INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Design of the storage capacity of artificial reservoirs

Artur Tiago Carvalho de Freitas Silva

Dissertation for obtaining the degree of Master in

Civil Engineering

Jury

President: José Câmara, PhD
Supervisor: Maria Manuela Portela, PhD
Member: António Carmona Rodrigues, PhD

October 2010

Acknowledgments

First and foremost, I wish to express my sincerest gratitude to Professor Maria Manuela Portela, whose expert guidance, confidence, and encouragement from the earliest to the concluding stages of the research, enabled me to fully develop an understanding of the subject.

I am also grateful to the Portuguese Water Resources Association (APRH) for awarding an honorable mention to a paper written in the scope of this dissertation. That recognition constituted an additional source of motivation during the development of my research.

A word of gratitude is also due to my friends and colleagues Ana Margarida Ricardo, Filipa Leite Rosa, João Véstia, José Miguel Delgado, José Pedro Matos, Pedro Morgado and Ricardo Canelas, whose assistance, advices and criticism were of great value to the outcome of this thesis.

Ana Filipa Dionísio deserves a special word of acknowledgment for being a continued source of cheer and inspiration.

Last but not the least, I would like to thank my family: my parents, Carlos and Maria, my brother Nuno and my sister Sofia, for providing a constant encouragement and support, at many levels, and from a distance.

Abstract

This study aims to establish design criteria that simultaneously meets the stochastic nature of the streamflow regime in Portuguese rivers and the dependency between such regime and the mean annual flow depth, in view of the preliminary design of the storage capacities of reservoirs. The data set consisted of 54 streamflow samples in Portuguese rivers. To comply with the previously stated objectives a procedure was implemented to generate annual and monthly synthetic streamflow series using, at the annual level, a probabilistic model based on the random sampling of the log-Pearson III distribution, and, at the monthly level, a disaggregation model, namely the method of fragments. For this purpose, a procedure was developed and tested for the automatic definition of the classes of fragments, thus conferring greater generality and robustness to the method. Such procedure was applied to generate a large number of synthetic series of monthly flows that preserved the statistical characteristics of the samples. For different operating conditions regarding the demand and performance, a simulation algorithm was applied to each synthetic series to estimate the storage capacity that would be required if an artificial reservoir for water supply was built at the location pertaining to each sample. The resulting collection of estimates of storage capacities were statistically analyzed using the Gumbel distribution in order to estimate the design storage capacities with specified values of reliability, defined as a non-exceedance probability. As a result, it was concluded that the design values of specific storage capacities are related to the mean annual flow depth of the sample series with significant correlation coefficients. The curves that express this relationship may be directly applied to the preliminary design of the storage capacity of reservoirs in Portuguese rivers. It is assumed that similar curves can also be applied in Southern European countries with hydrologic characteristics similar to those of Portugal.

Keywords: Reservoir capacity design, mean annual flow depth, synthetic streamflow series, method of fragments, Gumbel law, log-Pearson III law.

Resumo

O presente estudo visou estabelecer critérios fundamentados de projecto que simultaneamente atendem à natureza estocástica do regime fluvial dos cursos de água de Portugal Continental e à dependência entre tal regime e a altura do escoamento anual médio, tendo em vista o pré-dimensionamento de capacidades úteis de albufeiras. Os dados de base utilizados foram constituídos por 54 amostras de escoamentos mensais em rios portugueses. Para cumprir o anterior objectivo implementou-se um procedimento para gerar séries sintéticas de escoamentos anuais e mensais, utilizando, ao nível anual, um modelo probabilístico baseado na amostragem aleatória da lei log-Pearson III e, ao nível mensal, um modelo de desagregação, designadamente o método dos fragmentos. Para o efeito, foi desenvolvida e testada uma metodologia de definição automática das classes de fragmentos que reduz a intervenção subjectiva do modelador, conferindo maior generalidade e robustez ao método. Tal metodologia foi aplicada para gerar um número elevado séries sintéticas de escoamentos mensais, estatisticamente semelhantes às amostras. Considerando diferentes valores para o fornecimento de água e para as condições desse fornecimento, estimaram-se, por aplicação de um algoritmo de simulação a cada série sintética de escoamentos, as capacidades úteis que seriam requeridas se, no local a que se refere cada amostra, fosse construído um reservatório destinado ao abastecimento de um dado pedido uniforme da água. As estimativas assim obtidas no conjunto das séries sintéticas foram estatisticamente analisadas por aplicação da lei de Gumbel, com estimação das regularizações específicas associadas a diferentes garantias, entendidas como probabilidades teóricas de não-excedência. Em resultado do estudo efectuado, concluiu-se que as regularizações específicas de que é necessário dispor em Portugal Continental se relacionam com as alturas do escoamento anual médio com correlações notáveis, tendo-se estabelecido as curvas que exprimem essa dependência. Tais curvas podem ser aplicadas directamente, em fases preliminares de projecto, no dimensionamento de capacidades úteis de albufeiras criadas por barragens em rios portugueses. Admite-se que tenha sentido desenvolver procedimentos análogos aos propostos em países do Sul da Europa, com características hidrológicas semelhantes às que ocorrem em Portugal.

Palavras-chave: Dimensionamento da capacidade útil de albufeiras, altura do escoamento anual médio, séries sintéticas de escoamentos, método dos fragmentos, lei de Gumbel, lei log-Pearson III.

Contents

List of Tables

List of Figures

List of Symbols

Uppercase latin

C	Storage capacity
$\frac{C}{Q}$	Specific storage (percentage of the mean annual inflow
C_V	Coefficient of variation
D	Demand desired yield of a reservoir
$\frac{D}{Q}$	Draft (percentage of the mean annual inflow)
D_t	Demand during the tth time period
D'_t	Actual supply during the tth time period
D_k	Demand during the kth failure sequence
E_t	Losses by evaporation at time t
ER	Empirical reliability
F	Non-exceedance probability
H	Annual flow depth
$\overline{H}$	Mean annual flow depth
K	Total number of local minimums in a cumulative flow curve technique
L_t	Other losses from the reservoir at time t
M_k	k$^{\text{th}}$ Local maximum in a cumulative flow curve technique
MSD_i	Mean square deviation (of monthly or daily flows) in year i
$\overline{MSD}$	Mean of MSD_i
N	Total number of time periods
N	Mumber of years of a streamflow series
N_{fs}	Number of failure sequences
N_r	Number of failure periods
P_t	Precipitation over the reservoir at time t
$PRNG$	Pseudorandom number generator
Q	Inflow
Q_t	Inflow in time period t

$\overline{Q}$	Mean inflow
$\overline{Q}$	Mean annual inflow
R^2	Determination coefficient
R_t	Time-based reliability
R_V	Volumetric reliability
S_t	Storage at time t
SPA	Sequent peal algorithm
TR	Theoretical reliability
W	Logarithm of the streamflow
$\overline{W}$	Mean logarithm of the streamflow
X	Streamflow
$\overline{X}$	Mean annual streamflow
X_i	Annual streamflow in year i
$X_{i,j}$	Streamflow in month j of year i
$\breve{X}_m$	Estimated annual streamflow
X_k^*	k^{th} Annual streamflow (sorted from smallest to largest)
$X_{k,j}^*$	k^{th} Monthly streamflows of year k (pertaining to annual the annual streamflow X_k^*
$\hat{X}_i^{(m)}$	Annual streamflow in year i of the mth synthetic streamflow series.
$\hat{X}_{i,j}^{(m)}$	Monthly streamflow in month j year i of the mth synthetic streamflow series.

Lowercase latin

$c.c.$	Correlation coefficient
g_H	Skewness coefficient of the mean annual flow depth
g_W	Skewness coefficient of the logarithms of annual streamflows
g_X	Skewness coefficient of the annual stream flows
i	Year index
j	Month index
k	Failure sequence index
k	Serial correlation lag
k	Year index
ln	Natural logarithm
m_k	k^{th} local minimum in a cumulative flow curve technique
(m)	Synthetic series index (superscripted)
n_{frags}	Number of fragments in a class
r_k	Lag-k serial correlation
s_H	Standard deviation of the annual flow depth

s_{MSD}	Standard deviation of MSD_i
s_W	Standard deviation of the logarithms of the annual streamflow
s_X	Standard deviation of the annual streamflows
sh_k	Water shortage during the k$^{\text{th}}$ failure sequence
$s_{\hat{\theta}}$	Standard deviation of the M values of $\hat{\theta}^{(m)}$
t	Time period index.
$z_{1-\alpha/2}$	$1 - \alpha/2$ quantile of the Normal standard distribution
z_i	i$^{\text{th}}$ normally distributed random variable z, with mean zero and variance one

Uppercase greek

Δt	Time interval
Θ	Function
Ω	Function

Lowercase greek

α	Significance level
α	Parameter
β	Parameter
γ	Number of months in a year
γ	Number of days in a year
ϵ	Error
ε_t	Independent normal variate with mean zero and variance one
ζ_i	Probability factor of the Pearson III distribution obtained by applying the Wilson-Hilferty transformation to z_i
η	Vulnerability
θ	Statistic of the sample series
$\hat{\theta}^{(m)}$	Statistic of the m$^{\text{th}}$ synthetic series
$\overline{\hat{\theta}}$	Mean of the M values of $\hat{\theta}^{(m)}$
μ	Mean of the M estimates of $\frac{C}{Q}$
ξ	Independent normal variate with mean zero and variance $(1 - \phi^2)$
ξ	$U(0,1)$ random variable
σ	Standard deviation of the M estimates of $\frac{C}{Q}$
ϕ	Parameter
ϕ_i	Fragment pertaining to the year i
ϕ_k	Fragment pertaining to the year k (annual streamflows sorted from smallest to largest)

φ Resilience

Other symbols

$[\phi]$ Array of fragments

$1 - \alpha$ Confidence level

1 Introduction

1.1 Scope

The design of the storage capacity of artificial reservoirs is a classic problem in water resources management. From a hydrological point of view, its solution consists of finding the relationships between the inflow characteristics, the reservoir capacity, the controlled release and the performance of the system.

In Mainland Portugal, due to the irregularity of the streamflow regime, the supply of surface water is generally made from artificial reservoirs. Frequently, at preliminary stages of planning of water resources systems, it is necessary to make an expeditious evaluation of the feasibility of providing a given supply in order to match a certain demand. The design criteria available for making this evaluation are admittedly scarce.

Previous research on the subject was carried out in Portugal by Portela & Quintela (2000, 2001, 2002a,b, 2005a,b, 2006a,b), who established that the mean annual flow depth, $\overline{H}$, constitutes a powerful parameter for characterizing the streamflow variability in Portuguese rivers. Furthermore, those authors also showed that the mean annual flow depth is related, with significant correlations, to the storage capacities of reservoirs.

In the understanding that a sample of observed streamflows comprises a single event, namely the historical one, synthetic series obtained from that sample are considered as alternate events with a probability of occurrence that is admittedly equal to that of the observed event. The utilization of synthetic streamflow series is an important and useful tool for practitioners and researchers on water resources management, in that it enables the assessment of the uncertainty associated with natural hydrological phenomena in general, and with the flow regime in particular. By means of statistical analysis, it is possible to assign probabilistic criteria to the performance of the hydraulic systems in whose analysis synthetic flow series were applied.

The generation of synthetic series can be made simultaneously at different time levels, using a disaggregation technique. These techniques encompass a model for generating values at a given time level, namely the year, combined with a model to disaggregate these values into a lower level, namely the month, while achieving a preservation of the main statistical characteristics of the samples at the different time levels, such as the means, the standard deviations, and the skewness coefficients.

1.2   Objectives

Taking into account the relationships identified in previous studies regarding the dependency between the mean annual flow depth and the storage capacities of artificial reservoirs located in Mainland Portugal, the main objective of the research carried out in this dissertation was to provide more reliable and consistent criteria for the design of such storage capacities by taking into account the stochastic behaviour of the natural inflows and by expressing those capacities as a function of:

(a) the demanded supplies of water;

(b) the performance of the reservoirs while supplying those demands;

(c) the reliability associated with the supply of that demand, with the desired performance.

For accomplishing the proposed objective, a storage-yield procedure, based on the simulation algorithm, was applied to synthetic streamflow data, with the aim of estimating the reliability associated with the performances of water resources system.

For obtaining the aforementioned synthetic data, a methodology was conceived and tested for generating annual and monthly synthetic streamflow series, integrating a probabilistic generation model at the annual level, and a disaggregation model - the method of fragments - at the monthly level. For the purpose of applying that methodology to a large number of observed streamflow samples, with different lengths and statistical characteristics, a procedure was developed and tested for the automatic definition of the classes of fragments, resulting in a much more general and robust model, thus reducing the need of intervention of the modeler.

A data set utilized in the research consisted of 54 monthly streamflow samples from gauging stations geographically spread over Mainland Portugal.

1.3   Organization of the document

The present document is organized in six chapters. In Chapter 1 a brief introduction is made on the scope and objectives of the research.

In Chapter 2 a state-of-the-art review is made on the background investigation and theoretical concepts that support the research carried out in this dissertation.

In Chapter 3 the data set used in the research is presented and characterized. As previously stated such set comprehended 54 samples of monthly streamflows.

Chapter 4 describes of the proposed methodology for generating streamflow series and the reservoir storage-yield procedure applied to the data set.

The analysis and discussion of the obtained results is carried out in Chapter 5. The huge amount of information that supports such results should be stressed. In fact, in addition to the samples themselves, the results are supported by 1200 synthetic monthly streamflow series with a length equal to that of the corresponding sample, for each one of the 54 samples. As the average length of the streamflow samples is close to 35 years, a total of approximately 27 million monthly streamflows were estimated and analyzed in order to produce relationships between storage capacities, water demands, and reliability of the water resources systems.

Finally, Chapter 6 presents a summary and the conclusions achieved in this study.

2 Background

2.1 Introduction

In this chapter the main concepts that support the research carried out in the study of the storage capacities of artificial reservoirs are presented. Firstly a brief state-of-the-art analysis is made regarding the role of the mean annual flow depth as a regionalization parameter in hydrology. This concept is the basis of the regional study of reservoir storage made by Portela & Quintela (2006b), which, in turn, is the starting point for the research carried out in this dissertation. Then, a review of storage theory regarding the design of the storage capacities, the criteria of reservoir performance, and the measurement of the uncertainty associated with storage estimates is made. Finally, an overview of time-series modeling is presented, followed by a brief description of the methods utilized for generating synthetic time-series.

2.2 The mean annual flow depth as a regionalization parameter

Various studies carried out by Portela & Quintela (2000, 2001, 2002a,b, 2005a,b, 2006a,b), analyzed the relation between the mean annual flow depth, $\overline{H}$, i.e., the mean annual flow expressed as water depth over the watershed, and the temporal irregularity of the flow regimes of Portuguese rivers. The data used in those studies were long samples of the mean daily flow (the longest available records) at 24 stream gauging stations at a first stage Portela & Quintela (2000, 2001, 2002a,b), and at a later stage, at 54 stream gauging stations Portela & Quintela (2005a,b, 2006a,b). The gauging stations used were geographically spread over Mainland Portugal. The studies show that the mean annual flow depth is closely related to the temporal (within-year, as well as annual) variability of the flow regime. Then, $\overline{H}$ is a powerful parameter for the regionalization of hydrometric information (Portela & Quintela, 2006a,b).

The close relation between the temporal variability of the annual flow and the mean annual flow depth was demonstrated by the dependency that is verified between the coefficient of variation, C_V, of the annual flows, and the mean annual flow depth. The coefficient of variation of a series is a measure of its relative variability, and is expressed as a ratio between the standard deviation, s, and the mean of the series. The general form achieved for the equation that expresses the relationship between C_V and $\overline{H}$ is given by Equation (2.1).

$$C_V = \alpha \overline{H}^{-\beta} \tag{2.1}$$

As shown in Table 2.1, as the coefficient α and the exponent β take positive values, Equation (2.1) shows that the variability of the annual streamflow increases as $\overline{H}$ decreases. Then, the annual flows are more irregular when the watershed is located in a dryer area. In Mainland Portugal the dryer watersheds are mainly located in the Southern and Northeastern regions.

Table 2.1: Relationship between the coefficient of variation, C_V of annual flows, and the mean annual flow depth, $\overline{H}$. Values of α and β.

$\alpha = 4.895$	$\beta = 0.354$	(24 gauging stations)	(Portela & Quintela (2000, 2001, 2002a,b))
$\alpha = 4.285$	$\beta = 0.324$	(54 gauging stations)	(Portela & Quintela (2005a,b, 2006a))

The relationship between the relative temporal variability of within-year flows and $\overline{H}$ was described by the mean square deviation, MSD_i, of the monthly and daily flows expressed in a dimensionless form:

$$MSD_i = \sqrt{\frac{\sum_{j=1}^{\gamma} \left(\frac{H_{i,j} - \overline{H}_i}{\overline{H}_i} \right)^2}{\gamma}} \tag{2.2}$$

where:

- at a monthly level $\gamma = 12$ and the indexes i and j denote the year and month, respectively, and:

  MSD_i is the mean square deviation of the monthly flow in year i, in a dimensionless form;

  $\overline{H}_i$ is the mean monthly flow depth in year i;

  $H_{i,j}$ is the flow depth in month j of the year i;

- at a daily level, $\gamma = 365$, i denotes the year, j the day and:

  MSD_i is the mean square deviation of the daily flow in year i, in a dimensionless form;

  $\overline{H}_i$ is the mean daily flow depth in year i;

  $H_{i,j}$ is the flow depth in day j of the year i.

