The bivariate approach results in values of reservoir level that are more critical in terms of a risk assessment compared to the results obtained using the univariate approach.

Summarizing, the above statements indicate that the estimated floods based on the bivariate frequency analysis applied to the case study results in higher values of maximum water elevations and outflows compared with the univariate approach. If the risk of an existing dam is to be assessed, the bivariate approach would result in a greater risk of overtopping the dam for a given dam height and spillway. If a dam is to be designed, the consideration of a joint return period would result in estimated design parameters, i.e. maximum water height and outflow, which are higher compared to the values estimated using the univariate models.

As mentioned, the river used in this study has two flooding seasons. The two series of flood events are analyzed separately giving the possibility to compare the effect of some of the characteristics differing from one series to the other. The effect of the strength of the dependence between the variables is visible in the generation of random pairs of values. The pairs of variables characterizing the winter events, i.e. peak flow and volume, have a stronger dependence compared to the spring events. This results in random pairs of values that are closer to the observed pairs, and closer to the main diagonal when one variable is plotted against the second one. Whereas for the spring events, the random pairs of values are more dispersed and could result in floods that are not critical in terms

of maximum reservoir levels. The winter flood events, which have the strongest dependence between the variables, result in higher reservoir levels. The effect of the marginal univariate distributions is visible in the estimation of the pairs of values associated to a joint return period. The two copula models representing the winter and spring events show similar combinations of marginal distribution values associated to a joint return period. When these pairs of values are transformed to magnitudes of peak flow and volume, using the univariate models, the difference between the two models becomes visible. The variables describing the winter events are represented by two Generalized Pareto distributions, which result in estimations of random variables for the selected exceedance probability that do not differ significantly from the maximum observations. Whereas for the spring events, the univariate models representing the variables are the Weibull and Generalized Extreme Value distributions, resulting in estimations of variables that are much higher than the maximum observations. These effects are seen in the empirical distributions of maximum reservoir levels and maximum spilled flows, which show higher values for the spring events compared to the winter events. In the case of the winter floods, these values are lower, and although the curve of the bivariate analysis shows higher values than the univariate approach, they do not differ significantly.

The final conclusion of this work is that incorporating multiple variables to describe flood events, results in values that are more critical both for the risk assessment and for the design of hydraulic structures, such as dams.

7. FUTURE WORKS

The conclusions presented in this work are the result of characterizing the flood events using the peak flow and direct volume. Future research could involve studying the sensitivity of the results by characterizing the flood with other variables, like total volume, total duration of the event or time to peak. A similar approach as the one presented in this work could be followed by deriving the bivariate models using other combination of flood variables.

The copula models applied in this work include only some of the existing ones. Other copula models could be used, for example two parameter copulas and the results could be compared with the ones presented here. The method of estimation of the parameters of the copulas could also be compared with other methods proposed in the cited works. Another study could involve the analysis of the effect of the marginal distributions on the obtained results.

A new approach considering a multivariate copula model could be carried including three variables to describe the flood events. The flood could be characterized by: peak flow, volume and duration. This approach would include more information representing the flood event and the results could be important, especially regarding the variable duration, which is difficult to estimate.

Further effort should be given to the estimation of the time-characteristic variables of floods (duration and time to peak) in relationship with other flood characteristics, in order to have a better understanding of the flood process as a whole, and to reduce the uncertainties associated to the synthetic hydrographs.

8. ACKNOWLEDGMENTS

The access to the data used for this work was granted by the Subsecretaría de Recursos Hídricos de la República Argentina. They are gratefully acknowledged.