For each of the gauging stations, the number of values of MSD_i is equal to the number of years with flow records. By calculating the mean, $\overline{MSD}$, and the standard deviation, s_{MSD}, of the MSD_i series at the 54 gauging stations, and by plotting those statistics as a function of the mean annual flow depth of the corresponding station, Portela & Quintela (2005a,b) obtained the curves presented in Figure 2.1.
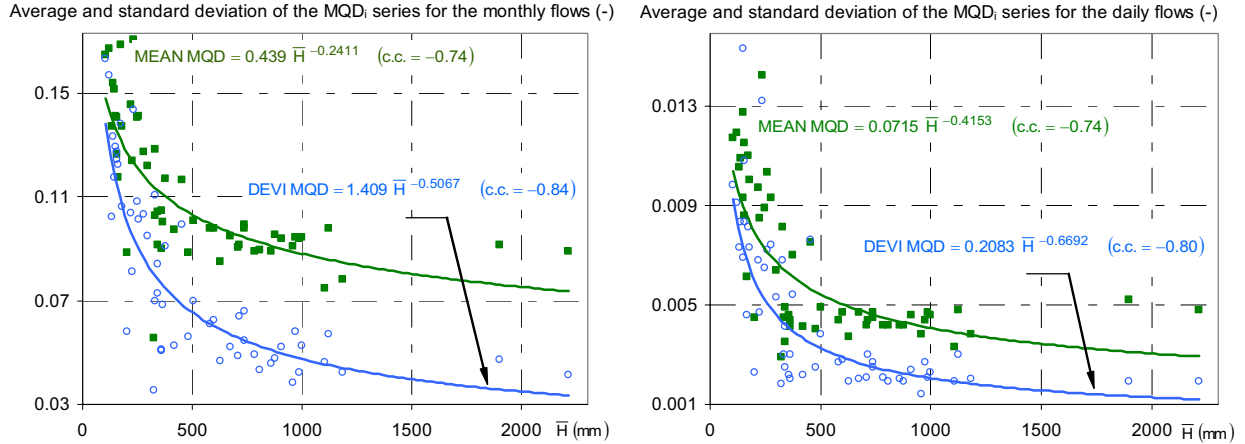
Figure 2.1: Averages and standard deviations of the mean square deviation series, in a dimensionless form, of the monthly flows (left) and of the daily flows (right) (reproduced from Portela & Quintela, 2006b).

Those curves show that, in average, as $\overline{H}$ increases, the relative temporal variability of the within-year flow regime decreases. Likewise, the dispersion of the mean square deviation around the average also decreases. Then it is shown that $\overline{H}$ also provides a measure of the relative temporal variability of within-year flows, which becomes attenuated as $\overline{H}$ increases, hence dryer regions have more irregular within-year flows.

Thus, Portela & Quintela (2000, 2001, 2002a,b, 2005a,b) conclude that, in Mainland Portugal, annual, monthly and daily flows at watersheds with close enough mean annual flow depths are similar, provided those series are expressed in dimensionless forms. These authors also believe that this similarity can be generalized to European regions having similar climatic characteristics, with emphasis to the Southern European zones (Portela & Quintela, 2006b).

2.3 Design of the storage capacity of reservoirs

2.3.1 Description of the problem and definition of terms

The design of the storage capacity of a reservoir is an old problem in water resources management. The unconstrained form of this problem poses the question: "how large must the storage capacity of a reservoir be in order to provide a steady supply of water of a demanded magnitude?". In this form, the problem may not have a solution. An obvious constraint that must be made to the problem is that the demanded supply must not be greater than the mean discharge of the stream, for it is impossible for a reservoir of any size to supply more water than it receives over a long period of time. Another constraint that must be made is that a reservoir may not be able to guarantee an uninterrupted provision even if the demand is smaller than the mean inflow discharge, due to the stochastic nature of the streamflow process. Because there is no record of

7

future inflows, the traditional approach to this problem has been to determine the storage capacity of the reservoir on the basis of the past streamflow record, and assume that the future will not be much different from the past (Klemeš, 1987).

The solution of the reservoir storage problem depends on several factors, such as the variability of the natural streamflows, the size of the demand, and the desired performance of the reservoir in meeting such demand. While it may be formulated in many different manners, it may be reduced to a basic problem: finding the relationship between inflow characteristics, reservoir capacity, controlled release, and the desired performance (McMahon & Adeloye, 2005, p. 42). The methods employed to estimate a solution of a reservoir storage problem are denominated reservoir storage-yield procedures.

The terminology involved in reservoir storage analyses may differ from author to author, hence it is necessary to categorize the terms used in this text as to which interpretation is adopted:

- The *storage capacity*, or *active storage* of a reservoir is the volume of water stored above the level of the lowest offtake. It is the total volume of water in the reservoir subtracted of the *dead storage* (volume of water below the lowest offtake).

- *Carryover storage* is the volume of water stored at the end of one year being transferred to the following year.

- *Yield* is the regulated flow supplied from a reservoir during a given time period.

- The *demand*, *demanded volume*, or *target*, is the desired yield of the reservoir. The synonymous term *draft* will be used in this text, for expressing the demand percentage of the mean inflow volume.

- *Spill* is regarded as the uncontrolled release of water from a reservoir. It occurs when the water stored in the reservoir is above the full supply level.

The terminology concerned with the performance of the reservoir will be addressed in 2.3.3.

2.3.2 Reservoir capacity-yield procedures

McMahon & Mein (1978) propose the classification of reservoir capacity-yield procedures into three main groups:

1. Critical period techniques.

2. Probability matrix methods.

3. Procedures based on generated synthetic data.

The methods of the third group are, however, essentially the same as in the first two groups, the only difference being that the input streamflows are synthetically generated rather than historical samples. McMahon & Adeloye (2005, p. 6), note that all techniques could be based on generated data.

The probability matrix methods are related to *Moran dam theory* (Moran, 1959). These methods will not be used in the dissertation, hence they will not be presented in this section. An exhaustive analysis of these methods may be found in McMahon & Mein (1978) and in McMahon & Adeloye (2005).

Critical period methods use historical or generated inflows and a projected demand to simulate the volumetric behaviour of a reservoir, that is, the time series of storage fluctuations versus time. Although there is no unique definition for *critical period*, it may be defined as the period during which the reservoir goes from full to empty without spilling (McMahon & Mein, 1978). In this thesis two types of critical period methods are presented: cumulative flow curve techniques and simulation (or behaviour) analysis. An up-to-date and more complete study of critical period methods for determining the storage capacity of artificial reservoirs can be made by consulting McMahon & Adeloye (2005) and Guimarães (2005).

Cumulative flow curve techniques

Cumulative flow curve techniques were originally developed as graphic methods that used the curve of accumulated flows of the streamflow time series to determine the storage capacity of a reservoir. This concept of design was introduced by Rippl (1883), and marks the beginning of rigorous reservoir storage theory (Klemeš, 1987; McMahon *et al.*, 2007a). Before Rippl's paper, the design of the size of a reservoir was governed by the sole criterion that it should be large enough to meet the demand in the driest year of the historical record. Rippl recognized the inadequacy of such approach and pointed out that the reservoir may not be filled up between two successive dry years and that a sequence of scarce streamflow periods would produce a negative cumulative effect, hence the reservoir inflow must be considered as a time series, rather than a separate treatment of annual inflows (Klemeš, 1987). At the time of Rippl's publication, numerical computations were done manually and were time consuming, and graphical methods were the main tool for increasing computational efficiency, therefore Rippl's method represented a considerable achievement.

The objective of a cumulative flow curve technique is to determine the minimum storage capacity, C, necessary to guarantee a full supply of a given demand, D, of a reservoir with a given inflow time series, Q, of length N. The methods have two basic assumptions: (i) that the reservoir is full at time zero, and (ii) that the historical inflow sample is representative of the future inflows.

The following is a description of four different graphical methods. It should be noted that while the graphical representation may differ from method to method, in the end the numerical results are concurrent. Each method's algorithm is essentially identical, and may be adapted for execution on a computer.

- *Rippl's method* consists of the following (Fig: 2.2):

  1. Draw the *net cumulative inflow curve* - the difference between cumulative inflows, Q and Q_t, and outflows or demand, D and D_t, t being the time period index and N the total number of time periods:

$$\sum(Q - D) = \sum_{t=1}^{N}(Q_t - D_t), \quad t = 1, 2, ..., N \tag{2.3}$$

  2. Moving backward from the end of the curve, locate the first local minimum, m_1.

  3. Locate the second local minimum, m_2.

  4. Locate the local maximum, M_1, between m_1 and m_2, and calculate the difference $(M_1 - m_1)$.

  5. Repeat the steps 2 through 4 until the beginning of the curve is reached.

  6. Calculate the storage capacity as $C = \max(M_k - m_k)$, with $k = 1, 2, ..., K$, where K is the total number of local minimums.

The main disadvantage of this approach is that the net cumulative inflow curve must be completely redrawn every time the demanded yield is altered.
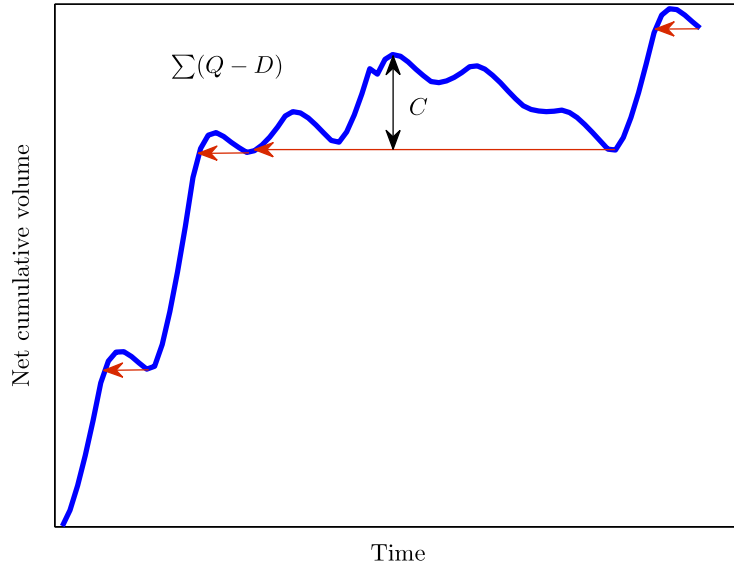
Figure 2.2: *Rippl's method* for determining the storage capacity of a reservoir, C, fed by the inflow series Q, necessary to continuously supply the target demand D.

- *Mass inflow curve method*, or simply *mass curve method* (Figure 2.3). This method is commonly, but incorrectly, referred to as 'Rippl's method' or 'Rippl diagram' (McMahon & Adeloye, 2005, p. 66;Klemeš, 1987). It may be consulted on many textbooks, such as McMahon & Mein (1978), under that denomination. The method consists of the following steps, where the variables have the previously presented meaning:

1. Draw the *mass or cumulative inflow curve*:

$$\sum Q = \sum_{t=1}^{N} Q_t, \quad t = 1, 2, ..., N \tag{2.4}$$

2. Draw a line of the cumulative demands:

$$\sum D = \sum_{t=1}^{N} D_t, \quad t = 1, 2, ..., N \tag{2.5}$$

3. Superimpose on the mass curve parallels of the demand line, tangential to each bump of the mass inflow curve.

4. Measure the successive differences between the parallels and the mass curve.

5. Determine the reservoir capacity corresponding to the maximum of the differences measured on the last step.

This method has the advantage of avoiding to redraw the mass curve every time a different demand is considered.
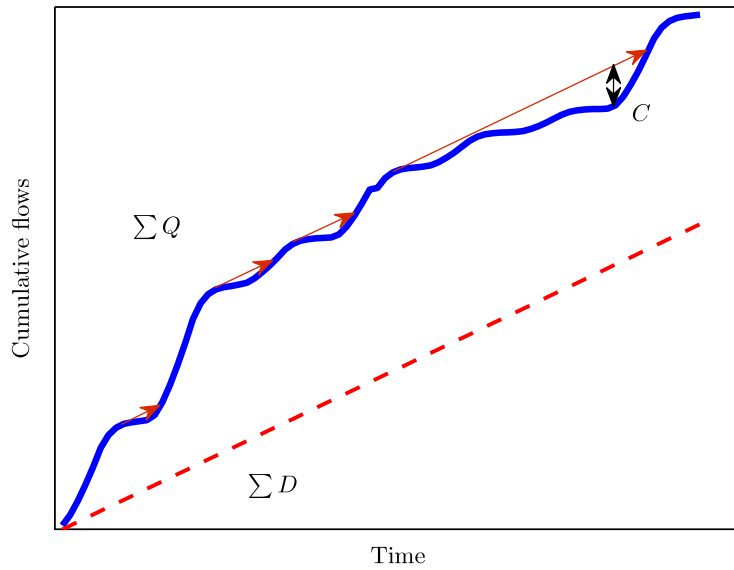
Figure 2.3: *Mass inflow curve method* for determining the storage capacity of a reservoir, C, fed by the inflow series Q, necessary to continuously supply the target demand D.

- *Residual mass curve method*, proposed by Sudler (1927) (Figure 2.4). The steps of the method's application are the following, where except for $\overline{Q}$, the variables have the previously presented meaning, $\overline{Q}$ being the mean of the inflow series.:

1. Draw the residual mass curve, obtained by subtracting the mean of the inflow series, $\overline{Q}$, to each of the series' entry:

$$\sum (Q - \overline{Q}) = \sum_{t=1}^{N} (Q_t - \overline{Q}), \quad t = 1, 2, ..., N \tag{2.6}$$

2. Superimpose on the residual mass curve, parallels of the residual cumulative demand line, obtained by subtracting the mean of the inflow series to the cumulative demand line:

$$\sum (D - \overline{Q}) = \sum_{t=1}^{N} (D_t - \overline{Q}), \quad t = 1, 2, ..., N \tag{2.7}$$

3. Measure the successive differences between the parallels and the residual mass curve.

4. Determine the reservoir capacity corresponding to the maximum of the differences measured on the last step.
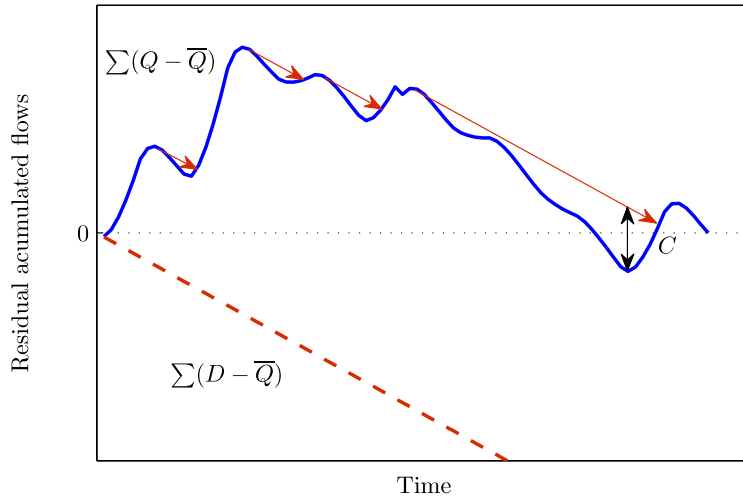
Figure 2.4: *Residual mass curve method* for determining the storage capacity of a reservoir, C, fed by the inflow series Q, necessary to continuously supply the target demand D.

- *Sequent peak algorithm* (SPA). This method, proposed by Thomas & Burden (1963), is very similar to Rippl's method, the only difference being that it moves forward from peak to peak, while Rippl's moves backwards (Figure 2.5):

1. Draw the net cumulative inflow curve - the difference between cumulative inflows and outflows according to Equation (2.3);

2. Moving forward from the beginning of the curve, locate the first local maximum, M_1.

3. Locate the second local maximum, M_2.

4. Locate the local minimum, m_1, between M_1 and M_2, and calculate the difference $(M_1 - m_1)$.

5. Repeat the steps 2 through 4 from peak to peak until the end of the curve is reached.

6. Calculate the storage capacity $C = \max (M_k - m_k)$, with $k = 1, 2, ..., K$, where K is the total number peaks.
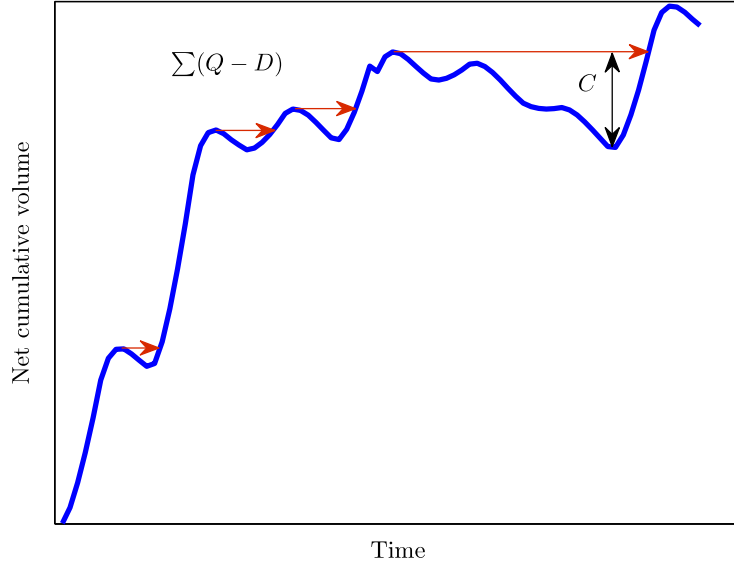
Figure 2.5: *Sequential peak algorithm* for determining the storage capacity of a reservoir, C, fed by the inflow series Q, necessary to continuously supply the target demand D.

There is a limitation inherent to the mass curve methods described, related to the assumption that the reservoir will be full at the beginning of its exploration. This practice tends to underestimate the storage capacity of the reservoir since it is seldom feasible (or even possible) to start the operation of a reservoir when its storage is at full capacity. To mitigate this problem, a two-cycle computation may be applied to the methodology, that is, given an N year period with records of streamflows, consider the inflow series Q_t, with $t = 1, 2, ..., 2N$, and $Q_{N+1} = Q_1, Q_{N+2} = Q_2, ..., Q_{2N} = Q_N$. Thomas & Burden (1963) incorporated the two-cycle computation into the sequent peak algorithm.

Figures from 2.2 to 2.5 represent applications of cumulative curve techniques considering uniform demands. However, it is possible to consider a seasonal structure of the draft. The designer may opt for a yearly, monthly, or daily time-step for the discretization of the inflow time series. Technical literature recommends the adoption of a monthly time-step for storage-yield procedures (McMahon & Adeloye, 2005, p. 24).

Simulation

The simulation method applied to the design of the storage capacities of artificial reservoirs is based on the replication of the reservoir exploitation during a period of time equal to the length of the inflow time series. The technique allows for considering the evaporation and other losses from the reservoir and, unlike the cumulative curve methods, it is useful in determining the performance of the reservoir, because it allows the designer to consider restrictions on the water supply when the reservoir is depleted, while the cumulative methods presuppose a complete fulfillment of the demand. Simulation is a trial and error method in which

a guess is made about the capacity, the exploitation of the reservoir is made for this capacity, and finally relevant performance measure are evaluated (McMahon & Adeloye, 2005, p. 114).

The simulation is performed along consecutive time periods, each one with the constant duration Δt. An initial value for the storage capacity is assumed, and the reservoir is considered to be full at time zero. The stored volume in each new time step calculated for each time step is computed by applying the mass balance equation:

$$S_{t+1} = S_t + Q_t - D_t - E_t + P_t - L_t \tag{2.8}$$

with the restriction:

$$0 \leq S_{t+1} \leq C \tag{2.9}$$

where S_t and S_{t+1} are the storage volumes at the beginning of time periods t and $t+1$, respectively, with $t+1 = t+\Delta t$. The remaining variables represent changes in the storage volume during the tth time period: Q_t is the inflow volume, D_t is the target demand, E_t represents the losses by evaporation, P_t is the precipitation over the reservoir, and L_t are other losses from the reservoir.

The simulation method is suited to make a behavioural analysis of the performance of the reservoir. This subject will be addressed in 2.3.3.

If this method is applied under the same assumptions as any of the cumulative curve methods, that is to say, considering the evaporation and other losses as null, and a guaranteed full supply, the results will coincide with those provided by the cumulative curve methods.

2.3.3 Reservoir performance criteria

In the design of the storage capacity of an artificial reservoir, it is important to evaluate how the system will perform under the expected supply demands and hydrologic conditions during its operating life. The criteria applied to measure the expected performance of a reservoir aim at characterizing particular aspects of unsatisfactory operating conditions, namely during low flow periods (McMahon & Adeloye, 2005, p. 12).

The most common criterion used for assessing the performance of a reservoir is reliability. Though there is no single definition of the reliability of a reservoir, it can be considered as a measure of the dependability of the system's requirements being met (Nagy *et al.*, 2002, p.68). The most common definition presented in the technical literature (McMahon & Mein, 1978, p. 17; McMahon & Adeloye, 2005, p. 13; Nagy *et al.*, 2002, p. 68; McMahon *et al.*, 2007b;McMahon *et al.*, 2007a) defines the *reliability*, R_T, as the percentage of time units in which the specified demand is met:

$$R_T = \left(1 - \frac{N_r}{N}\right) \times 100 \tag{2.10}$$

where N_r is the number of periods where the supply is not able to meet the demand and restrictions on water use are made (*failure* in the supply) and N is the total number of time periods in the analysis. Then N_r/N is the empirical probability of failure of the demand being met, and R_T is its complement (non-failure probability). It should, however, be stressed that this concept deals with an empirical frequency instead of a theoretical probability.

The quality of the information given by the reliability as formulated in Equation (2.10) is questionable, mainly because it masks two important aspects of the behaviour of the system: the duration of the water shortage event (i.e.: a continuous sequence of failures), and the volume shortfall associated with each failure. In water supply, it is better to cope with a long series of minimal shortfalls that could easily be absorbed by the system, than to cope with a single shortfall of a crippling magnitude (McMahon & Adeloye, 2005, p. 88), yet failures are statistically weighed independently of the volume shortage. It is also a non-realistic concept, for in the real operation of a reservoir, restrictions on water supply would be applied long before it would be permitted to be empty.

The concept of reliability expressed by Equation (2.10) will be referred to as *time-based reliability* for the remainder of this dissertation. It is the performance criterion used by Portela & Quintela (2006a,b).

An alternative definition of reliability is the *volumetric reliability* proposed by McMahon & Mein (1978, p. 17) and McMahon & Adeloye (2005, p. 15):

$$R_V = \left(1 - \frac{\displaystyle\sum_{t=1}^{N_r}(D_t - D'_t)}{\displaystyle\sum_{t=1}^{N} D_t}\right) \times 100 \tag{2.11}$$

where D_t and D'_t are, respectively, the target demand and the actual supply during the tth time period, and the remaining variables have the same meaning as before. Note that if D_t is 100% satisfied, $D_t = D'_t$.

Another reservoir performance criterion is the *vulnerability* which measures the severity of shortfalls, based on the assumption that the period with the largest shortfall will be the most severe in terms of its impacts on water supply. Hashimoto *et al.* (1982) define the vulnerability, η, of a reservoir as the average of the maximum shortfalls occurring in each of the continuous failure sequences (McMahon & Adeloye, 2005, p. 17):

$$\eta = \frac{\displaystyle\sum_{k=1}^{N_{fs}} \frac{\max(sh_k)}{D_k}}{N_{fs}} \tag{2.12}$$

15

where N_{fs} is the total number of failure sequences, max sh_k is the maximum water shortage in the kth failure sequence, and D_k is the supply demand during the same sequence. The division by the demand, D_k is merely a formality in order to obtain a dimensionless descriptor.

The criterion to evaluate the ability of recovery of a reservoir from a failure sequence is the *resilience*. There are many definitions of the resilience (McMahon & Adeloye, 2005, p.16), but the most widely adopted was the one proposed by Hashimoto *et al.* (1982):

$$\varphi = \frac{N_{fs}}{N_r} \tag{2.13}$$

where φ is the resilience, N_{fs} has the same meaning as in Equation (2.12), and N_r has the same meaning as in Equation (2.10). By this definition, the resilience is the reverse of the average duration of failure sequences (N_r/N_{fs}), consequently the longer the average duration of failure sequences, the more difficult it is for a reservoir to recover from a failure.

The concepts of time-based reliability, volumetric reliability, vulnerability, and resilience provide information on different characteristics of the behaviour of a reservoir. In fact:

- time-based reliability characterizes the frequency of failures;
- volumetric reliability and vulnerability characterize the magnitude of the shortfall volumes;
- resilience characterizes the duration of the water shortage events.

Each of these characteristics should be considered when making a proper assessment of the performance of a reservoir, yet they cannot by ascertained by a single criterion (Vaz, 1984, p. 165) .

It is important to note that not all reservoir storage-yield procedures can be subjected to an analysis under these criteria. A cumulative flow curve method, for instance, designs a reservoir for the full supply of the target demand, *i.e.*, a failure-free performance. The simulation method is, of the foregoing procedures, the sole which is suited to make a behavioural analysis.

McMahon *et al.* (2006) made an extensive examination of different reservoir performance metrics, using both historical and stochastically generated streamflow data as inflows to hypothetical reservoirs in four rivers – Earn river in the United Kingdom, Hatchie river in the United States, Richmond river in Australia and the Vis river in South Africa.

One of the conclusions of that study is that the vulnerability η and the resilience φ, as formulated by Equations (2.12) and (2.13), respectively, do not have a monotonic behaviour with an increasing storage capacity for a fixed yield, nor with an increasing yield with a fixed storage capacity. This results from the effect of averaging in both criteria's definition: the vulnerability metric η can increase with a reduction of the number of failures sequences N_{fs} without a significant reduction of the numerator of Equation (2.12), meaning that,

for a fixed reservoir capacity, decreasing the draft would result in a more vulnerable system; the resilience estimator φ suffers from the similar averaging in Equation (2.13) as in Equation (2.12). The resulting trends of vulnerability and resilience, relative to the variation of storage capacity with a fixed draft, or *vice-versa*, are distorted and do not have the same qualities as the reliability metrics R_T and R_V, which have a monotonic formulation. This phenomenon is exemplified in Figure 2.6, based on a monthly streamflow sample (from October 1955 to September 2004) from the *Cidadelhe* gauging station at the *Côa* river in Portugal, considering a fixed draft of 60% and increasing values of the storage volume. In order to simplify the comparison of the trends displayed by all four performance metrics in Figure 2.6, the vulnerability, η, and the resilience, φ, are represented as percentages rather than fractions.
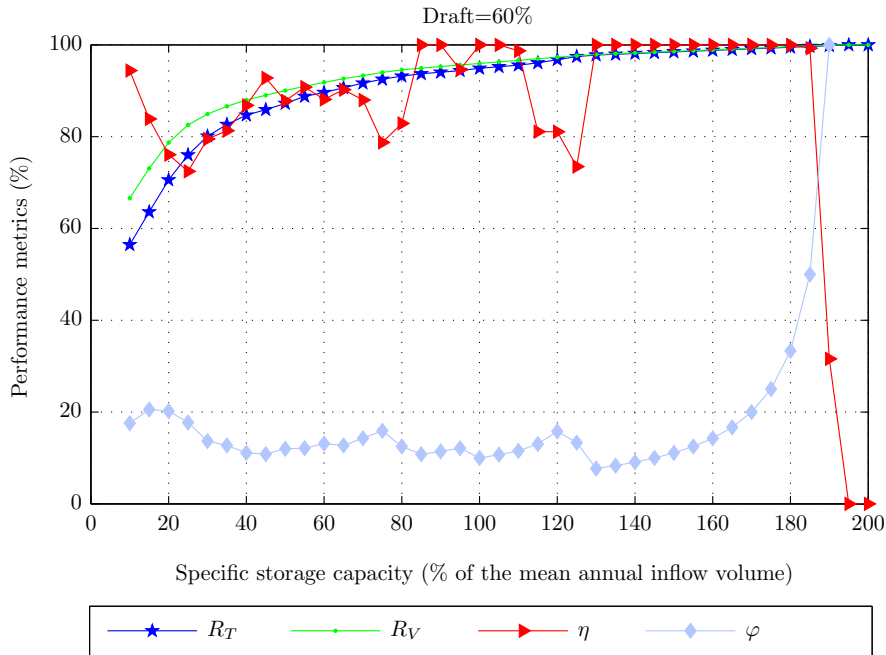
Figure 2.6: *Côa* river in Portugal. Comparison of four reservoir performance metrics for increasing storage capacities and a fixed draft of 60% (R_T - time-based reliability; R_V - volumetric reliability; η - vulnerability; φ - resilience).

Furthermore, McMahon *et al.* (2006) studied the hypothesis that the resilience φ and the vulnerability η have a complementary relationship. Those authors applied behaviour analyses to the previously mentioned four rivers (Earn, Hatchie, Richmond, and Vis) considering different values of the storage capacities and demanded drafts, in a total of 31 reservoir storage-yield combinations. By plotting the values of vulnerability against the values of resilience, McMahon *et al.* (2006) obtained the results of Figure 2.7.

The fitted lines in Figure 2.7 are:

- historical data: $\eta = 1.019 - 1.028\varphi$ $(R^2 = 0.97)$;

- stochastic data: $\eta = 1.034 - 1.048\varphi$ $(R^2 = 0.98)$;

which those authors stated that can both be approximated to $\eta \approx 1.00 - \varphi$, which implies that only one of them should by explicitly estimated during storage-yield analyses. R_2 stands for the determination coefficient.
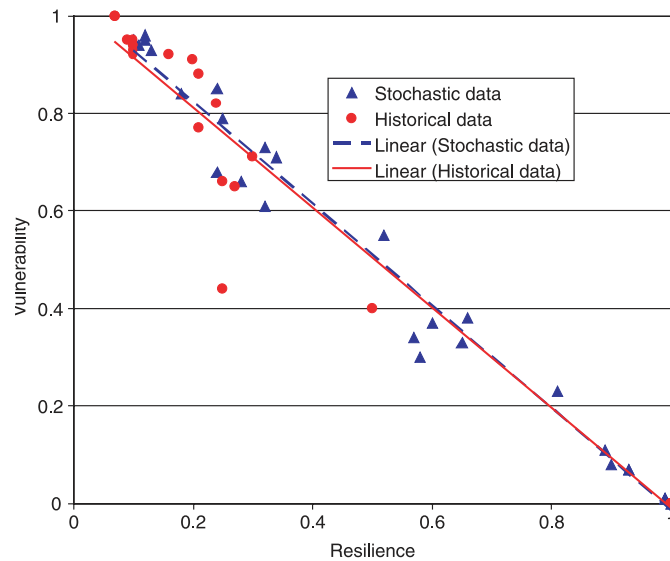
Figure 2.7: Relationship between vulnerability, η, and resilience, φ, for hypothetical reservoirs located on four rivers, based on a behaviour analysis for both historical and generated streamflows (reproduced from McMahon *et al.*, 2006).

McMahon *et al.* (2007b) repeated this analysis, based on a much larger data set than that used in McMahon *et al.* (2006) (a global set of 729 unregulated rivers with at least 25 years of continuous data). This analysis considered the storages equal to the mean annual flows and values of the draft of 30%, 50% and 75% of the mean annual flow. Figure 2.8 shows the values of vulnerability plotted against the values of resilience and the curves fitted to those results.

The analysis presented in Figure 2.8 suggests that the complementary relationships between these two metrics (slope of -1.00) is not particularly strong. Furthermore McMahon *et al.* (2007b) conclude that the line of best fit through all the data in the figure has a slope of -0.76 and that the relationship is not very linear.

18
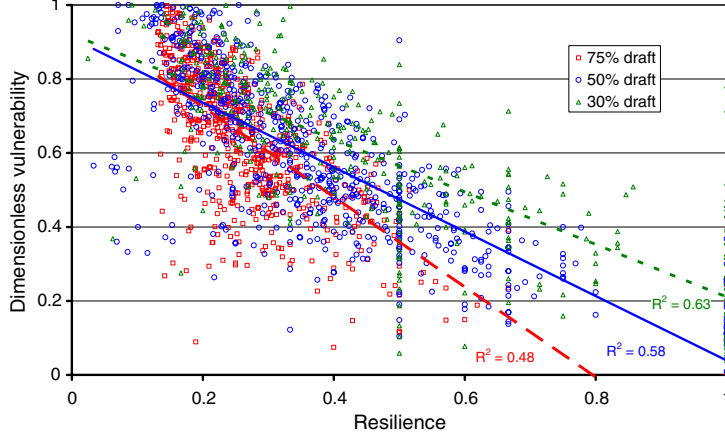
Figure 2.8: Relationship between vulnerability, η, and resilience, φ, for hypothetical reservoirs located on 729 rivers, for storages equal to mean annual flow and for 30%, 50% and 75% target draft (reproduced from McMahon *et al.*, 2007b).

2.3.4 Uncertainty associated with estimates of storage capacities

If the demanded draft of a reservoir is above a certain percentage of the mean annual flow, the reservoir will need resort to over-year regulation. This means that the designed storage capacity should allow for the inflows of a wet year to be transferred to dry years. The occurrence of over-year regulation implies that within-year storage is a function, not only of the inflow of that year, but also of the storage that is carried over from the previous year. Consequently an inflow time series constitutes a single event in a reservoir storage problem. Under this understanding, the use of a single streamflow series, as is the historical sample, cannot result in more than one estimate of the reservoir capacity.

Synthetic streamflow series represent alternative events to the historical streamflow series. The subject of generation of synthetic flow series will be addressed in Section 2.4.

The application of a storage-yield procedure to each inflow series, having defined a specific yield and performance, produces a single estimate of the reservoir storage capacity. A collection of estimates of such capacities enables the application of statistical analysis to the definition of the design storage capacity. By this approach, the probability distribution of the storage volumes can be identified. Then, to each collection of estimates of the storage capacity, C, has a probability density function, $f(c)$, and a probability distribution function $F(c) = P(C \leq c)$, the reliability is the probability that the required storage capacity required not be greater than $c*$, hence $reliability = F(c*)$. The probability of failure is the probability that the storage capacity required will be greater than $C*$, hence $risk = 1 - F(c*)$.

It is important not to misinterpret the concept of reliability as a non-exceedance probability of the storage capacity estimates, with the concept of reliability as a reservoir performance criterion, such as time-based reliability, R_T, as formulated in Equation (2.10). Because R_T is based upon a frequency, and not upon a

19
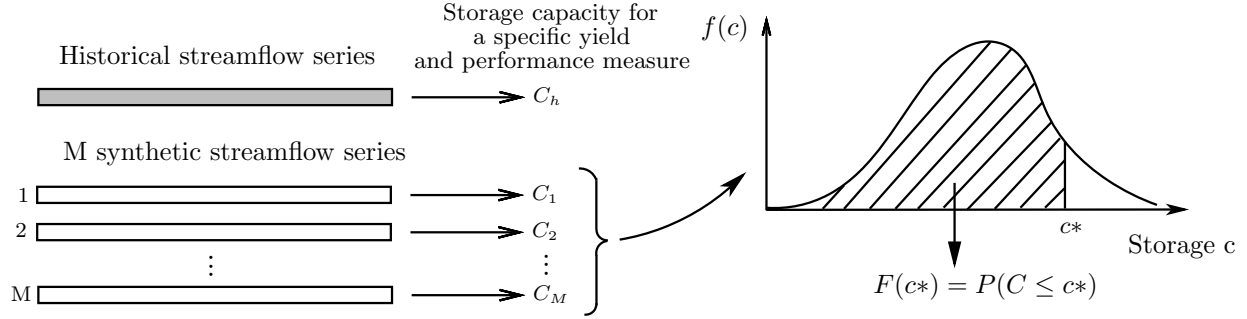
Figure 2.9: Procedure for determining the uncertainty associated with estimates of the storage capacity.

theoretical probability distribution, it is an empirical reliability. This distinction is made because of the need to differentiate this notion of reliability from the probability of non-exceedance, F, of a given storage capacity. F is a theoretical reliability based on the statistical treatment of the storage estimates obtained by applying a behaviour analysis to a large number of synthetic flow series. Therefore, in the context of the design of the storage capacity, these two notions of reliability will be denominated **empirical reliability** (ER), and **theoretical reliability** (TR), and will be expressed as *percentages*, for the remainder of this dissertation.