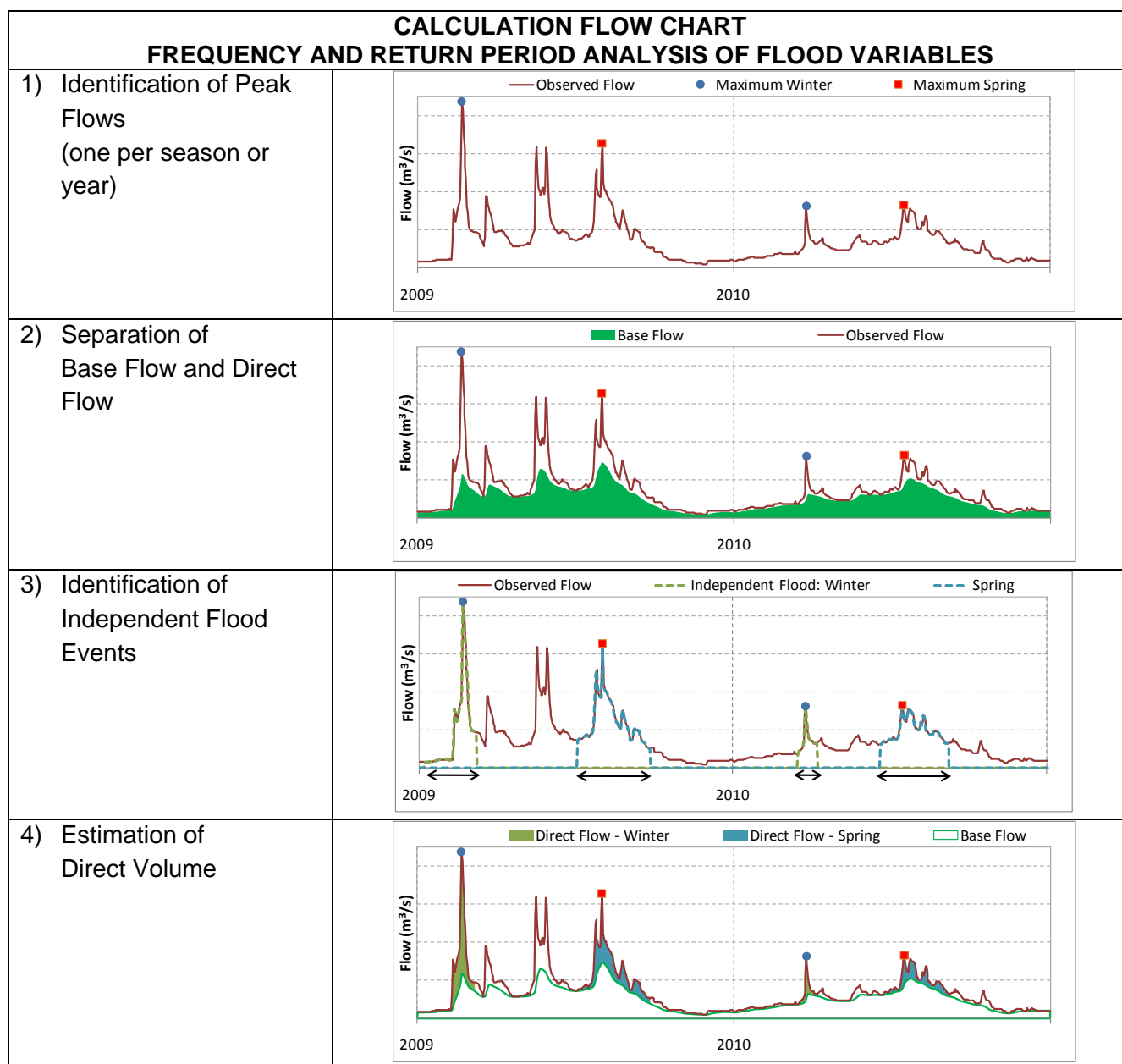
REFERENCES

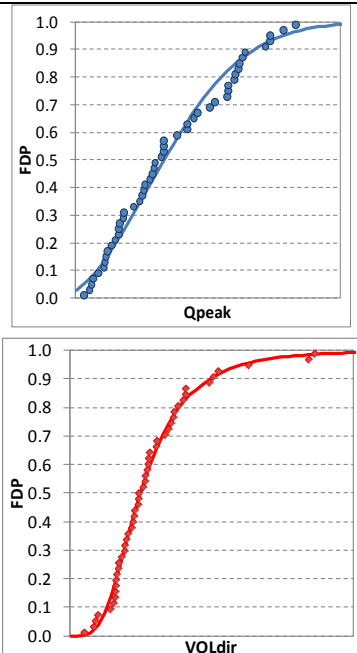
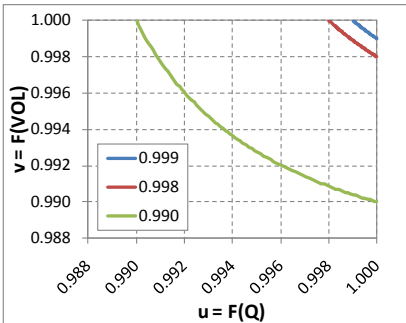
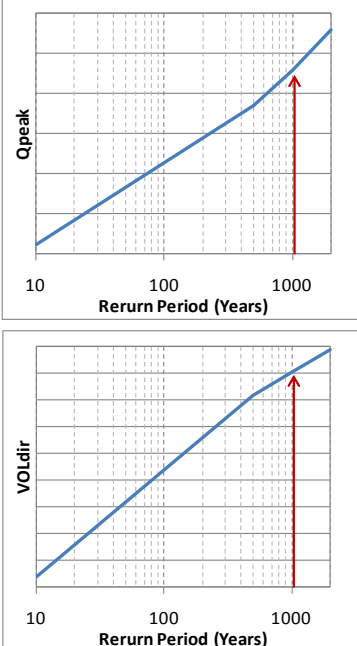
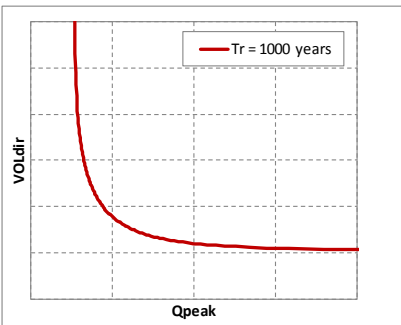
- Aldama, A., and A. Ramírez (1999). A new approach for dam design flood estimation. In *Proceeding of the IAHR Congress, Graz, Austria, 22-27 August, 1999*.
- Arnold, J. G., P. M. Allen, R. Muttiah, and G. Bernhardt (1995). Automated Baseflow Separation and Recession Analysis Techniques. *Groundwater* 33 pp.1010-1018.
- Arnold, J.G., and P.M. Allen (1999). Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association* 35(2) pp. 411-424.
- Asquith, William H. (2011). *Distributional Analysis With L-Moment Statistics Using the R Environment for Statistical Computing*. ISBN-13: 978-1463508418. 2011.
- Bacchi, B., A. Brath, and N. T. Kottegoda (1992). Analysis of Relationships Between Flood Peaks and Flood Volumes Based on Crossing Properties of River Flo Processes. *Water Resources Research* 28(10) pp. 2773-2782.
- Bell, F., and S. Om Kar (1969). Characteristic responses times in design flood estimation. *Journal of Hydrology* 8 pp. 173-196 .
- Bender, J., Mundersbach C., and J. Jensen (2011). Synthetic Hydrograph Generator for Reservoirs: Options and Limitations. *European Geosciences Union – General Assembly 2011*.
- Benedetti, J. (2000). Los Sistemas de Información Geográfica en el Instituto Geográfico Militar. X Congreso Nacional de Cartografía, Contribuciones Científicas, pp. 314-322. Instituto Geográfico Militar. Buenos Aires.
- Bergmann, H., O. Hable, and R. Krainer (2000). Statistical Approach of Overtopping Probability for Dams with Consideration of the Reservoir Management. 26th National and 3rd International Hydrology and Water Resources Symposium of The Institution of Engineers, Australia.
- Bhunya, P.K., R. Berndtsson, C.S.P. Ojha, and S.K. Mishra (2007). Suitability of Gamma, Chi-square, Weibull, and Beta distributions as synthetic unit hydrographs. *Journal of Hydrology* 334 (1-2) pp. 28-38.
- Bobée, B., and F. Ashklar (1991). The gamma family and derived distributions applied in hydrology. *Water Resources Publications* (Littleton, Colo., U.S.A.). ISBN 0918334683.
- Bureau of Reclamation (1987). Design of small dams. A Water Resources Technical Publication, United States Department of Interior. U.S. Government Printing Office. Superintendent of Documents, Mail Stop: SSOP. Washington, DC 20402-9328.
- Chow, V. T., D. R. Maidment, and L. W. Mays (1988). *Applied hydrology*. McGraw-Hill, New York. ISBN 0-07-010810-2.
- Chowdhary, H., L. A. Escobar, and V. P. Singh (2011). Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrology Research Vol. 42 No. 2–3* pp. 193–216.
- Cunderlik, J.M., and D.H. Burn (2004). Linkages between regional trends in monthly maximum flows and selected climatic variables. *Journal of Hydrologic Engineering ASCE* 9(4) pp. 246-256.
- Cunnane, C. (1979). A note on the Poisson assumption in partial duration series models. *Water Resources Research* 15 (2) pp. 489–494.
- D’Agostino, R.B., and M. A. Stephens (1986). *Goodness-of-fit techniques*. Marcel Dekker, Inc. ISBN 0-8247-7487-6.
- De Michele, C., G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso (2005). Bivariate Statistical Approach to Check Adequacy of Dam Spillway. *Journal of Hydrologic Engineering* 10(1) pp. 50-57.
- Delignette-Muller, M. L., R. Pouillot, J.B. Denis, and C. Dutang (2010). *fitdistrplus: help to fit of a parametric distribution to non-censored or censored data*.
- Descroix, L. M., K. Esteves, J.L. Souley Yéro, M. Rajot, S. Malam Abdou, J. M. Boubkraoui, N. Lapetite, I. Dessay, O. Zin, A. Amogu, I. Bachir, E. Bouzou Moussa, E. Le Breton, and I. Mamadou (2011). Runoff evolution according to land use change in a small Sahelian catchment. *Hydrology and Earth System Sciences Discussions* 8 pp. 1569-1607.
- Durante, F., and G. Salvadori (2009). On the construction of multivariate extreme value models via copulas. *Environmetrics* (2009).
- England, J.F. Jr. (2003). Simplified probabilistic extreme flood hydrographs for dam safety. Bureau of Reclamation, Denver, CO pp 1-11.
- Fealy, R., and J. Sweeney (2005). Detection of a possible change point in atmospheric variability in the North Atlantic and its effect on Scandinavian glacier mass balance. *International Journal of Climatology* 25 pp. 1819-1833.
- Favre, A.C., S. El Adlouni, L. Perreault, N. Thiémonge, and B. Bobée (2004). Multivariate hydrological frequency analysis using copulas, *Water Resources Research* 40 (W01101) pp. 1-12.
- Gaál, L., J. Szolgay, S. Kohnová, J. Parajka, R. Merz, A. Viglione and G. Blöschl (2012). Flood timescales: Understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research* 48 (W04511) pp. 1-21.
- Genest, C., and A. Favre (2007a). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrological Engineering* 12 pp. 347–368.
- Genest, C., A. Favre, J. Béliveau, and C. Jacq ues (2007b). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research* 43 (W09401) pp. 1-12.
- Genest, C., B. Remillard, and D. Beaudoin (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics* 44 pp. 199–214.
- Gobierno de la Provincia de Neuquén (2008). Aprovechamiento Multipropósito Chihuido I, Convocatoria pública a inversores.
- Goodarzi, E., M. Mirzaei, L. T. Shui, and M. Ziaei (2011). Evaluation dam overtopping risk based on univariate and bivariate flood frequency analysis. *Hydrology and Earth System Sciences Discussion* 8 pp. 9757–9796.