2.3.5 Previous research on reservoir storage in Portugal

As was introduced in Section 2.2, Portela & Quintela (2000, 2001, 2002a,b, 2005a,b) demonstrated that the mean annual flow depth, $\overline{H}$, is intrinsically related with the temporal variability of the streamflows in Portuguese rivers, therefore it constitutes a a powerful regionalization parameter for hydrometric data.

Using a data set of monthly streamflow samples in 24 gauging stations (Portela & Quintela, 2002a,b), and, later, 54 gauging stations (Portela & Quintela, 2006a,b), those authors applied the simulation algorithm to the historical samples estimate the specific storage capacities, for fixed values of time-based reliability and draft, and plotted the results against the mean annual flow depth of each sample.

The results of these studies allowed the authors to establish the following relationship:

$$\Theta\left(\frac{C}{\overline{Q}}, \frac{D}{\overline{Q}}, ER, \overline{H}\right) = 0 \tag{2.14}$$

where, $\frac{C}{\overline{Q}}$ is the specific storage[1], $\frac{D}{\overline{Q}}$ is the draft, ER is the empirical reliability (time-based reliability, R_T, was used), and $\overline{H}$ the mean annual flow depth.

---

[1]The storage volume is often represented as a percentage of the mean annual inflow. This normalized quantification of the storage estimate is denominated *specific storage*. For the purpose of simplifying the notation, the specific storage will be designated by the fraction $\frac{C}{\overline{Q}}$, albeit representing a percentage. Likewise, the draft, which is also often represented as a percentage o the mean annual inflow, will be represented by $\frac{D}{\overline{Q}}$.

Figure 2.10 shows the results obtained in Portela & Quintela (2006b) for an empirical reliability of 90% and the different drafts considered. It is apparent that, for each pair of values of empirical reliability and draft, the specific storage of the reservoir increases as the mean annual flow depth decreases, seeing that the streamflow regime has more variability in these cases, according to the relationships presented in Item 2.2.

For each reliability-draft combination, those authors obtained equations of the following type with significant correlation coefficients:

$$\frac{C}{\overline{\overline{Q}}} = \alpha \overline{H}^{\beta} \tag{2.15}$$

where the coefficient α is positive and the exponent β is negative.
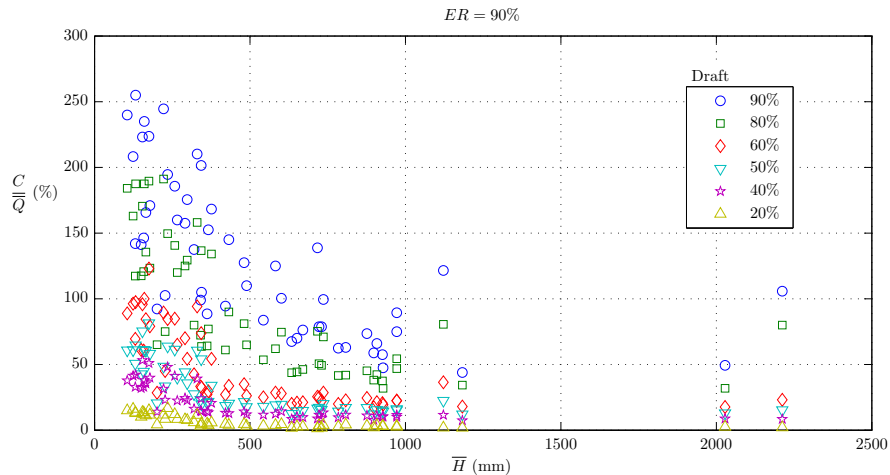
Figure 2.10: Estimates of the specific storages obtained by Portela & Quintela (2006b) for 54 gauging stations. Empirical reliability of 90% and drafts from 20% to 90%.

Figure 2.11 shows the results obtained in Portela & Quintela (2006b) for the empirical reliability of 85% as well as the curves that represent Equation (2.15) for the considered drafts.

Table 2.2 contains the values of the parameters α and β and of the correlation coefficients, *c.c.*, of the curves obtained by Portela & Quintela (2006b).

Portela & Quintela (2006a,b) concluded that those curves could be applied to Portuguese rivers to make a preliminary evaluation of the storage capacity required by a reservoir to supply a target draft with a given empirical reliability.
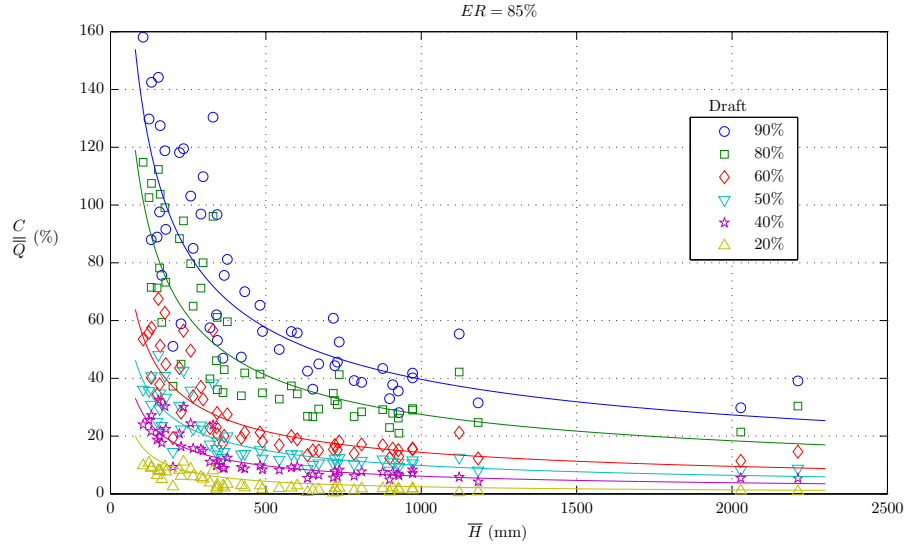
Figure 2.11: Estimates of the specific storages obtained by Portela & Quintela (2006b) for 54 gauging stations. Empirical reliability of 85% and drafts from 20% to 90%. Fitted curves of the type of Equation (2.15).

Table 2.2: Results obtained by Portela & Quintela (2006b). Parameters α and β and correlation coefficients, $c.c.$, of the curves defined by Equation (2.15) for different values of empirical reliability, ER, and target draft, $\frac{D}{\overline{Q}}$.

ER (%)	$\frac{D}{\overline{Q}}$ (%)	α	β	$c.c.$
	90	2141.8	-0.430	-0.747
	80	3134.7	-0.556	-0.800
95	60	3309.8	-0.682	-0.843
	50	2824.2	-0.723	-0.861
	40	2093.3	-0.748	-0.869
	20	1086.7	-0.835	-0.888
	90	2360.4	-0.500	-0.786
	80	2805.4	-0.591	-0.817
90	60	2443.0	-0.687	-0.855
	50	1618.6	-0.681	-0.862
	40	1124.0	-0.693	-0.881
	20	755.1	-0.829	-0.880
	90	1618.1	-0.537	-0.853
	80	1512.1	-0.580	-0.861
80	60	851.5	-0.591	-0.863
	50	679.9	-0.613	-0.881
	40	621.4	-0.669	-0.891
	20	1352.3	-1.025	-0.855

22

2.4 Generation of synthetic flow series

2.4.1 An overview of time series modeling

The process of mathematical modeling in Hydrology can take two approaches: a deterministic approach, and a non-deterministic one. Deterministic models have no element of chance, there is a unique correspondence between a given input and the resulting outputs. Non-deterministic models on the other hand, allow for randomness to induce indeterminacy in the model, so for the same given input the model may take several different paths and produce a different output, though some paths are more probable than others. Deterministic models can be further divided into two groups: empirical models which provide cause-effect relationships between input and output variables and are based on experience, and physically based models which aim to reproduce the physical laws to which the natural processes are subjected to.

Likewise, non-deterministic models are also divided into two groups: probabilistic models when the hydrologic process has a purely random behaviour, and the temporal and/or spatial sequence of the variables may be disregarded; stochastic models, when besides the random component, the process has a deterministic component that allows for the variables' sequence to be preserved.

The aforementioned categorization of mathematical hydrologic models is summarized in Figure 2.12. It is a widely adopted categorization, but it should be noted that some authors consider probabilistic models as a subgroup of stochastic models. This categorization was presented in Quintela & Portela (2002) and is the one adopted in this dissertation.

$$
\text{Mathematical hydrologic models}
\begin{cases}
\text{Deterministic}
\begin{cases}
\text{Empirical} \\
\text{Physically based}
\end{cases} \\[2ex]
\text{Non-Deterministic}
\begin{cases}
\text{Probabilistic} \\
\text{Stochastic}
\end{cases}
\end{cases}
$$
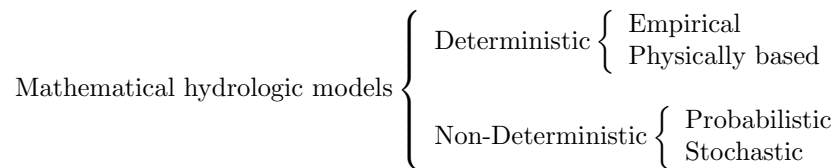
Figure 2.12: Categorization of mathematical hydrologic models, after Quintela & Portela (2002)

Synthetic time series generation is an important tool for practitioners and researchers on water resources management and planning, namely in the field of reservoir storage design. The use of synthetic inflow series in reservoir storage-yield procedures is important for estimating the reliability associated with each particular value of storage capacity.

Consider an observed series of N years of annual streamflow, X_t, with $t = 1, 2, ..., N$. Although the streamflow process is continuous in time, it is usual to consider its observations as discrete variables (Salas *et al.*, 1980, p. 2).

The aim of time series modeling is to find a model that reproduces the statistical properties of the sample.

For instance, if X is normal with mean μ and variance σ^2, the mathematical model that represents this series can be written as

$$X_t = \mu + \sigma\varepsilon_t \quad , \quad t = 1, 2, ... \tag{2.16}$$

where ε_t is normal with mean zero and variance one and $\varepsilon_1, \varepsilon_2...$ are random and X_t is a function only of the random variable ε_t; thus X_t is also random and Equation (2.16) is a non-deterministic probabilistic model with the parameters μ and σ. The model would become stochastic if ε_t in Equation (2.16) would be represented by the stochastic model

$$\varepsilon_t = \phi\varepsilon_{(t-1)} + \xi_t \tag{2.17}$$

where ξ_t is an independent series with mean zero and variance $(1 - \phi^2)$ and ϕ is the parameter of the model. ε_t is a function of both ξ_t and the same variable ε at a time $t-1$, thus it is a stochastic series. In this case the parameters of the model of X_t would be μ, σ and ϕ. Since the parameters of the above models are constant in time, the models are stationary sequential generation models. Non-stationary models use time-varying parameters.

A good practice of time series modeling requires a diagnostics check of the chosen type and form of the model. This is done by comparing the synthetic series to the observed sample and determining if the main statistics have been preserved. The synthetic series are considered to be valid alternative sequences to the sample series if the sample's statistics are preserved, thus enabling the use of the generated series for the purpose for which they were needed (Guimarães, 2005, p. 81). If M synthetic series are generated, the comparison of the sample series' statistic, θ, with the synthetic series' statistics, $\hat{\theta}^{(m)}$ ($m = 1, 2, ...M$), can be made by obtaining the mean $\left(\overline{\hat{\theta}}\right)$ and standard deviation $\left(s_{\hat{\theta}}\right)$ of $\hat{\theta}$ and by establishing a confidence interval of $(1 - \alpha)$ for θ,

$$\left]\overline{\hat{\theta}} - z_{1-\alpha/2} \; s_{\hat{\theta}}; \overline{\hat{\theta}} + z_{1-\alpha/2} \; s_{\hat{\theta}}\right[\tag{2.18}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the Normal standard distribution, α being the significance level. If the statistic θ of the sample series belongs to that interval, it is considered to be preserved in the synthetic series.

For the design of the storage capacity of a reservoir, as was the objective of this research, monthly synthetic flow series were utilized. The generation of the monthly streamflow series can take a two-level approach, comprehending a generation model for the annual synthetic series, and a disaggregation model to obtain the corresponding monthly flow series.

2.4.2 Generation models

The selection of an appropriate model for generating synthetic time series depends on the characteristics of the sample.

If the sample series is built upon a variable independent in time, the generation of synthetic series can be carried out with a non-deterministic probabilistic model, such as the simple Monte Carlo sampling procedure. This procedure consists of a random sampling of a probability density function as input of the mathematical model to produce a number of possible outcomes (Ribeiro, 1996, p. 72, Guimarães, 2005, p. 65).

In the case that the sample series is dependent in time, the model used should be a stochastic model. The most adopted stochastic models are:

- Autoregressive models, AR(p), and periodic autorregressive models, PAR(p), of order p, are short memory models capable of preserving low-order moments of the sample series (Guimarães, 2005, p. 30);

- Autorregressive-moving average models, ARMA(p,q), and periodic autorregressive-moving average models, PARMA(p,q), add a moving average component of order q to the AR models. These are long memory models suitable for series with high order correlations.

AR(p) and ARMA(p,q) models are applicable to annual streamflow series, while PAR(p) and PARMA(p,q) models are more suited for monthly streamflow series. An exhaustive study of these and other stochastic models can me made by consulting the textbooks by Salas *et al.* (1980) and Box *et al.* (1994).

2.4.3 Disaggregation models

The models presented in 2.4.2 have been designed to preserve the statistical characteristics of each sample. A model applied to a sample of annual flows is directed to preserve the statistics of the annual flows. Likewise, a model applied to a sample of monthly flows is directed to preserve the statistics of the monthly flows. However, the foregoing models cannot preserve the statistics simultaneously at both time levels (year and month). For instance, if the generated monthly flows are aggregated to obtain the corresponding annual flows, there is no assurance that the historical annual statistics would be preserved. In fact, they are seldom preserved (Santos, 1983, p. 2).

Disaggregation techniques are aimed to preserve the statistical properties of generated time series at more than one level. These techniques consist of a combination of a model for generating values of a variable at a given time level, and a model for disaggregating these values into a lower level. More often, these levels have a temporal nature, the upper level being, for example, the year and the lower level the months.

There are a number of disaggregation models available. The following is a brief reference of some of these models.

- The Valencia and Shaake method, proposed by Valencia & Schaake Jr (1972) (Silva, 1989, p. 13) was the first formal disaggregation model to be developed and has become widely accepted in stochastic hydrology. The basis of the method is the simultaneous generation of the monthly or seasonal flows of a given year, by disaggregation of the respective annual flow. This model has the attributes of preserving the annual and seasonal statistics and the additive property, that is, the aggregation of the generated seasonal flows result in the generated annual flows. This model has, however, some shortcomings, namely, its large number of parameters.

- The step disaggregation model proposed by Santos (1983) is based on a sequential disaggregation of annual flows, preserving the additive quality of the monthly flows. This model has the advantage of using a minimum number of parameters to preserve the covariance structure of the seasonal series.

- The method of fragments, proposed by Svanidze in 1961 (Svanidze, 1980), was very popular within the former USSR, but did not have similar popularity elsewhere (Santos, 1983, p. 19). This method considers the within-year distribution of monthly flows and assumes that two years with a similar annual flow value will have a similar within-year distribution of the monthly flows. The method of fragments is capable of generating monthly flow series while preserving monthly and annual statistics and the additive property.

3 Streamflow data

The data set used in research consists of continuous monthly and equivalent annual streamflow records from 53 gauging stations, geographically spread over Mainland Portugal. The locations of the stations are shown in Fig. 3.1, on page 29. The 53 gauging stations were selected from a list of 54 stations used in previous studies carried out by Portela & Quintela (2005a,b, 2006a,b), however the length of some of the streamflow samples has been extended for the current data set. Although *Moinho do Bravo* has a long streamflow sample (between 1934/35 and 1989/90), there is a considerably long gap in the records (between 1959/60 and 1975/76). Consequently the streamflows preceding and the ones succeeding this gap were considered as two separate streamflow samples. As a result, the total number of streamflow samples used was 54. The hydrological year (starting on October 1$^{\text{st}}$) was adopted as the annual time-step. Table 3.1 summarizes the names of the gauging station, the location, the period with records and the catchment area for each sample. The Code column lists the reference of the gauging stations in the SNIRH online database (*Sistema Nacional de Informação de Recursos Hídricos* - http://snirh.pt).

The average length of the samples is close to 35 years, while the shortest and the longest samples respectively have the lengths of 11 (*Pte. da Ota*) and 73 years (*Pte. Juncais*). The catchment area varies from 4 to 3718 km^2 (*Louçainha* and *Castanheiro*, respectively).

The coefficient of variability C_V of the annual flows was computed for each sample and a regression analysis was made to determine the relationship between C_V and $\overline{H}$, adopting the form given by Equation (2.1). The graphical representation of this analysis as well as the equation that expresses the relationship, including the correlation coefficient, cc, are shown in Fig. 3.2, on page 30. The figure includes the comparison of the curve with the ones previously established by Portela & Quintela (2005a,b, 2006a) based on a data set of 54 gauging stations (of which, 53 stations were used in the current research, though with slightly different recording periods), and by Portela & Quintela (2000, 2001, 2002a,b), based on 24 gauging stations. The latter curves were presented in Table 2.1, on page 6.

Table 3.1: Streamflow samples: gauging stations, period of records of continuous annual and monthly data, and catchment area.

| Sample No. | Gauging station | | | Period of records | Area (km^2) |
	Code	Name	Catchment/Watercourse	(number of years)	
1	26J/01	Albernoa	Guadiana/Terges	1970/71 - 1994/95 (25)	177
2	27J/01	Monte da Ponte	Guadiana/Cobres	1959/60 - 1993/94 (35)	701
3	24H/01	S.Domingos	Sado/Rib^a Algalé	1934/35 - 1958/59 (25)	59
4	24L/01	Amieira	Guadiana/Degebe	1944/45 - 1990/91 (47)	1474
5	24H/03	Torrão do Alentejo	Sado/Xarrama	1961/62 - 1989/90 (29)	465
6	19D/04	Pte. da Ota	Tejo/Rib^a de Ota	1979/80 - 1989/90 (11)	56
7	19C/02	Pte. Barnabé	Tejo/Alenquer	1979/80 - 1991/92 (13)	114
8	25G/02	Moinho do Bravo (1)	Sado/Rib^a Corona	1934/35 - 1958/59 (25)	218
9	24I/01	Odivelas	Sado/Rib^a de Odivelas	1934/35 - 1969/70 (36)	431
10	25G/02	Moinho do Bravo (2)	Sado/Rib^a Corona	1976/77 - 1989/90 (14)	218
11	18L/01	Couto de Andreiros	Tejo/Rib^a de Seda	1974/75 - 1992/93 (19)	244
12	12E/01	Pte. Azenha Nova	Mondego/Rib^a de Foja	1975/76 - 1987/88 (13)	51
13	30F/02	Vidigal	Algarve/Rib^a do Farelo	1938/39 - 1963/64 (26)	19
14	19M/01	Monforte	Tejo/Rib^a de Avis	1955/56 - 1988/89 (34)	136
15	31K/03	Bodega	Algarve/Rib^a de Alportel	1952/53 - 1988/89 (37)	132
16	06O/03	Q. das Laranjeiras	Douro/Sabor	1942/43 - 2005/06 (64)	3464
17	29L/01	Monte dos Fortes	Guadiana/Rib^a Odeleite	1961/62 - 1992/93 (32)	288
18	23I/01	Flor da Rosa	Sado/Xarrama	1934/35 - 1965/66 (32)	278
19	28L/02	Vascão	Guadiana/Rib^a Vascão	1960/61 - 1982/83 (23)	403
20	08O/02	Cidadelhe	Douro/Côa	1955/56 - 2003/04 (49)	1685
21	30G/01	Mte. dos Pachecos	Algarve/Rib^a de Odelouca	1961/62 - 1982/83 (22)	386
22	18E/01	Pte. Freiria	Tejo/Maior	1976/77 - 1989/90 (14)	184
23	10P/01	Castelo Bom	Douro/Côa	1957/58 - 2003/04 (47)	897
24	13F/02	Pte. Casével	Mondego/Ega	1975/76 - 1989/90 (15)	146
25	21C/01	Pte. Pinhal	Tejo/Rib^a de Loures	1977/78 - 1988/89 (12)	79
26	06M/01	Castanheiro	Douro/Tua	1958/89 - 2003/04 (46)	3718
27	05M/01	Murça	Douro/Tinhela	1970/71 - 2003/04 (34)	265
28	11I/06	Pte. Tábua	Mondego/Mondego	1937/38 - 1978/79 (42)	1552
29	10K/01	Pte. Sta Clara-Dão	Mondego/Dão	1921/22 - 1988/89 (68)	177
30	10L/01	Pte. Juncais	Mondego/Mondego	1918/19 - 1990/91 (73)	604
31	10L/01	Caldas S. Gemil	Mondego/Dão	1952/53 - 1989/90 (38)	617
32	04J/05	Pte. Cavez	Douro/Tâmega	1957/58 - 2005/06 (49)	1951
33	08L/01	Quinta do Rape	Douro/Távora	1976/77 - 2003/04 (28)	170
34	03N/01	Rebordelo	Douro/Rabaçal	1955/56 - 2002/03 (48)	857
35	12H/03	Pte. Mucela	Mondego/Alva	1938/39 - 1989/90 (52)	666
36	09G/01	Pte. Vale Maior	Vouga/Caima	1935/36 - 1972/73 (38)	188
37	11H/03	Açude Saimilo	Mondego/Dão	1939/40 - 1974/75 (36)	1371
38	06I/02	Pte. Canavezes	Douro/Tâmega	1955/56 - 1986/87 (32)	3180
39	03K/01	Vale Giestoso	Douro/Beça	1957/58 - 2005/06 (49)	77
40	10M/03	Videmonte	Mondego/Mondego	1975/76 - 1996/97 (22)	121
41	08J/01	Castro Daire	Douro/Pavia	1945/46 - 2003/04 (59)	291
42	07I/04	Cabriz	Douro/Rib^a S. Paio	1966/67 - 1996/97 (31)	17
43	09I/02	Pte. Vouzela	Vouga/Vouga	1956/57 - 1973/74 (18)	649
44	03P/01	Vinhais - Qt.Ranca	Douro/Tuela	1955/56 - 1996/97 (42)	455
45	04J/04	Cunhas	Douro/Beça	1949/50 - 2005/06 (57)	338
46	06K/01	Ermida - Corgo	Douro/Corgo	1956/57 - 2005/06 (50)	291
47	11M/01	Pai Diz	Mondego/Mondego	1973/74 - 1995/96 (23)	50
48	05K/01	S. Marta do Alvão	Douro/Louredo	1955/56 - 2005/06 (51)	52
49	09H/01	Pedre Ribeiradio	Vouga/Vouga	1962/63 - 1979/80 (18)	928
50	13H/03	Louçainha	Mondego/Simonte	1959/60 - 1983/84 (25)	4
51	08H/02	Fragas da Torre	Douro/Pavia	1945/46 - 2005/06 (61)	660
52	17F/02	Pte. Nova	Tejo/Almonda	1976/77 - 1989/90 (14)	102
53	11L/01	Manteigas	Tejo/Zêzere	1948/49 - 1995/96 (48)	28
54	03H/04	Covas	Cávado/Homem	1955/56 - 1973/74 (19)	116
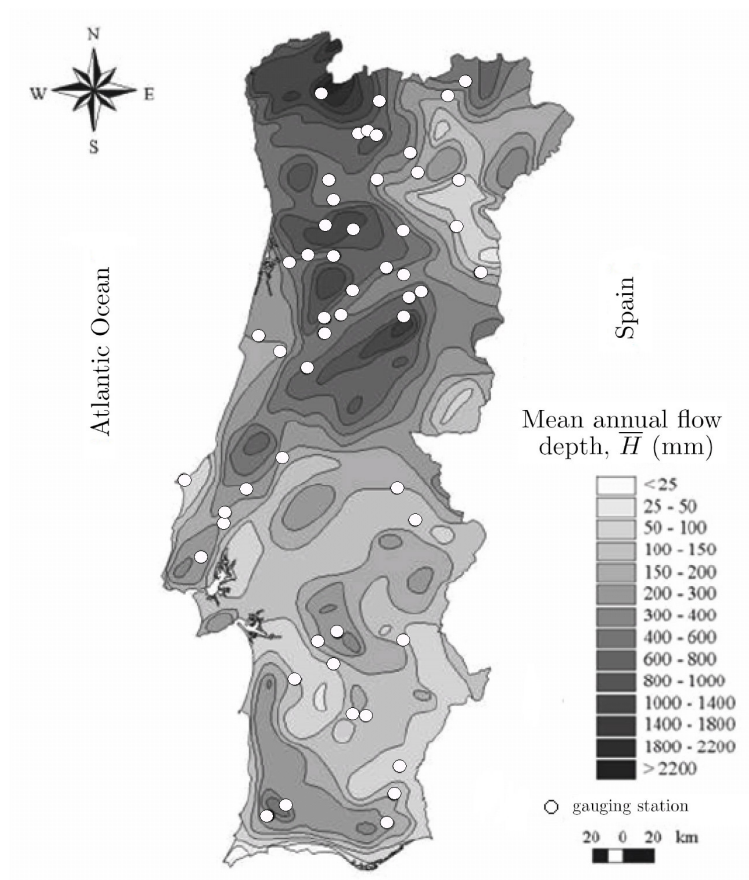
Figure 3.1: Mainland Portugal. Location of the 53 stream gauging stations. Contour map of the mean annual flow depth, after Portela & Quintela (2002b).

The curves represented in Figure 3.2 do not differ significantly from each other. Hence the hypothesis that the mean annual flow depth $\overline{H}$ is closely related to the temporal variability of the flow regime, in accordance with the various studies carried out by Portela & Quintela (2000, 2001, 2002a,b, 2005a,b), is validated for the current data set. The values of the mean, standard deviation, coefficient of variation, and skewness coefficient of annual flow depths ($\overline{H}$, s_H, C_V, and g_H, respectively), as well as the values of the mean and standard deviation of the annual flow volumes ($\overline{X}$ and s_X, respectively), of each streamflow sample are presented in Table 3.2, on page 31.

29
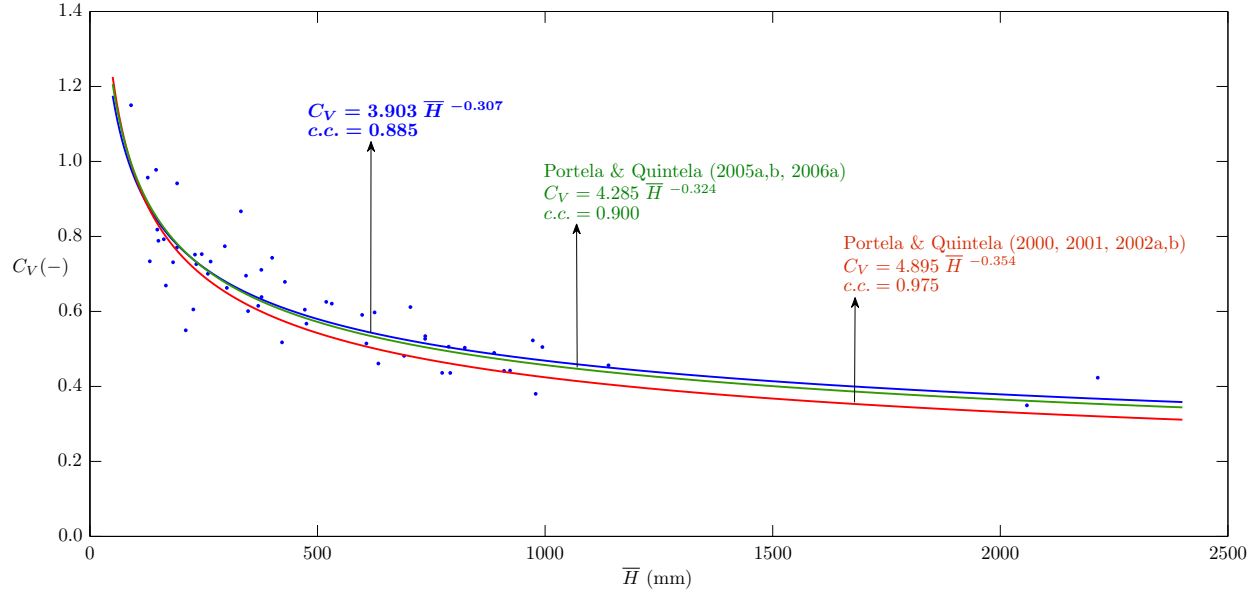
Figure 3.2: Relationship between the coefficients of variation of annual flows, C_V, and the mean annual flow depth, $\overline{H}$. Results based on the current data set (blue curve and the corresponding dots), and previous results (green and red curves).

According to Quintela (1967) (Silva, 1989, p. 51), the annual streamflow series of Mainland Portugal river is an independent time series if the hydrological year (starting on October 1st) is adopted as a time-step, provided that an analysis of the correlation structure of the sample does not invalidate this assumption. The correlation structure of a time-series may be investigated by computing the lag-k autocorrelation coefficient r_k, which was computed, for each streamflow sample, using Equation (A.6), presented in Appendix A, defined by Salas *et al.* (1980, p. 38). If $r_k = 0$ for $k \neq 0$, then the time-series is independent in time. However, due to the variability of the annual streamflows, r_k is expected not to be equal to zero but to fluctuate around zero, resulting in the need to use statistical criteria to decide if the r_k values differ significantly from zero. Anderson (1942) defined the probability limits for the serial correlation of an independent series, for a confidence level of 95% (Salas *et al.*, 1980, p. 49):

$$r_k(95\%) = \frac{-1 \pm 1.96\sqrt{N - k - 1}}{N - k} \tag{3.1}$$

where N is the sample length and k is the lag.

Table 3.3 contains the values of the lag-one and lag-two autocorrelation coefficients, and the corresponding Anderson (1942) probability limits.

Table 3.2: Streamflow samples: Main statistical characteristics of the annual flows.

Sample No.	Name	$\overline{H}$ (mm)	s_H (mm)	$\overline{X}$ (dam^3)	s_X (dam^3)	$g_{H,X}$ (-)	C_V (-)
1	Albernoa	90.4	104.0	16002	18402	1.869	1.150
2	Monte da Ponte	127.1	121.6	89087	85245	1.068	0.957
3	S.Domingos	131.5	96.5	7761	5694	0.324	0.734
4	Amieira	145.3	142.1	214233	209450	0.995	0.978
5	Torrão do Alentejo	148.0	121.1	68809	56304	0.572	0.818
6	Pte. da Ota	150.5	118.7	8428	6645	0.709	0.788
7	Pte. Barnabé	162.4	128.7	18513	14672	0.931	0.793
8	Moinho do Bravo (1)	166.9	111.7	36389	24346	0.158	0.669
9	Odivelas	182.7	133.6	78752	57575	0.352	0.731
10	Moinho do Bravo (2)	190.9	147.0	41624	32048	0.110	0.770
11	Couto de Andreiros	191.6	180.4	46749	44017	0.802	0.942
12	Pte. Azenha Nova	210.6	115.7	10739	5903	0.364	0.550
13	Vidigal	227.2	137.6	4318	2614	0.275	0.605
14	Monforte	230.8	173.5	31387	23590	0.645	0.752
15	Bodega	233.7	169.7	30852	22401	0.617	0.726
16	Q. das Laranjeiras	245.7	185.0	851107	640723	1.087	0.753
17	Monte dos Fortes	259.1	181.5	74632	52283	0.692	0.701
18	Flor da Rosa	265.2	194.4	73727	54052	0.204	0.733
19	Vascão	296.4	229.4	119459	92449	0.674	0.774
20	Cidadelhe	300.9	199.4	506950	336010	0.386	0.663
21	Mte. dos Pachecos	331.7	287.6	128046	111015	0.919	0.867
22	Pte. Freiria	343.0	238.5	63111	43887	0.515	0.695
23	Castelo Bom	347.5	208.8	311726	187310	0.456	0.601
24	Pte. Casével	369.8	227.4	53988	33202	0.451	0.615
25	Pte. Pinhal	376.6	267.7	29749	21149	1.388	0.711
26	Castanheiro	377.1	240.7	1401978	895045	1.069	0.638
27	Murça	400.3	297.5	106080	78831	1.768	0.743
28	Pte. Tábua	422.0	218.4	654987	339001	0.543	0.518
29	Pte. Sta Clara-Dão	428.4	290.8	75824	51469	1.668	0.679
30	Pte. Juncais	472.1	285.5	285150	172424	1.287	0.605
31	Caldas S. Gemil	475.8	270.0	293547	166582	0.663	0.567
32	Pte. Cavez	519.2	324.7	1012890	633543	1.198	0.625
33	Quinta do Rape	531.4	329.9	90343	56082	0.388	0.621
34	Rebordelo	598.1	353.3	512542	302796	1.030	0.591
35	Pte. Mucela	607.4	312.5	404554	208138	0.360	0.514
36	Pte. Vale Maior	625.4	373.6	117572	70235	1.220	0.597
37	Açude Saimilo	633.8	292.2	868884	400606	0.643	0.461
38	Pte. Canavezes	690.3	332.5	2195048	1057224	0.774	0.482
39	Vale Giestoso	704.1	430.6	54215	33159	1.152	0.612
40	Videmonte	736.4	388.0	89107	46952	0.055	0.527
41	Castro Daire	736.8	393.5	214396	114522	0.891	0.534
42	Cabriz	773.8	337.5	13155	5738	0.677	0.436
43	Pte. Vouzela	787.9	398.6	511352	258679	1.144	0.506
44	Vinhais - Qt.Ranca	791.6	345.1	360177	157043	0.568	0.436
45	Cunhas	823.5	414.4	278348	140064	0.711	0.503
46	Ermida - Corgo	888.1	434.7	258443	126489	0.823	0.489
47	Pai Diz	909.7	401.6	45485	20079	-0.007	0.441
48	S. Marta do Alvão	923.0	408.1	47996	21222	0.884	0.442
49	Pedre Ribeiradio	973.1	508.6	903007	471998	0.567	0.523
50	Louçainha	979.3	372.4	3917	1490	0.267	0.380
51	Fragas da Torre	994.1	502.1	656081	331357	0.944	0.505
52	Pte. Nova	1139.3	519.8	116204	53017	-0.489	0.456
53	Manteigas	2058.5	719.5	57638	20145	0.899	0.350
54	Covas	2214.1	937.2	256832	108713	0.132	0.423

Table 3.3: Samples of annual streamflows. Lag-one and lag-two serial correlation coefficients.