- Grimaldi, S., and F. Serinaldi (2006). Asymmetric copula in multivariate flood frequency analysis. *Advances in Water Resources* 29 (8) pp. 1155–1167.
- Hamed, K.H. (2009). Enhancing the effectiveness of prewhitening in trend analysis of hydrologic data. *Journal of Hydrology* 368 pp. 143–155.
- Hirsch, R.M., J.R. Slack, and R.A. Smith (1982). Techniques of trend analysis for monthly water quality data. *Water Resources Research* 18(1) pp. 107-121.
- Hofert, M., and M. Maechler (2011). Nested Archimedean Copulas Meet R: The nacopula Package. *Journal of Statistical Software*, 39(9), 1-20. URL <http://www.jstatsoft.org/v39/i09/>.
- Klein, B., M. Pahlow, Y. Hundecha, C. Gattke, and A. Schumann (2008). Probabilistic Analysis of Hydrological Loads to Optimize the Design of Flood Control Systems. 4th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability Toronto, Ontario, Canada, May 6-8, 2008.
- Klein, B. (2009). Ermittlung von Ganglinien für die risikoorientierte Hochwasserbemessung von Talsperren. PhD Thesis, Ruhr-Universität Bochum, Schriftenreihe Hydrologie/ Wasserwirtschaft, Nr. 25.
- Klein, B., M. Pahlow, Y. Hundecha, and A. Schumann (2010). Probability Analysis of Hydrological Loads for the Design of Flood Control Systems Using Copulas. *Journal of Hydrologic Engineering* 15 (5) pp. 360–369.
- Klein, B., A. Schumann, and M. Pahlow (2011). Flood Risk Assessment and Management. Chapter 8: Copulas – New Risk Assessment Methodology for Dam Safety. Springer. ISBN: 9789048199167.
- Kojadinovic, I. (2009). Introduction to the theory and practice of copulas – New York, July, 2009. Department of Statistics, The University of Auckland.
- Kojadinovic, I., and J. Yan (2010). Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *Journal of Statistical Software*, 34(9), 1-20. URL <http://www.jstatsoft.org/v34/i09/>.
- Lang, M., T.B.M.J. Quarda, and B. Bobée (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225 pp. 103–117.
- LAWA (1997): Pegelvorschrift – Stammtext. Länderarbeitsgemeinschaft Wasser (LAWA). Bonn und Berlin: 1997.
- Lettenmaier, D.P., E.F. Wood, and J.R. Wallis (1994). Hydro-climatological trends in the continental United States 1948-1988. *Journal of Climate* 7(4) pp. 586-607.
- Maidment, D.R. (1993). *Handbook of Hydrology*. McGraw-Hill, Inc. ISBN 007039732-5.
- McCuen, R. H. (1998). *Hydrologic Analysis and Design*. Prentice Hall. ISBN 0131349589.
- Mediero, L., A. Jiménez-Álvarez, and L. Garrote (2010). Design flood hydrographs from the relationship between flood peak and volume. *Hydrology and Earth System Sciences* 14 pp. 2495–2505.
- Michiels, F., and A. De Schepper (2008). A Copula Test Space Model: How To Avoid the Wrong Copula Choice. Research Paper RPS–2007–027. Department of Mathematics, Statistics and Actuarial Sciences, University of Antwerp, Belgium.
- Moyano, C., and R. Diaz (2006). Mapa argentino de zonas áridas, semiáridas y subhúmedas. Proyecto del CONAPHI.
- Nagy, I.M., K. Asante-Duah, and I. Zsuffa (2002). *Hydrological Dimensioning and Operation of Reservoirs – Practical Design Concepts and Principles*. Kluwer Academic Publishers. ISBN 1402004389.
- Nathan, R. J., and T. A. McMahon (1990). Evaluation of Automated Techniques for Baseflow and Recession Analysis. *Water Resources Research* 26(7):1465-1473.
- Natural Environment Research Council (1975). *Flood Studies Report, Volume IV, Hydrological Data*. NERC, London.
- Nijssen, D., A. Schumann, M. Pahlow, and B. Klein (2009). Planning of technical flood retention measures in large river basins under consideration of imprecise probabilities of multivariate hydrological loads. *Natural Hazards and Earth Systems Sciences* 9 pp. 1349–1363.
- Novak, P., A.I.B. Moffat, C. Nalluri, and R. Narayanan (2007). *Hydraulic Structures*. Taylor & Francis, Inc. ISBN 978-0-415-38626-5.
- Pettitt, A.N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics* 28 (2) pp. 126-135.
- R Development Core Team (2012). *R: A Programming Environment for Data Analysis and Graphics*. R version 2.15.0 (2012-03-30).
- Salvadori, G. (2004). Bivariate return periods via 2-Copulas. *Statistical Methodology* 1 pp. 129–144.
- Salvadori, G., and C. De Michele (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research* 40 (W12511) pp.1-17.
- Salvadori, G., and C. De Michele (2007a). On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering* Vol. 12 (4) pp.369-380.
- Salvadori, G., C. De Michele, N.T. Kottegoda, and R. Rosso (2007b). *Extremes in Nature: An Approach Using Copulas*. Springer. ISBN 9781402044144.
- Salvadori, G., C. De Michele, and F. Durante (2011). On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences* 15 pp. 3293-3305.
- Seoane, R., and A. Callau (2009). Statistical analysis and tendency of maximum flows in argentinian rivers. *Estrategias integradas de mitigación y adaptación a cambios globales* pp.134-141. ISBN: 9789879641392.
- Serinaldi, F., and S. Grimaldi (2011). Synthetic design hydrographs based on distribution functions with finite support. *Journal of Hydrologic Engineering-ASCE* 16 pp. 434–445.
- Shiau, J. T. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment* 17 (2003) 42–57.
- Shiau, J. T., H. Y. Wang, and C. T. Tsai (2006). Bivariate frequency analysis of floods using copulas. *Journal of the American Water Resources Association* 42 (6) pp. 1549-1564.
- Singh, V. P. (1996). *Hydrology of Disasters*. Kluwer Academic Publishers. ISBN: 0792340922.

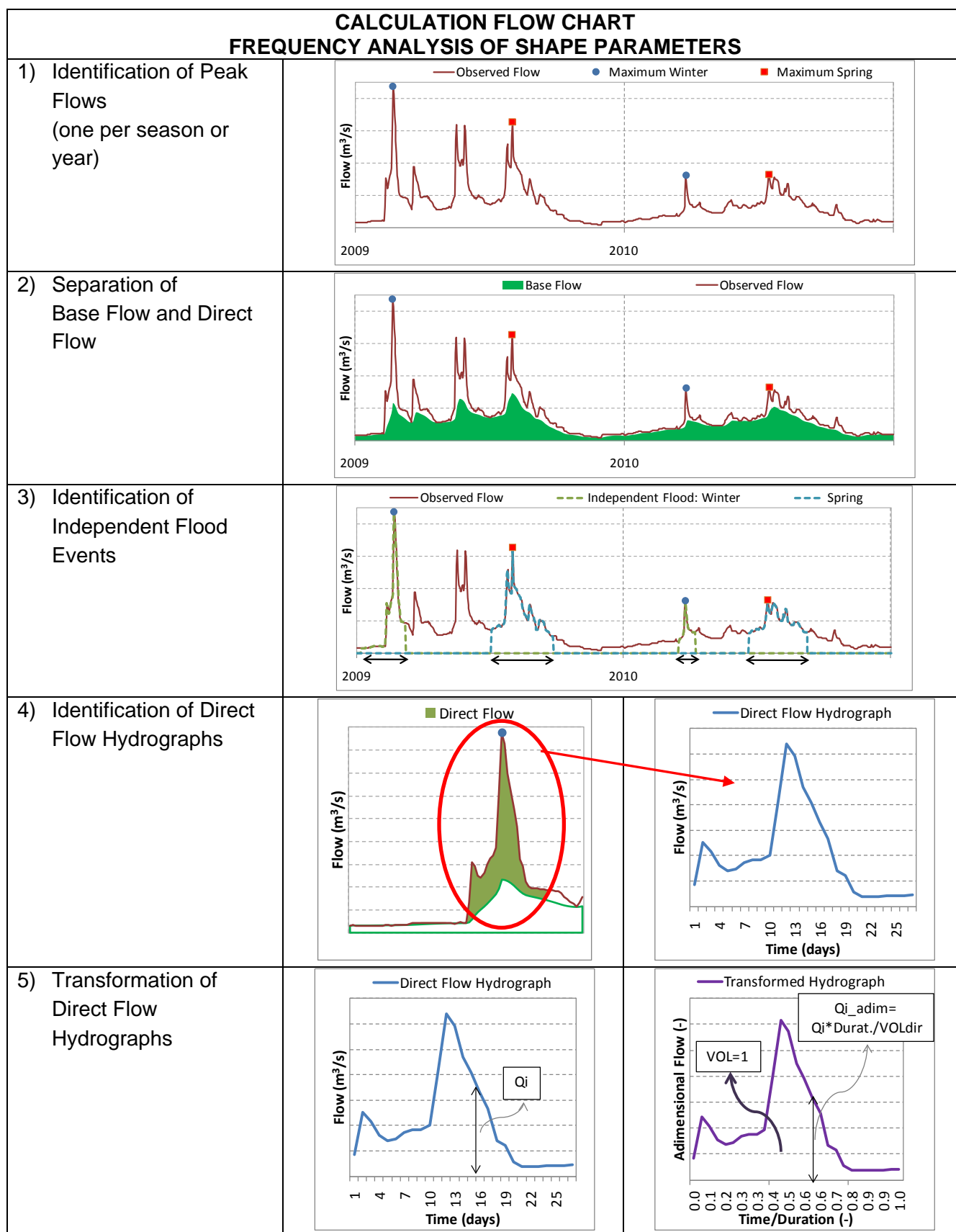
- SSRH (2002). Atlas Digital de los Recursos Hídricos de la República Argentina. CD-Rom. Subsecretaría de los Recursos Hídricos de la Nación. Buenos Aires.
- UNEP (1997). United Nations Environment Programme. World Atlas of desertification. Edited by Middleton N. and Thomas D.
- UNESCO (2003). Water for People, Water for Life, The United Nations World Water Development Report.
- US Water Resource Council (1976). Guidelines For Determining Flood Flow Frequency Bulletin 17 of the Hydrology Subcommittee of the U.S. Water Resources Council, Washington, DC.
- Valicenti, J. L. (2001). Cuenca del Río Neuquén – Análisis del Fenómeno Precipitación-Escorrentía. Autoridad Interjurisdiccional de las Cuencas de los Ríos Limay, Neuquén y Negro.
- Vandenberghe, S., M.J. van den Berg, B. Gräler, A. Petroselli, S. Grimaldi, B. De Baets, and N.E. C. Verhoest (2012). Joint return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrology and Earth System Sciences Discussions* 9 (5). pp. 6781-6828.
- Viglione, A., G.B. Chirico, R. Woods, and G. Blöschl (2010). Generalised synthesis of space-time variability in flood response: An analytical framework. *Journal of Hydrology*. 394 pp.198–212.
- Wang, C., N-B. Chang, and G-T. Yeh (2009). Copula-based flood frequency (COFF) analysis at the confluences of river systems. *Hydrological Processes* 23 pp. 1471–1486.
- Westmacott, J.R., and D.H. Burn (1997). Climate changes effects on the hydrologic regime within the Churchill-Nelson River Basin. *Journal of Hydrology* 202 pp. 263-279.
- Willems, P. (2009). A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental Modelling & Software* 24 pp. 311–321.
- Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software*, 21(4), 1-21. URL <http://www.jstatsoft.org/v21/i04/>.
- Yue, S. (2000). The bivariate lognormal distribution to model a multivariate flood episode, *Hydrological Processes* 14(14) pp. 2575–2588.
- Yue, S., T.B.M.J. Ouarda, B. Bobée, P. Legendre, and P. Bruneau (2002). Approach for Describing Statistical Properties of Flood Hydrograph. *Journal of Hydrologic Engineering* Vol. 7 (2) pp. 147-153.
- Yue, S., P. Pilon, and B. Phinney (2003). Canadian streamflow trend detection: Impacts of serial and cross-correlation. *Hydrological Sciences Journal* 48(1) pp. 51–63.
- Zhang, L., and V. P. Singh (2006). Bivariate Flood Frequency Analysis Using the Copula Method. *Journal of Hydrologic Engineering* 11 (2) pp. 150-164.