| Sample No. | Name | Lag-k serial correlation coefficients ($^\alpha$) | |
		lag-one, r_1	lag-two, r_2
1	*Albernoa*	0.0720 (-0.4333 - 0.3500)	0.0883 (-0.4432 - 0.3562)
2	*Monte da Ponte*	0.0897 (-0.3606 - 0.3017)	-0.1033 (-0.3663 - 0.3057)
3	*S.Domingos*	-0.0958 (-0.4333 - 0.3500)	-0.3942 (-0.4432 - 0.3562)
4	*Amieira*	0.1386 (-0.3076 - 0.2641)	-0.1758 (-0.3111 - 0.2667)
5	*Torrão do Alentejo*	0.1378 (-0.3994 - 0.3280)	-0.0561 (-0.4072 - 0.3331)
6	*Pte. da Ota*	-0.1736 (-0.6880 - 0.4880)	-0.0419 (-0.7271 - 0.5049)
7	*Pte. Barnabé*	0.1695 (-0.6250 - 0.4584)	-0.0745 (-0.6544 - 0.4726)
8	*Moinho do Bravo (1)*	-0.1284 (-0.4333 - 0.3500)	-0.3573 (-0.4432 - 0.3562)
9	*Odivelas*	-0.0246 (-0.3551 - 0.2980)	-0.3860 (-0.3606 - 0.3017)
10	*Moinho do Bravo (2)*	0.1173 (-0.5992 - 0.4454)	0.0549 (-0.6250 - 0.4584)
11	*Couto de Andreiros*	0.2917 (-0.5045 - 0.3934)	-0.1894 (-0.5200 - 0.4024)
12	*Pte. Azenha Nova*	-0.1104 (-0.6250 - 0.4584)	-0.0820 (-0.6544 - 0.4726)
13	*Vidigal*	0.0418 (-0.4241 - 0.3441)	-0.3654 (-0.4333 - 0.3500)
14	*Monforte*	0.0479 (-0.3663 - 0.3057)	-0.0768 (-0.3723 - 0.3098)
15	*Bodega*	0.0431 (-0.3499 - 0.2943)	-0.3253 (-0.3551 - 0.2980)
16	*Q. das Laranjeiras*	-0.0540 (-0.2608 - 0.2291)	0.0190 (-0.2630 - 0.2308)
17	*Monte dos Fortes*	0.1809 (-0.3786 - 0.3140)	-0.1509 (-0.3852 - 0.3185)
18	*Flor da Rosa*	-0.0894 (-0.3786 - 0.3140)	-0.3140 (-0.3852 - 0.3185)
19	*Vascão*	0.3201 (-0.4537 - 0.3628)	-0.1256 (-0.4650 - 0.3698)
20	*Cidadelhe*	-0.0731 (-0.3008 - 0.2591)	0.1122 (-0.3041 - 0.2616)
21	*Mte. dos Pachecos*	0.3175 (-0.4650 - 0.3698)	-0.2390 (-0.4772 - 0.3772)
22	*Pte. Freiria*	0.2670 (-0.5992 - 0.4454)	0.2275 (-0.6250 - 0.4584)
23	*Castelo Bom*	-0.0919 (-0.3076 - 0.2641)	0.0929 (-0.3111 - 0.2667)
24	*Pte. Casével*	0.1141 (-0.5762 - 0.4333)	-0.0278 (-0.5992 - 0.4454)
25	*Pte. Pinhal*	0.1520 (-0.6544 - 0.4726)	-0.2658 (-0.6880 - 0.4880)
26	*Castanheiro*	-0.1164 (-0.3111 - 0.2667)	0.1188 (-0.3148 - 0.2694)
27	*Murça*	-0.1033 (-0.3663 - 0.3057)	0.0519 (-0.3723 - 0.3098)
28	*Pte. Tábua*	0.0297 (-0.3267 - 0.2780)	-0.0793 (-0.3310 - 0.2810)
29	*Pte. Sta Clara-Dão*	0.0197 (-0.2526 - 0.2227)	0.0690 (-0.2546 - 0.2243)
30	*Pte. Juncais*	-0.0055 (-0.2433 - 0.2155)	-0.0731 (-0.2450 - 0.2169)
31	*Caldas S. Gemil*	0.0646 (-0.3449 - 0.2908)	0.0239 (-0.3499 - 0.2943)
32	*Pte. Cavez*	-0.0133 (-0.3008 - 0.2591)	0.0177 (-0.3041 - 0.2616)
33	*Quinta do Rape*	0.0287 (-0.4072 - 0.3331)	-0.0564 (-0.4154 - 0.3385)
34	*Rebordelo*	-0.1034 (-0.3041 - 0.2616)	0.0696 (-0.3076 - 0.2641)
35	*Pte. Mucela*	0.2310 (-0.2914 - 0.2521)	-0.0930 (-0.2944 - 0.2544)
36	*Pte. Vale Maior*	0.2998 (-0.3449 - 0.2908)	0.0852 (-0.3499 - 0.2943)
37	*Açude Saimilo*	-0.0452 (-0.3551 - 0.2980)	-0.2056 (-0.3606 - 0.3017)
38	*Pte. Canavezes*	0.0147 (-0.3786 - 0.3140)	-0.1201 (-0.3852 - 0.3185)
39	*Vale Giestoso*	-0.0418 (-0.3008 - 0.2591)	0.0814 (-0.3041 - 0.2616)
40	*Videmonte*	-0.0143 (-0.4650 - 0.3698)	-0.1468 (-0.4772 - 0.3772)
41	*Castro Daire*	-0.1014 (-0.2724 - 0.2379)	0.0314 (-0.2749 - 0.2398)
42	*Cabriz*	-0.0125 (-0.3852 - 0.3185)	-0.0608 (-0.3921 - 0.3231)
43	*Pte. Vouzela*	-0.0204 (-0.5200 - 0.4024)	-0.0756 (-0.5369 - 0.4119)
44	*Vinhais - Qt.Ranca*	-0.1009 (-0.3267 - 0.2780)	0.0045 (-0.3310 - 0.2810)
45	*Cunhas*	-0.0385 (-0.2774 - 0.2417)	0.0981 (-0.2801 - 0.2437)
46	*Ermida - Corgo*	-0.0471 (-0.2975 - 0.2567)	0.0094 (-0.3008 - 0.2591)
47	*Pai Diz*	-0.0344 (-0.4537 - 0.3628)	-0.2009 (-0.4650 - 0.3698)
48	*S. Marta do Alvão*	-0.0928 (-0.2944 - 0.2544)	0.0063 (-0.2975 - 0.2567)
49	*Pedre Ribeiradio*	-0.1136 (-0.5200 - 0.4024)	0.0862 (-0.5369 - 0.4119)
50	*Louçainha*	0.0482 (-0.4333 - 0.3500)	-0.1096 (-0.4432 - 0.3562)
51	*Fragas da Torre*	-0.0510 (-0.2676 - 0.2343)	0.0049 (-0.2699 - 0.2360)
52	*Pte. Nova*	-0.0435 (-0.5992 - 0.4454)	-0.1072 (-0.6250 - 0.4584)
53	*Manteigas*	0.0648 (-0.3041 - 0.2616)	0.0275 (-0.3076 - 0.2641)
54	*Covas*	0.2572 (-0.5045 - 0.3934)	0.1956 (-0.5200 - 0.4024)

($^\alpha$) The values in parentheses represent the Anderson (1942) probability limits for a confidence level of 95%, as defined in Equation (3.1).

The annual streamflows in gauging stations *Pte. Vale Maior* and *Odivelas* (samples no. 36 and 27), respectively) present r_k values that are slightly outside of the 95% confidence interval (for $k = 1$ and $k = 2$, respectively). It was assumed that this was due to the samples' variability of the annual streamflows and it was admitted that the hypothesis of temporal independence of annual streamflows was valid for those two samples. The remaining streamflow samples all display r_k values within the corresponding 95% confidence intervals.

4 Methodology

4.1 Generation of synthetic monthly streamflow series

4.1.1 General considerations

The methodology applied to generate synthetic monthly streamflow series consisted of a two-level approach, comprehending a generation model for the annual synthetic series, and a disaggregation model to obtain the corresponding monthly flow series.

In Chapter 3, the temporal independence of the annual streamflows at the several gauging stations, while adopting the hydrological year, was confirmed (Table 3.2), hence the modeling of the annual flow series used a probabilistic model. The model is based on the random sampling of the probability density function of the log-Pearson III law, which is the Pearson III law applied to the logarithms of the random variable. Because this distribution can only take values between 0 and ∞, the generation of negative flows is naturally avoided. The log-Pearson III law has a flexible distribution, assuming a number of different shapes depending on the mean, variance and skewness of the sample, hence the treatment of skewed data is assured (Chow *et al.*, 1988, p. 375).

The disaggregation of the generated annual flows into monthly flows used the method of fragments, proposed by Svanidze in 1961 (Santos, 1983). A brief description of this model is presented in Section 2.4.3.

Sections 4.1.2 and 4.1.3 describe the generation model and the disaggregation model, respectively, used to generate annual and monthly streamflow series with the lengths equal to that of the samples.

Guimarães (2005, p. 175) indicates that the number of generated series, M, in reservoir storage estimation studies, should be 1200. This indication was adopted in the methodology proposed in this dissertation. Naturally, in order to obtain M synthetic series, the described methodology must be carried out M times for each sample.

Given that the mean length of the samples is 35 years, an average of 500 000 monthly flows were generated for each sample, which results in a total of nearly 30 million monthly and annual flows generated for the 54 samples.

4.1.2 Generation of annual streamflows

The model for generating annual flows consists of a random sampling of the log-Pearson III distribution, which is the Pearson III distribution applied to the logarithms of the flows, that is, if $\ln(X)$ follows the Pearson III distribution, then X follows the log-Pearson III distribution. The first step of the model is then to determine the series of logarithms of the annual flows:

$$W_i = \ln(X_i + c) \tag{4.1}$$

where W_i is the series of the natural logarithms of the annual flows, X_i. The constant c was added to the annual flows to avoid the occurrence of null flows to which the logarithmic transformation is not applicable. This constant adopted a fixed value of $c = 0.0001$ in order to minimize the effect that its addition has on the characteristics of the samples.

In fact, rather than annual flows, the statistical model generates logarithms of annual flows. The method of moments was used to determine the statistical parameters of each sample of logarithms of the annual flows, namely the mean, $\overline{W}$, the standard deviation, s_W, and the skewness coefficient, g_W, according to Equations (A.1), (A.2) and (A.4) in Appendix A (the latter two are unbiased estimators).

The synthetic series of N logarithms of annual flows $\hat{W}$ are generated according to the following expression

$$\hat{W}_i = \overline{W} + \zeta_i s_W \tag{4.2}$$

where ζ_i is the probability factor of the Pearson III distribution which may be obtained by applying the Wilson-Hilferty transformation (Arsénio, 2003, p. 50; Naghettini & Pinto, 2007, p. 321) to the z_i random variables. The Wilson-Hilferty transformation is given by the following equation, where $i = 1, ..., N$:

$$\zeta_i = \left\{ \left[\frac{g_W}{6} \left(z_i - \frac{g_W}{6} \right) + 1 \right]^3 - 1 \right\} \frac{2}{g_W} \tag{4.3}$$

In the previous equation, z_i designates the ith normally distributed random variable z, with mean zero and variance one:

$$z \sim \mathcal{N}(0,1) \tag{4.4}$$

The values of z_i were obtained using the built-in function `RANDN` in Matlab, which uses a PRNG (PseudoRandom Number Generator) - Marsaglia's ziggurat algorithm - which generates numbers drawn from a normal distribution with mean zero and standard deviation one (the period is approximately 2^{64}) (MathWorks, 2008). As an example, Figure 4.1 shows fifty numbers generated by this algorithm. The sequence of numbers drawn by a PRNG is determined by the initial state of the generator. Setting the generator to the same fixed state allows computations to be repeated; while setting the generator to different states leads to unique computations. The state of the PRNG does not change any statistical properties of the drawn numbers. The initial state of the PRNG can be set using a seed state. To insure that the PRNG is starting from a different

state on each run, a different seed was assigned for each streamflow sample, and the PRNG was reset to that seed at the beginning of the generation model for said sample. It is important not to reset the state of the PRNG after each series is generated, otherwise the M series would be identical. Table 4.1 shows the seed assigned to each streamflow sample.
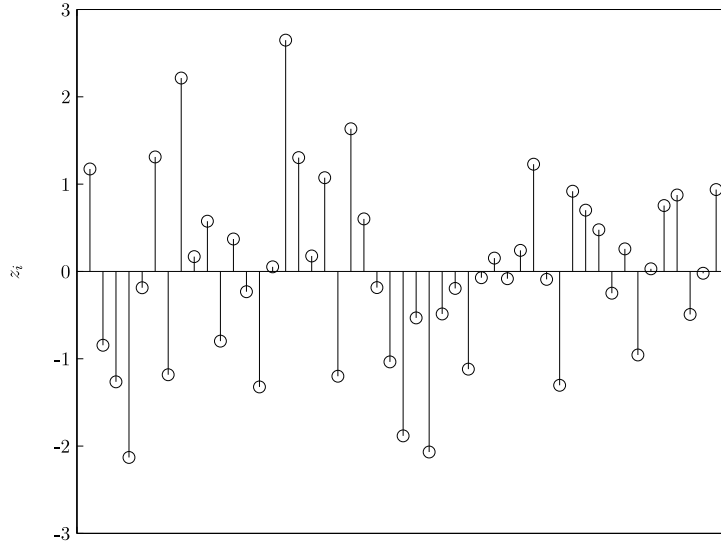
Figure 4.1: Fifty random Gaussian numbers, with mean zero and variance one, generated using Marsaglia's Ziggurat algorithm.

Table 4.1: Pseudorandom number generator seed numbers assigned to each sample.

Sample No.	Name	Seed No.	Sample No.	Name	Seed No.
1	*Albernoa*	535646	28	*Pte. Tábua*	796525
2	*Monte da Ponte*	379587	29	*Pte. Sta Clara-Dão*	517609
3	*S.Domingos*	293555	30	*Pte. Juncais*	520968
4	*Amieira*	359503	31	*Caldas S. Gemil*	762340
5	*Torrão do Alentejo*	606021	32	*Pte. Cavez*	99434
6	*Pte. da Ota*	257589	33	*Quinta do Rape*	904016
7	*Pte. Barnabé*	351194	34	*Rebordelo*	145591
8	*Moinho do Bravo (1)*	178293	35	*Pte. Mucela*	223848
9	*Odivelas*	863914	36	*Pte. Vale Maior*	687465
10	*Moinho do Bravo (2)*	226977	37	*Açude Saimilo*	836139
11	*Couto de Andreiros*	590645	38	*Pte. Canavezes*	813847
12	*Pte. Azenha Nova*	710596	39	*Vale Giestoso*	36539
13	*Vidigal*	145024	40	*Videmonte*	849561
14	*Monforte*	342105	41	*Castro Daire*	281855
15	*Bodega*	297686	42	*Cabriz*	142423
16	*Q. das Laranjeiras*	147535	43	*Pte. Vouzela*	701025
17	*Monte dos Fortes*	913973	44	*Vinhais - Qt.Ranca*	389499
18	*Flor da Rosa*	967751	45	*Cunhas*	661656
19	*Vascão*	250960	46	*Ermida - Corgo*	927758
20	*Cidadelhe*	459484	47	*Pai Diz*	114505
21	*Mte. dos Pachecos*	633042	48	*S. Marta do Alvão*	726129
22	*Pte. Freiria*	298262	49	*Pedre Ribeiradio*	901758
23	*Castelo Bom*	144756	50	*Louçainha*	279357
24	*Pte. Casével*	581073	51	*Fragas da Torre*	163983
25	*Pte. Pinhal*	717723	52	*Pte. Nova*	813399
26	*Castanheiro*	342669	53	*Manteigas*	451579
27	*Murça*	994234	54	*Covas*	344959

It should be noted that, when the skewness is zero, ζ is equal to z, that is, the log-Pearson III distribution takes the shape of the Log Normal law. This characteristic of the log-Pearson III distribution also insures the treatment of non-skewed data.

By inversion of the logarithmic transformation, and subtraction of the constant c, the synthetic series of annual flows are obtained:

$$\hat{X}_i = e^{\hat{W}_i} - c \tag{4.5}$$

The model is carried out M times until there are M synthetic annual flow series are generated: $\hat{X}_i^{(m)}$, with $m = 1, 2, ..., M$, M being fixed on 1200, as previously justified.

4.1.3   Monthly disaggregation of streamflows

The method of fragments considers that the within-the-year distribution of streamflows is identical in years with close values of annual flow volumes. In this method, for a given year k the observed monthly flows are divided by the corresponding annual flow volume, X_k, the resulting set of twelve standardized monthly flows constitutes the *fragment* pertaining to the year k, ϕ_k:

$$\phi_k = \frac{X_{k,j}}{X_k} = \left[\begin{array}{ccccc} \dfrac{X_{k,1}}{X_k} & \dfrac{X_{k,2}}{X_k} & \cdots & \dfrac{X_{k,11}}{X_k} & \dfrac{X_{k,12}}{X_k} \end{array} \right] \tag{4.6}$$

where $X_{k,j}$ are the monthly flows ($j = 1, 2, ..., 12$).

The application of the method of fragments requires that the fragments are constituted and assembled into classes beforehand. For the constitution of the array of fragments, $[\phi]$, the annual flows must previously be ranked from smallest to largest:

$$[\phi] = \left[\begin{array}{c} \phi_1 \\ \vdots \\ \phi_k \\ \vdots \\ \phi_N \end{array} \right] = \left[\begin{array}{ccc} \dfrac{X_{1,1}^*}{X_1^*} & \cdots & \dfrac{X_{1,12}^*}{X_1^*} \\ \vdots & \ddots & \vdots \\ \dfrac{X_{k,1}^*}{X_k^*} & \cdots & \dfrac{X_{k,12}^*}{X_k^*} \\ \vdots & \ddots & \vdots \\ \dfrac{X_{N,1}^*}{X_N^*} & \cdots & \dfrac{X_{N,12}^*}{X_N^*} \end{array} \right] \tag{4.7}$$

where X_k^* represents the k$^{\text{th}}$ annual flow such that $X_{k-1}^* \leq X_k^* \leq X_{k+1}^*$, and $X_{k,j}^*$ the monthly flows of that year.

In Figure 4.2, a typical fragment for Mainland Portugal is presented as an example. The sum of its values is, evidently, equal to one.