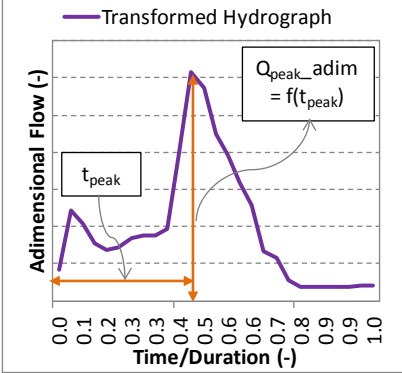
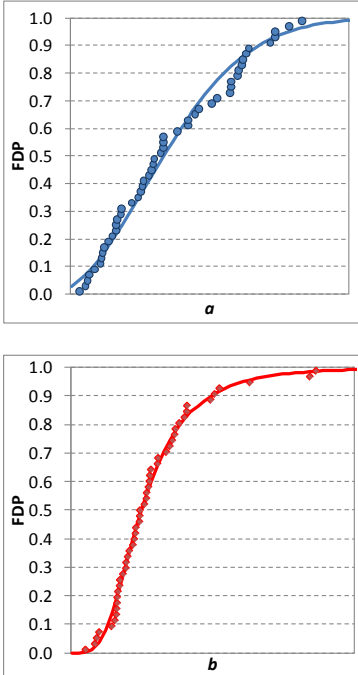
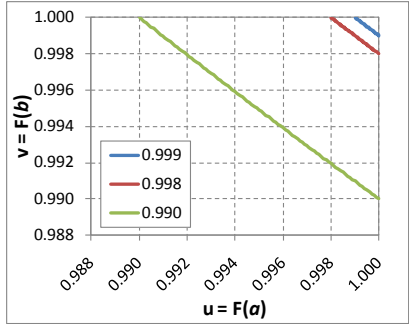
APPENDIX A



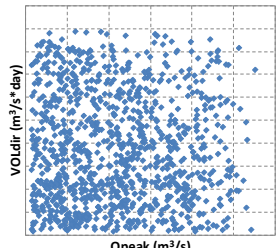
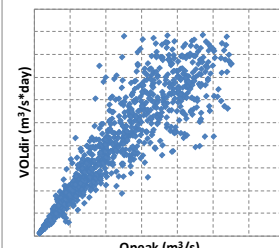
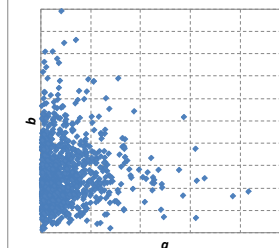
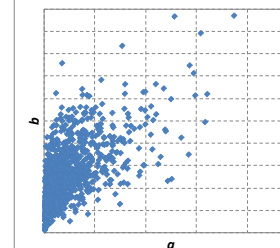
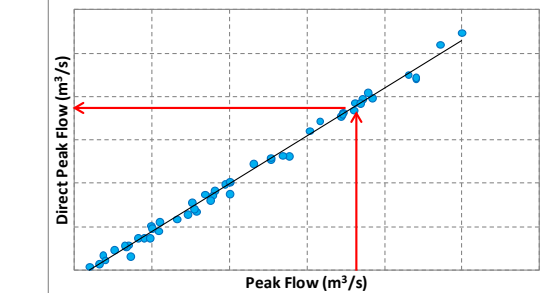
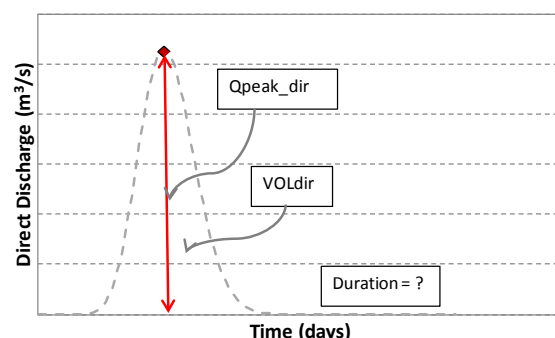
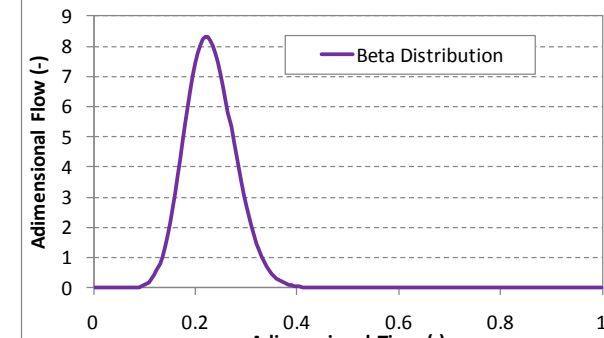
<p>5) Frequency Analysis of the series: Peak Flow and Direct Volume</p>	<p style="text-align: center;">UNIVARIATE</p> 	<p style="text-align: center;">BIVARIATE</p> 
<p>6) Return Period Analysis of the series: Peak Flow and Direct Volume</p>		

APPENDIX B

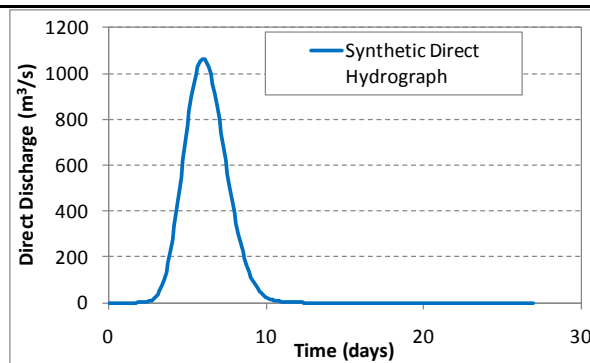


<p>6) Estimation of adimensional: time to peak: t_{peak} and peak flow: $f(t_{peak})$, for each of the hydrographs.</p>		
<p>7) Estimation of Beta distribution parameters: a and b, for each of the hydrographs.</p>	$t_{peak} = \frac{a - 1}{a + b - 2}$ $f(t_{peak}) = \frac{(a - 1)^{(a-1)} * (b - 1)^{(b-1)}}{(a + b - 2)^{(a+b-2)} * B(a, b)}$	
<p>8) Frequency Analysis of the series: a and b</p>	<p style="text-align: center;">UNIVARIATE</p> 	<p style="text-align: center;">BIVARIATE</p> 

APPENDIX C

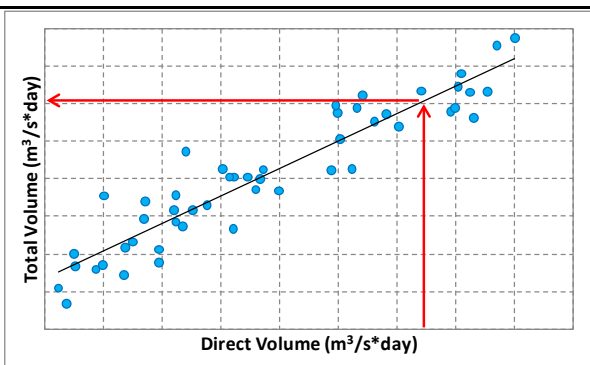
CALCULATION FLOW CHART RISK ASSESSMENT FOR DAM SAFETY			
1A) Flood Magnitude: 1000 random pairs of Qpeak-VOLdir		1B) Flood Shape: 1000 random pairs of a-b	
UNIVARIATE	BIVARIATE	UNIVARIATE	BIVARIATE
			
2A) Estimation of Direct and Base Peak Flows, using the Qpeak (Total Peak Flow)		2B) Estimation of the adimensional time to peak and peak flow, using a-b	
		$t_{peak} = \frac{a - 1}{a + b - 2}$ $f(t_{peak}) = \frac{(a - 1)^{(a-1)} * (b - 1)^{(b-1)}}{(a + b - 2)^{(a+b-2)} * B(a, b)}$	
3A) Magnitude of Direct Flow Hydrograph		3B) Shape of Direct Flow Hydrograph	
			
4) Estimation of the duration: Combination between flood magnitude and shape			
$Q_{peak_dir} = f(t_{peak}) \frac{VOLdir}{Duration}$		$Duration = f(t_{peak}) \frac{VOLdir}{Q_{peak_dir}}$	

5) Estimation of Direct Flow Hydrograph

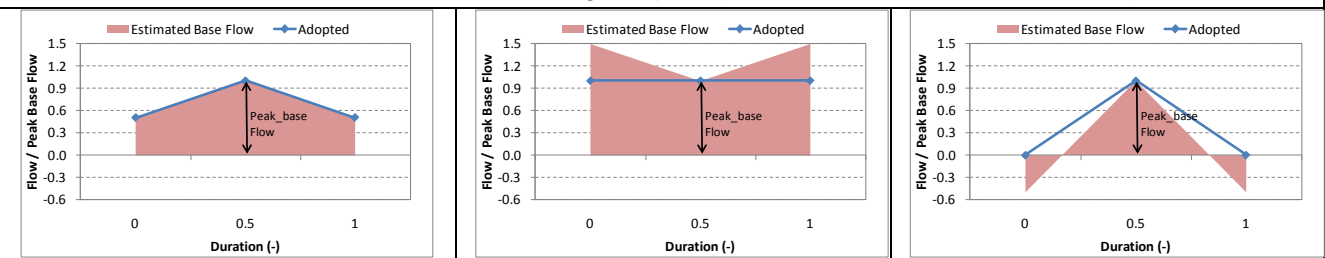


6) Estimation of Volume corresponding to the Base Flow:

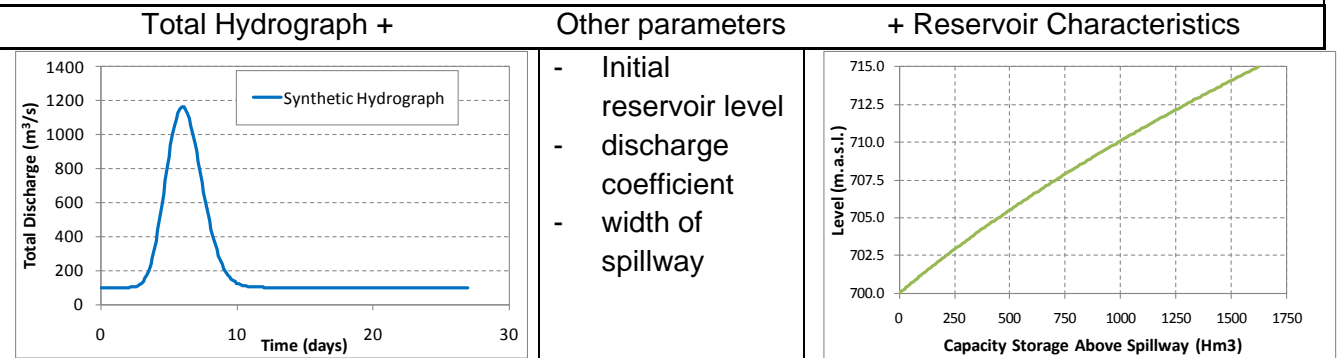
$$\text{Total Volume} = \text{Base Volume} + \text{Direct Volume}$$