The classes of fragments are defined by the clustering of annual flows with similar values, and they may or may not have a constant amplitude. In the state of the art involving the method of fragments, there is no

known rule for the number and size of classes to adopt, which constitutes a source of ambiguity concerning the utilization of the method.

$$\phi_k = X_{k,j}/X_k$$
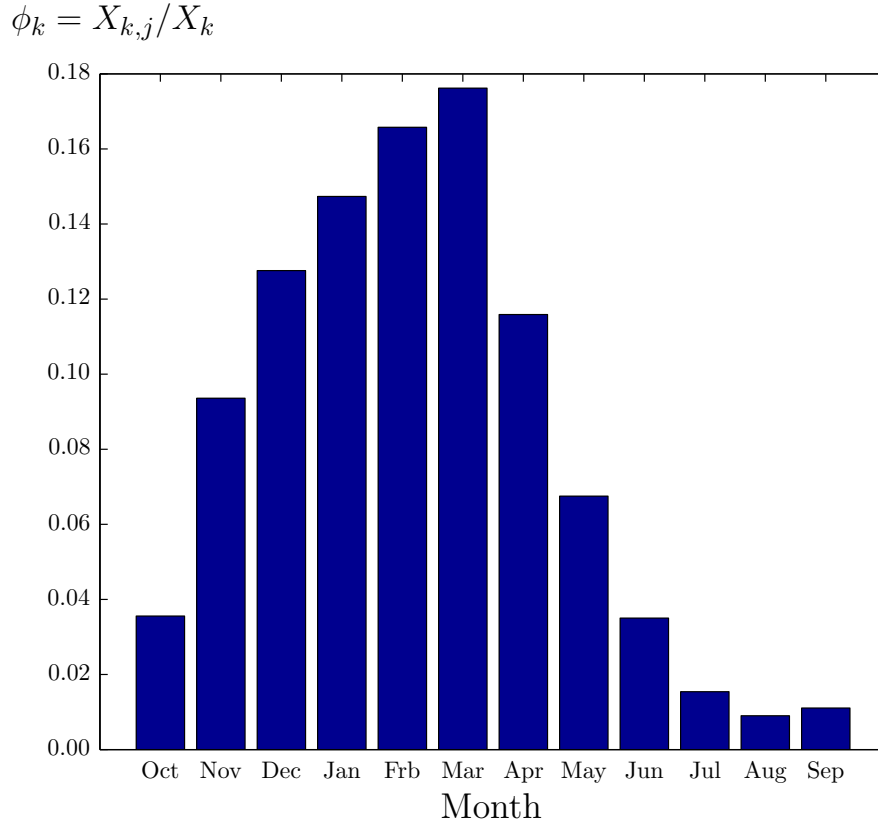
Figure 4.2: Example of a fragment for Mainland Portugal

Arsénio (2003) implemented the method of fragment to generate synthetic series of daily flows, having performed a sensitivity analysis involving different criteria for the definition of the classes of fragments. As a result Arsénio (2003, p. 84) suggested that the definition of the classes of fragments be made by trial and error until an assembly of fragments is found that leads to a more effective preservation of the statistics of the samples. However, that suggestion was not followed in the procedure developed for this dissertation, not only because it would require an individual analysis of each one of the more than fifty streamflow samples (which amounts to a very time-consuming process), but also because what was intended in the research carried out in this dissertation was to apply the same procedure to all the streamflow samples, and to make a global evaluation of said procedure.

A number of alternate approaches for defining the classes of fragments was devised and compared, in terms of robustness regarding the preservation of the statistical parameters of the samples and its ability to be automated, the one that proved best considers the classes of fragments defined as probability intervals. The

procedure applied is as follows:

1. Consideration of a series of nine equally spaced non-exceedance probabilities of the annual streamflow, F_m $(m = 1, ..., 9)$, with an increment of 10%, that is: $F_1 = 10\%$; $F_2 = 20\%$; ...;$F_9 = 90\%$.

2. Estimation of the annual flows, $\breve{X}_m$, pertaining to the aforementioned non-exceedance probabilities, by inverting the probability distribution function of the log-Pearson III law, attending to the statistical characteristics of each sample - Equations (4.2) and (4.3) - followed by the inversion of the logarithmic transformation. This results in ten equally probable classes with limits represented by $[0, \breve{X}_1[, [\breve{X}_1, \breve{X}_2[, [\breve{X}_2, \breve{X}_3[, ..., [\breve{X}_8, \breve{X}_9[, [\breve{X}_9, +\infty[$.

3. Distribution of the fragments among the successive classes, while simultaneously verifying if there are any empty classes. In case each class has at least one fragment, the procedure for the definition of the classes is complete and one can proceed to the monthly disaggregation model, otherwise one needs to take the next additional step.

4. Redefinition of empty classes:

   - If the first class is empty, it is included in the next class, by changing its upper limit to the value of $[0, \breve{X}_2[$. If it remains empty, it is included it in the next class and so on and so forth.

   - If the last class is empty, it is included in the previous class, by changing the lower limit to the value of $[\breve{X}_8, +\infty[$. If it remains empty, it is included in the previous class and so on and so forth.

   - If an intermediate class, $[\breve{X}_m, \breve{X}_{m+1}[$, is empty, half of the probability step that the defines this class is attributed to the two bordering classes. This results in that the three original classes $[\breve{X}_{m-1}, \breve{X}_m[, [\breve{X}_m, \breve{X}_{m+1}[, [\breve{X}_{m+1}, \breve{X}_{m+2}[$ are substituted by the following two new ones $[\breve{X}_{m-1}, \breve{X}_\partial[, [\breve{X}_\partial, \breve{X}_{m+1}[$ where $\breve{X}_\partial$ is the estimate of the annual flow with a non-exceedance probability that is the mean of the probability limits of the eliminated class. If the class remains empty, the process is repeated by attributing half of the probability interval to the bordering classes and consequently redefining these bordering classes' limits.

After the classes of fragments are defined, and having previously generated the synthetic series of annual flows, the method for generating the monthly flows proceeds with the identification, for each year i of the synthetic annual flow series, $\hat{X}_i$, of the class to which that annual flow belongs - class $m + 1$ such that $\hat{X}_i \in \left[\breve{X}_m, \breve{X}_{m+1}\right[$.

The next step is to select the fragment ϕ_i to be used to disaggregate the annual flow $\hat{X}_i$. If the class to which the previous step refers has only one fragment, said fragment is selected. Otherwise, that is, if the class has two or more fragments, the fragment is randomly selected, in accordance with the following procedure, provided that those fragments are sorted from smallest to largest value of the corresponding annual flows:

i A single random number ξ between zero and one is generated using the Matlab built-in function RAND, which uses the PRNG Mersenne Twister algorithm by Nishimura and Matsumoto, and draws numbers from a uniform distribution (MathWorks, 2008). The initial set of this PNGR is reset at the beginning of the generation model for each streamflow sample, to the seed numbers shown in Table 4.1.

ii The selected fragment number, that is, the number of order of the fragment in the class, is the integer part (floor) of the following operations: $\xi \ n_{frags} + 1$, where n_{frags} designates the number of fragments in the class.

The random selection of a fragments from a class with more than one fragment is done without replacement until the class runs out of fragments. When this happens, the class is refilled with the respective fragments.

Each generated annual flow, $\hat{X}_i$, is disaggregated into generated monthly flows in accordance with the following system:

$$\begin{bmatrix} \hat{X}_{i,1} & \hat{X}_{i,2} & \cdots & \hat{X}_{i,11} & \hat{X}_{i,12} \end{bmatrix} = \begin{bmatrix} \dfrac{X_{i,1}}{X_i} & \dfrac{X_{i,2}}{X_i} & \cdots & \dfrac{X_{i,11}}{X_i} & \dfrac{X_{i,12}}{X_i} \end{bmatrix} \hat{X}_i \tag{4.8}$$

which, for the purpose of simplifying the notation, will also be designated by:

$$\hat{X}_{i,j} = \phi_i \hat{X}_i \tag{4.9}$$

where $i = 1, 2, ..., N$ and $j = 1, 2, ..., 12$.

The disaggregation procedure is repeated until M N-year synthetic streamflow series, with an annual and monthly time-step, are generated:

$$\hat{X}_{i,j}^{(m)} = \phi_i^{(m)} \hat{X}_i^{(m)} \tag{4.10}$$

where, as before, the indices i and j refer to the year and the month, respectively; $\hat{X}_i^{(m)}$ represents the m$^{\text{th}}$ synthetic series of annual flows; $\hat{X}_{i,j}^{(m)}$, the corresponding synthetic series of monthly flows; and ϕ_i, the fragment used to disaggregate $\hat{X}_i^{(m)}$, being, in all cases, $m = 1, 2, ..., M$. As aforementioned, the adopted total number, M, of synthetic monthly series of N years was 1200, as indicated by (Guimarães, 2005, p. 175).

The method of fragments has a limitation that should be mentioned: when one wants to disaggregate a generated annual flow volume that is smaller than the smallest volume of the sample, or larger than the largest volume of the sample, the available fragments are only the ones on the first or the last class, respectively. This limitation may reduce the variability of the within-the-year distribution of the synthetic monthly flow series, as pointed out by (Guimarães, 2005, p. 46).

4.1.4 Assessment of the quality of the generated series

The quality of generation and disaggregation models was evaluated by comparing the synthetic series with the samples in terms of the main historical statistics, as described in Section 2.4.1. The comparison utilized

confidence levels, computed by Equation (2.18).

At the annual level the model generates logarithms of the random variable (annual flow volumes), hence, the statistics to be used when appraising the quality of the model refer to the logarithms, namely the means, standard deviations and skewness coefficients of the logarithms of annual flows. However, an additional analysis was carried out by comparing the statistics of the random variable itself (annual streamflows, in this case).

At the monthly level, the analyzed statistics were again the means, standard deviations and skewness coefficients of the monthly flows in each month of the hydrologic year.

At both temporal levels, a confidence level of $1 - \alpha = 95\%$ was adopted for the confidence intervals.

4.2   Reservoir storage-yield analyses

The estimation of the storage volumes used the simulation algorithm described in Section 2.3.2 which is based on the application of the mass balance equation (Equation 2.8) to compute the reservoir storage capacity. This method is suited to make a behaviour analysis of the performance of a reservoir, that is, one can allow the water supply to be less than the demanded volume and apply a number of different criteria to analyze the performance of the reservoir in meeting such demand.

The algorithm applies Equation (2.8) to simulate the state of the storage based on a continuous streamflow series combined with a known or assumed demand, precipitation over the surface of the reservoir and water loss.

The application of this method aims to obtain an estimate of the reservoir storage capacity. This has to be done iteratively: first an initial estimate of the reservoir capacity, C, is assumed and the initial condition of the reservoir (initial water content) is decided; a simulation using the time series data is accomplished and the performance of the reservoir is computed; if the performance is unsatisfactory, then a new estimate of the capacity is made and the simulation is repeated; iterations are then carried out until the estimated capacities fulfill the desired level of performance.

The performance criterion considered in the procedure was the time-based reliability R_T (Equation 2.10), in part because it is the most common criterion, but also because it has a monotonic behaviour when the storage varies and the demand is fixed, or *vice versa*, as was exemplified in Item 2.3.3. For the purpose of avoiding misinterpretation involving the different concepts of reliability mentioned in this research, the time-based reliability will be represented by the acronym ER - *empirical reliability*, as discussed in Item 2.3.4.

The procedure did not consider the variations in storage due to precipitation over the reservoir, losses by evaporation, or other losses (apart from the spill that occur when the reservoir is full) which were neglected.

Given these conditions, the mass balance equation becomes:

$$S_{t+1} = S_t + Q_t - D_t \qquad (4.11)$$

where S_t is the storage, Q_t is the inflow (directly taken from the streamflows, X), and D_t is the demanded volume, all at the instant t, and t takes a monthly time-step, and the restriction $0 \leq S_{t+1} \leq C$ is made.

Furthermore, the simulations carried out followed the assumptions:

- The initial storage condition is that of a full reservoir ($S_1 = C$).

- The demand is fixed and uniform during the year, being defined by a given percentage of the mean annual flow of the streamflow series, $\frac{D}{\overline{Q}}$.

- The simulation does not handle storage dependent processes, that is, no adjustment is made to the demand in case the water level in the reservoir falls below a certain level.

- The simulation does not consider any sedimentation in the reservoir, that is, the storage capacity of the reservoir does not vary in time.

The following exemplifies the storage-yield analysis based on the simulation model applied to the streamflow records at the *Moinho do Bravo (2)* gauging station (sample no.10) with a recording period of 14 years (168 months) and a mean annual flow volumes of 41624 dam^3. The desired empirical reliability, ER, is of 90%, and the desired *draft*, *i.e.* the demand expressed as a percentage of the mean annual inflow volume uniformly distributed over 12 months, is 60%.

The analysis was carried out using a monthly time-step, as described. Because the desired empirical reliability, ER, is 90%, the iterative process will stop when the estimated storage capacity leads to an operation in which only 10% of the months will see restrictions on water use. However, 10% of the months does not equal an integer number of months ($0.9 \times 168 = 151.2$), this means that the problem does not have an exact solution because ER can never be exactly 90%; hence ER should be formally described as $90\% \pm \epsilon$ where ϵ is an error. In this case $\epsilon = \frac{|151-151.2|}{168} \times 100 = 0.12\%$. The goal can then be understood as finding the smallest storage capacity that leads to a reservoir operation in which only $168 - 151 = 17$ of the 168 months have restrictions applied to the water supply.

Figure 4.3 shows the behaviour diagram, that is, the fluctuation of water in storage during the period of operation. The reservoir is empty on two occasions, for a total of 17 months. The estimated storage capacity is 45914.73 dam^3.

The storage capacity is then represented as a percentage of the mean annual flow volume of the time series. This normalized quantification of the storage estimate is denominated *specific storage*. For the purpose

of simplifying the notation, the specific storage will be designated the fraction $\frac{C}{\overline{Q}}$, albeit representing a percentage. Likewise, the draft will be represented by $\frac{D}{\overline{Q}}$.

For the present example, $\frac{C}{\overline{Q}} = 110.31\%$
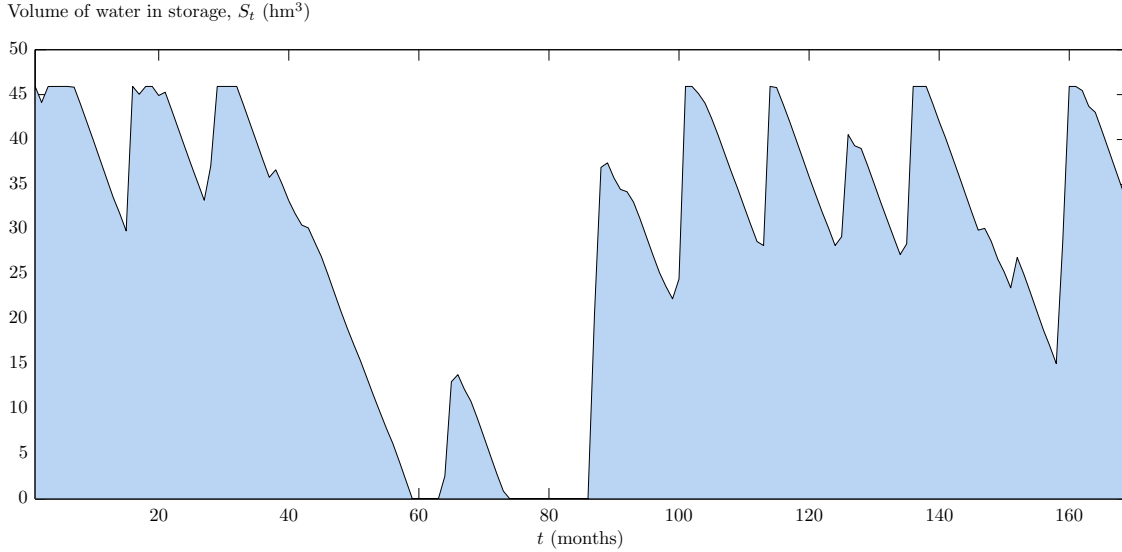
Volume of water in storage, S_t (hm³)

Figure 4.3: Behaviour diagram of a hypothetical reservoir on the *Moinho do Bravo* gauging station.

The storage-yield analyses were carried out considering:

- values of empirical reliability, ER, of 100% (full supply), 95%, 90%, and 80%;

- values of the draft, $\dfrac{D}{\overline{Q}}$, of 90%, 80%, 60%, 50%, 40%, and 20%.

The procedure exemplified based on the *Moinho do Bravo* gauging station was applied to the 54 streamflow samples, each resulting in one estimate of the reservoir specific storage capacity, $\frac{C}{\overline{Q}}$, for each combination of the values of ER and of $\frac{D}{\overline{Q}}$. Estimates of the vulnerability, η, and the resilience, φ, as defined by Equations (2.12) and (2.13), respectively, were also computed.

Furthermore, the procedure was applied to the 1200 synthetic flow series relative to each streamflow sample, resulting in 1200 estimates of the reservoir specific storage capacity $\left(\frac{C}{\overline{Q}}\right)^{(m)}$, with $m = 1, 2, ..., 1200$, for each combination of the values of ER and of $\frac{D}{\overline{Q}}$. The resulting total number of behaviour analyses carried out was near 1.6 Million.

It should be stressed that the behavioural analyses applied **to each synthetic flow series** at each specific gauging station considered that the mean annual flow volume, $\overline{Q}$, used in the adimensionalization of the demanded draft, $\frac{D}{\overline{Q}}$, and of the specific storage, $\frac{C}{\overline{Q}}$, was **the one pertained to that series**, thus differing among synthetic series and also differing from the historical sample.

4.3 Establishment of design curves

A large number of estimates of the specific storage capacity of a reservoir enables the application of a statistical analysis to define the design storage capacity, as illustrated in Figure 2.9 (although it is not expressed in the mentioned figure, the storage capacity C, may be expressed in a dimensionless form, as the specific storage, $\frac{C}{\overline{Q}}$). Under this understanding, and taking into account the previous research carried out on reservoir storage in Portuguese rivers (see Item 2.3.5), it is expected that the results of the behavioural analyses applied to a large number, M, of synthetic streamflow series, would allow for the following relationship to be established:

$$\Omega\left(\frac{C}{\overline{\overline{Q}}}, \frac{D}{\overline{\overline{Q}}}, ER, TR, \overline{H}\right) = 0 \tag{4.12}$$

where TR is the theoretical reliability, understood as a theoretical non-exceedance probability of the specific storage, and the remaining variables have the same meaning as in Equation (2.14).