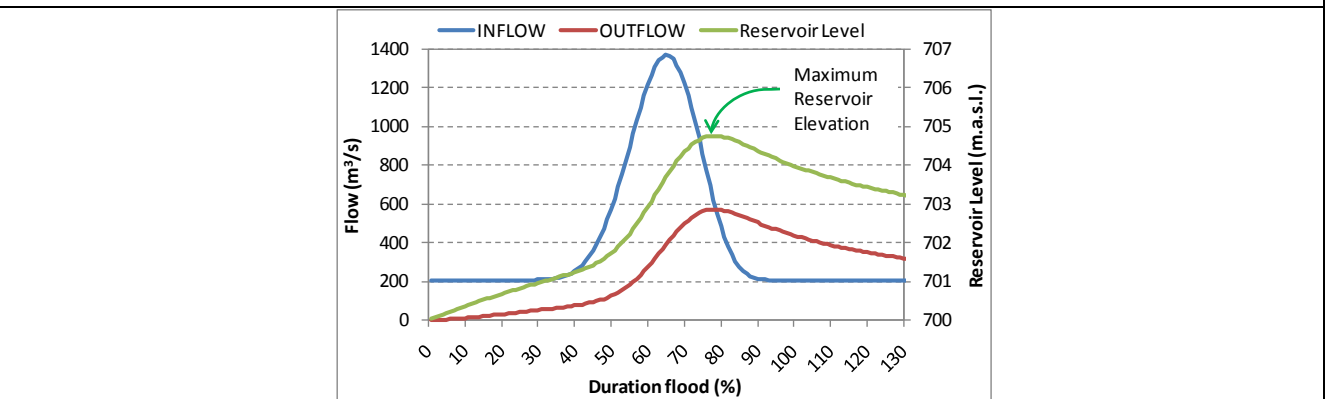
7) Adding up the Base flow, using the Base Peak Flow, Base Volume and Duration of the event. The base flow will take one of the following shapes. (for further details see Table 3.2)



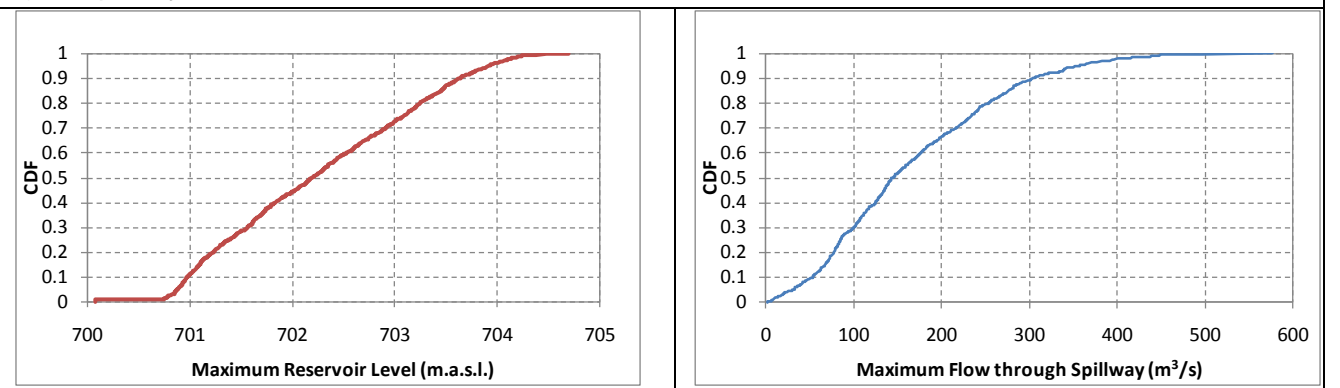
8) Routing of the total hydrograph through the reservoir



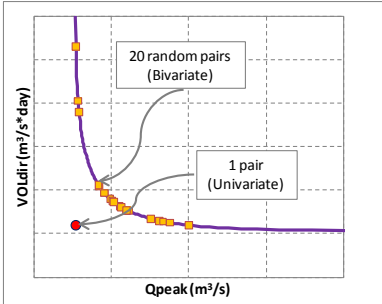
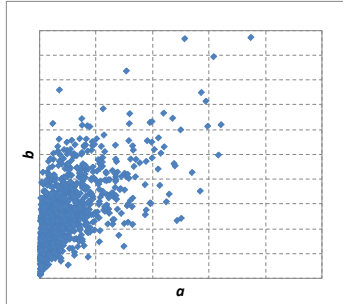
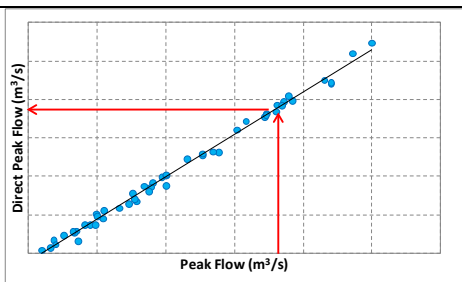
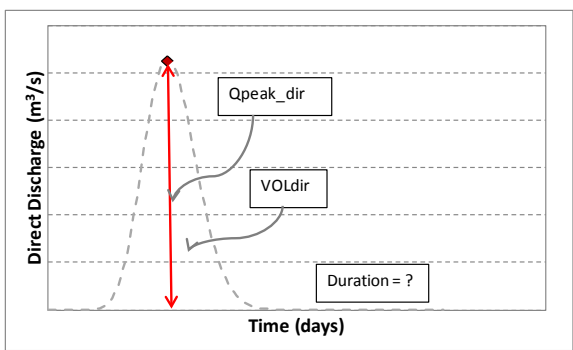
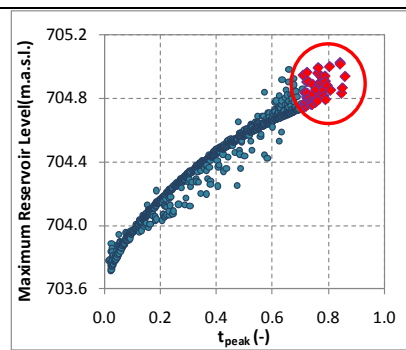
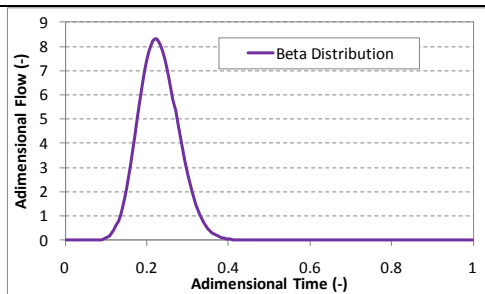
Resulting Curves



9) Frequency of Maximum Reservoir levels and outflows



APPENDIX D

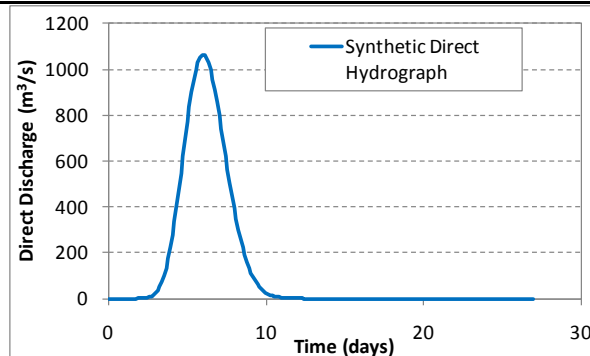
CALCULATION FLOW CHART DAM DESIGN PROCEDURE		
1A) Flood Magnitude: Pairs of Qpeak-VOLdir associated to a return period: $Tr=1000$ years		1B) Flood Shape: 1000 random pairs of a-b
UNIVARIATE: 1 pair of Qpeak-VOLdir	BIVARIATE: 20 random pairs of Qpeak-VOLdir	BIVARIATE
		
2A) Estimation of Direct and Base Peak Flows, using the Qpeak (Total Peak Flow)		2B) Estimation of the adimensional time to peak and peak flow, using a-b
		$t_{peak} = \frac{a - 1}{a + b - 2}$ $f(t_{peak}) = \frac{(a - 1)^{(a-1)} * (b - 1)^{(b-1)}}{(a + b - 2)^{(a+b-2)} * B(a, b)}$
3A) Magnitude of Direct Flow Hydrograph		3B*) Selection of “critical” flood shapes: 50 (“critical” in terms of maximum reservoir levels)
		
		3B**) Shape of Direct Flow Hydrograph
		

4) Estimation of the duration: Combination between flood magnitude and shape

$$Q_{peak_dir} = f(t_{peak}) \frac{VOL_{dir}}{Duration}$$

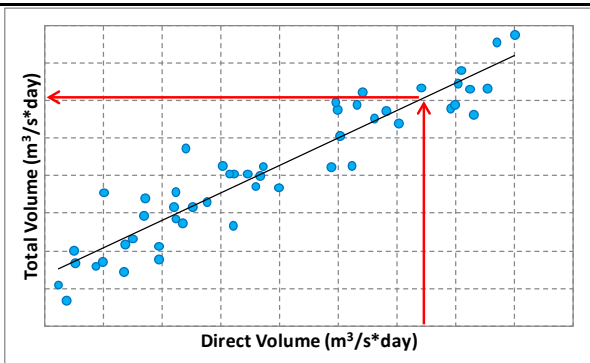
$$Duration = f(t_{peak}) \frac{VOL_{dir}}{Q_{peak_dir}}$$

5) Estimation of Direct Flow Hydrograph



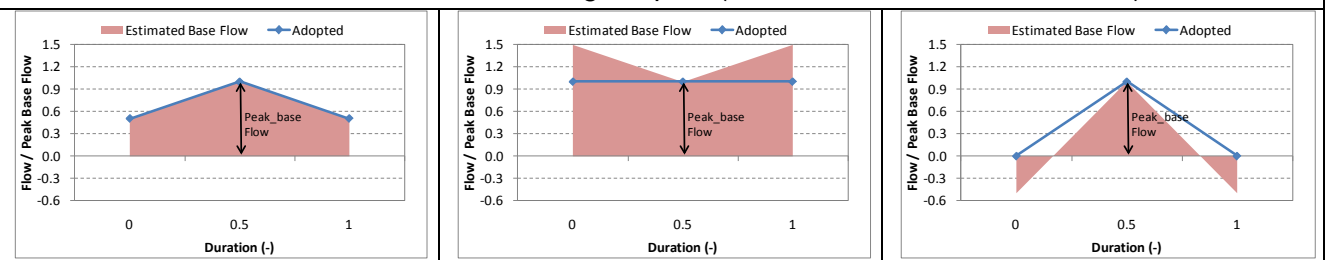
6) Estimation of Volume corresponding to the Base Flow:

Total Volume = Base Volume + Direct Volume

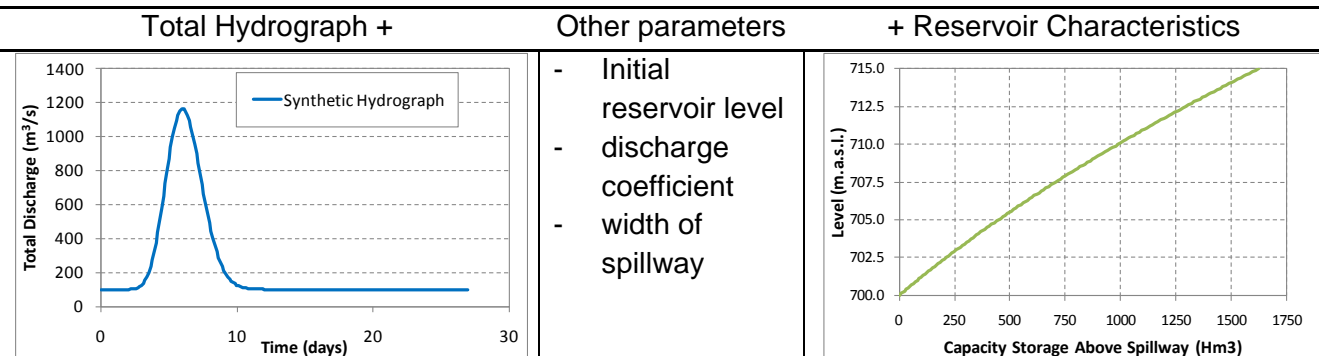


7) Adding up the Base flow, using the Base Peak Flow, Base Volume and Duration of the event.

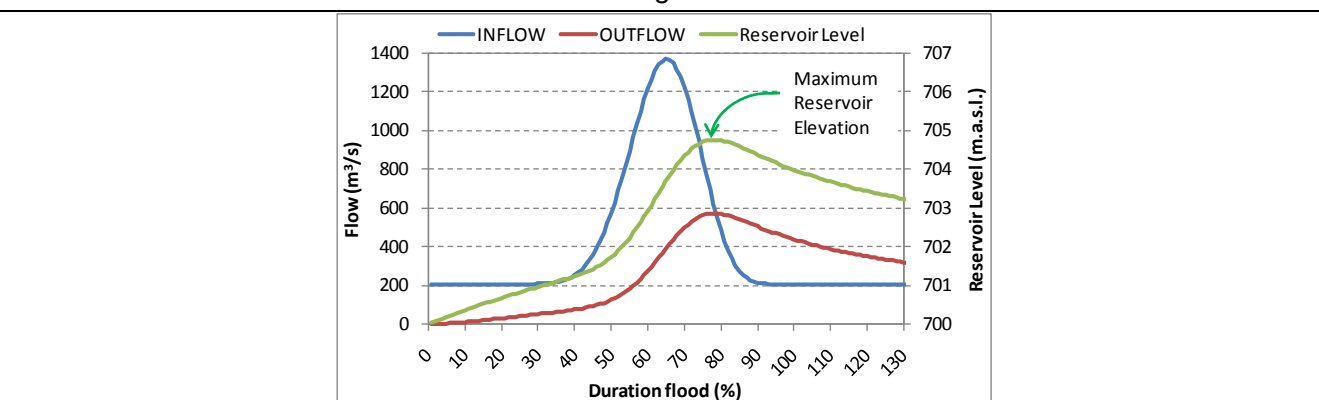
The base flow will take one of the following shapes. (for further details see Table 3.2)



8) Routing of the total hydrograph through the reservoir



Resulting Curves



9) Frequency of Maximum Reservoir levels and outflows