For the purpose of applying a statistical analysis to the results of the storage-yield analyses, the mean, μ, and standard deviation, σ, of each sample of 1200 estimates of the specific storage, $\frac{C}{\overline{Q}}$, were estimated according to Equations (A.1) and (A.2) in Appendix A.

According to (Ribeiro, 1996, p. 79) and (Guimarães, 2005, p. 83), the distribution that better fits a large number of estimates of storage capacities of reservoirs is the Gumbel distribution. The probability factor of the Gumbel distribution is given by the following equation:

$$K_G = -\frac{\sqrt{6}}{\pi}\left\{0.5772 + \ln\left[\ln\left(\frac{1}{F}\right)\right]\right\} \tag{4.13}$$

where F is the non-exceedance probability.

Then the specific storage capacity associated with a determined value of F may be estimated by:

$$\frac{C}{\overline{\overline{Q}}} = \mu + K_G\sigma \tag{4.14}$$

where, for each river station, μ and σ are the mean and the standard deviation of the 1200 estimates of $\frac{C}{\overline{\overline{Q}}}$, as previously mentioned.

The values of the theoretical probability, TR considered in the statistical analyses that were carried out were 99%, 95%, 90%, and 80%.

In accordance with the curves developed by Portela & Quintela (2006a,b), which express the specific reservoir storage capacity as a function of the mean annual flow depth, $\overline{H}$, for each combination of values of ER, TR, and $\frac{D}{\overline{Q}}$ (a total of 96 possible combinations), equations of the type of Equation (2.15) were obtained with significant correlation coefficients.

5 Results and discussion

5.1 Assessment of the quality of the generated streamflow series

5.1.1 Previous considerations

As described in Item 4.1, for each of the gauging stations presented in Table 3.1, a two-level streamflow generation model was applied to generate 1200 synthetic annual and monthly streamflow series with an average length close to 35 years.

The results of the models at both temporal levels were evaluated in terms of the capacity of the generated synthetic series to preserve the main statistical characteristics of the corresponding samples. This assessment is made by comparing the synthetic series with the samples in terms of the main historical statistics, as described in Section 2.4.1. The comparison utilized confidence levels, computed by Equation (2.18).

In this analysis, diagrams were obtained which contain the relative position of the limits of the aforementioned confidence intervals, and the estimated samples' statistics.

A confidence level of $1 - \alpha = 95\%$ was adopted for the confidence intervals.

5.1.2 Results at the annual level

The generation of the annual streamflows takes place in the domain of the logarithmic transforms of the random variable (annual streamflows), hence, the confidence intervals and the samples' statistics to be used when appraising the quality of the model refer to the logarithms, namely the means, standard deviations and skewness coefficients of the logarithms of annual flows.

Figure 5.1, on the next page, displays the representation of those confidence intervals, together with the values of the samples' statistics. As a way of simplification, those values are referred to as historical values. In the figure, each sample is identified in the horizontal axis by the number attributed in Table 3.1.

The analysis of Figure 5.1 reveals that, for every sample, the historical values of the statistics are contained within the corresponding confidence intervals.
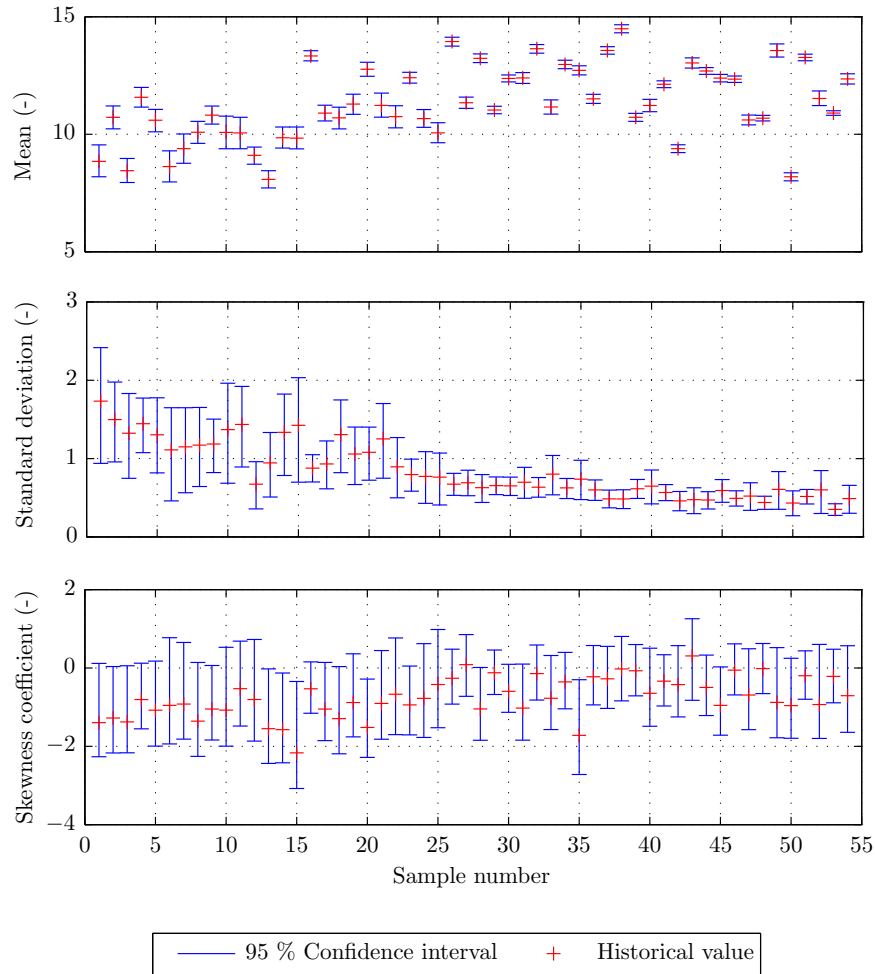
Figure 5.1: Confidence intervals at 95% of the means (top), the standard deviations (center), and the skewness coefficients (bottom) of the logarithms of annual streamflows. Comparison with the historical values of the same statistics.

An additional analysis was made by obtaining similar diagrams but in the domain of the random variable itself (the annual streamflow, in this case). However the representation of such diagrams in a single figure is not feasible due to the huge amplitude of the values of the means and of the standard deviations when referred to flow units (see Table 3.2 for the observed flow).

Notwithstanding this limitation, the appraisal of the preservation of the aforementioned statistical characteristics, however estimated on the basis of the annual flows themselves, showed that the inversion of the logarithmic transformation did not compromise the preservation of the means, of the standard deviations, or even, in the majority of the cases, of the skewness coefficient of the annual streamflows. In fact, the only case where the skewness coefficients was not preserved was the *Flor da Rosa* gauging station (sample no. 18).

48

Such results, along with those pertained to the logarithms of streamflows, are summarized in Table 5.1 which specifies the samples (identified by the numbers designated in Table 3.1) whose statistics were not preserved by the generation model.

Table 5.1: Results of the analysis of the preservation of the annual streamflow samples' statistical characteristics: samples - identified by the numbers designated in Table 3.1 - whose means, standard deviations, and skewness coefficients were not preserved by the respective synthetic series.

Variable	Statistic		
	Mean	Standard deviation	Skewness coefficient
Logarithm of the annual streamflow	–	–	–
Annual streamflow	–	–	18

5.1.3 Results at the monthly level

At the monthly level, the analyzed statistics were again the means, standard deviations and skewness coefficients of the monthly flows in each month of the hydrologic year.

It should be mentioned that the calculation of the skewness coefficients was not always possible due to the fact that, in some samples, the flow in a certain month is null in every year of the record. This is the case of the monthly flows in September and August at *Couto de Andreiros* and *Vidigal* gauging stations (Samples no. 11 and 13), respectively. Furthermore, the disaggregation process generated a number of synthetic monthly flow series of N years that present months with null flows in every year. In these cases it is also impossible to estimate the skewness coefficient. In order to standardize the procedures, if null flows in one or more months always occur in one or more of the 1200 synthetic series generated from a given sample, the confidence interval of the skewness coefficients of the flows in those months were not established, and therefore the analysis of the preservation of the skewness coefficient of monthly flows is not made in such months. Table 5.2 specifies the months (16 in total) that are in this situation. It should, however, be stressed that, apart from the skewness coefficient, the analysis of the preservation of the means and standard deviations focused on $54 \times 12 = 648$ samples of monthly flows, a number that, for the skewness coefficient is thus reduced to 632.

The application of the models at the monthly level showed that, for all 54 analyzed samples, the means of the monthly flow were always preserved. The standard deviations and the skewness coefficients were not preserved in a total of 3 months and 10 months, respectively.

Figures 5.2 to 5.4, on page 51, exemplify the confidence intervals and the historical values of the analyzed statistics of the monthly flows in *Albernoa*, *Açude Saimilo* and *Covas* (samples no. 1, 37 and 54). These gauging stations were selected as an example because they display values of the mean annual flow depth, $\overline{H}$, that are somewhat representative of the lower, medium, and high values in Mainland Portugal: 90.4, 633.8 and 2214.1 *mm*, respectively.

Table 5.2: Cases where the preservation of the skewness coefficient of flows was not analyzed in one or more months

Sample No.	Gauging station		Months where the preservation of the
	Code	Name	skewness coefficient was not analyzed [a]
2	27J/01	*Monte da Ponte*	August
5	24H/03	*Torrão do Alentejo*	August
8	25G/02	*Moinho do Bravo (1)*	July; August
9	24I/01	*Odivelas*	August; September
11	18L/01	*Couto de Andreiros*	July; August; **September**
13	30F/02	*Vidigal*	**August**; September
15	31K/03	*Bodega*	September
17	29L/01	*Monte dos Fortes*	September
19	28L/02	*Vascão*	September
21	30G/01	*Mte. dos Pachecos*	August; September

[a] The months written in bold indicates that the historical flow in that month was null in every year of that sample.

The remaining results of the analysis at the monthly level are displayed in Table 5.3 which specifies the samples (identified by the numbers designated in Table 3.1), and months whose statistics were not preserved by the generation and disaggregation models. It should be clarified that the months which were not subjected to the analysis of the preservation of the skewness coefficient (Table 5.2) do not take part in the analysis underlying Table 5.3.

Table 5.3: Results of the analysis of the preservation of the streamflow samples' statistical characteristics: samples - identified by the numbers designated in Table 3.1 - whose means, standard deviations, and skewness coefficients were not preserved by the respective synthetic series.

Month	Statistic		
	Mean	Standard deviation	Skewness coefficient
October	–	8; 35	17
November	–	–	–
December	–	–	15
January	–	–	30; 35
February	–	–	–
March	–	–	–
April	–	–	44
May	–	–	9; 18
June	–	–	–
July	–	–	–
August	–	18	–
September	–	–	20; 23; 51

The results represented in Figures 5.1 to 5.4, and the ones contained in Tables 5.1 and 5.3, clearly show that, generally, the proposed methodology for generating annual and monthly synthetic streamflow series accurately preserves the main statistical characteristics of the historical streamflow records in Portuguese rivers. Hence, the inclusion of that methodology in studies concerning the design of the storage capacities of artificial reservoirs in Mainland Portugal is validated.
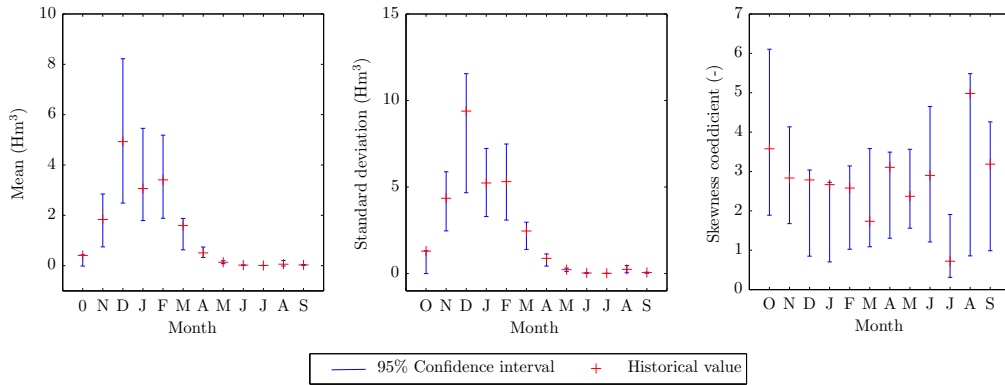
Figure 5.2: Confidence intervals at 95% of the means, the standard deviations, and the skewness coefficients of the monthly flows in *Albernoa* (sample no. 1).
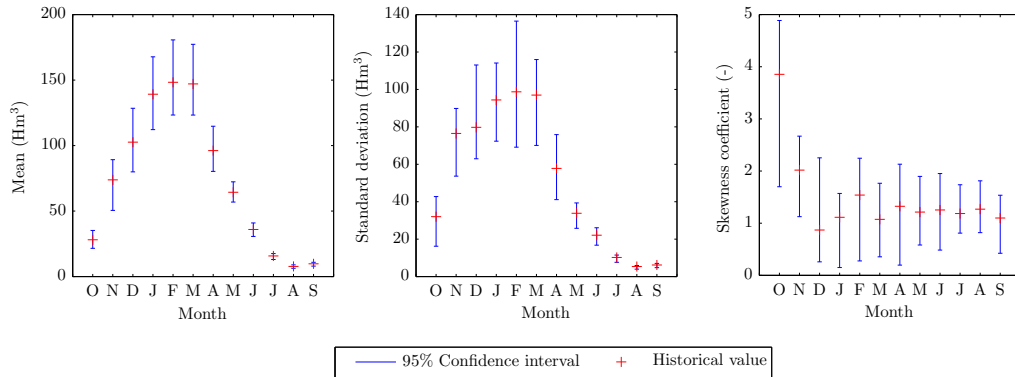

Figure 5.3: Confidence intervals at 95% of the means, the standard deviations, and the skewness coefficients of the monthly flows in *Açude Saimilo* (sample no. 37).
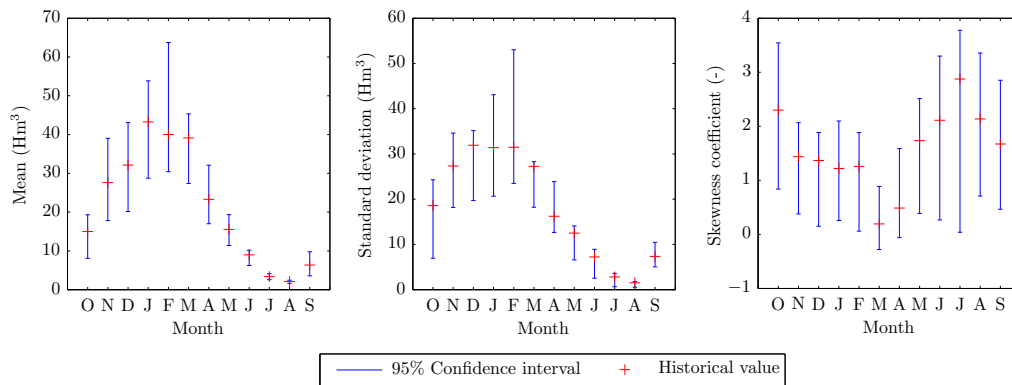

Figure 5.4: Confidence intervals at 95% of the means, the standard deviations, and the skewness coefficients of the monthly flows in *Covas* (sample no. 54).

5.2 Results of the storage-yield analyses

5.2.1 General remarks

The results from the storage-yield analyses have two different underlying conditions: either they were obtained directly from the historical streamflow samples or they resulted from the synthetic streamflow series. Therefore the presentation of such results was organized into separate items, according to the underlying condition of the streamflow series.

Another remark concerns the use of Equation (2.15). Along the presentation of the results there are references of curves resulting from that equation. Though it might by omitted, it should be stressed that the parameters of those curves were obtained be linear regression analysis applied to the logarithmic transforms of the variables that appear in Equation (2.15), followed by the inversion of the logarithmic transformation.

5.2.2 Reservoir vulnerability and resilience

5.2.2.1 Results based on historical streamflow samples

As discussed in Item 2.3.3, the reservoir performance metrics vulnerability, η, and resilience, φ, defined by Equations (2.12) and (2.13), respectively do not have a monotonic behaviour with an increasing storage capacity for a fixed yield, or *vice-versa*, as is exemplified by Figure 2.6. Therefore it was not feasible to use either of these metrics as a stopping criterion on the storage-yield procedure described in Item 4.2. Nevertheless it is possible to compute the values of those metrics when the stopping condition (defined by the empirical reliability, ER) is reached. However, their values can only be estimated for $ER < 100\%$, because their calculation presupposes the existence of shortages in the water supply.

For the 54 streamflow samples (historical data), Figure 5.5 shows the relationship between the vulnerability, η, and the mean annual flow depth, $\overline{H}$, for the empirical reliabilities, ER, of 95%, 90% and 80% and the target drafts of 20% to 90%. The analysis of this figure shows that, generally, under the same operation conditions regarding the target draft and the reliability of the supply, the vulnerability of reservoirs decreases as the mean annual flow depth increases. This trend is particularly visible for empirical reliabilities of 90% and 80%. It should be noted that for $\overline{H} > 500$ mm there is no case where the vulnerability is 1.0, this means that if $\overline{H} > 500$ mm, the water supply is never zero, for the considered reliability-draft combinations. Furthermore, although the lower estimates of the vulnerability occur in the $ER = 95\%$ graph, the same graph contains higher estimates of vulnerability, particularly for $\overline{H} > 500$ mm, than the $ER = 80\%$ graph, therefore increasing reliability does not always reduce vulnerability.
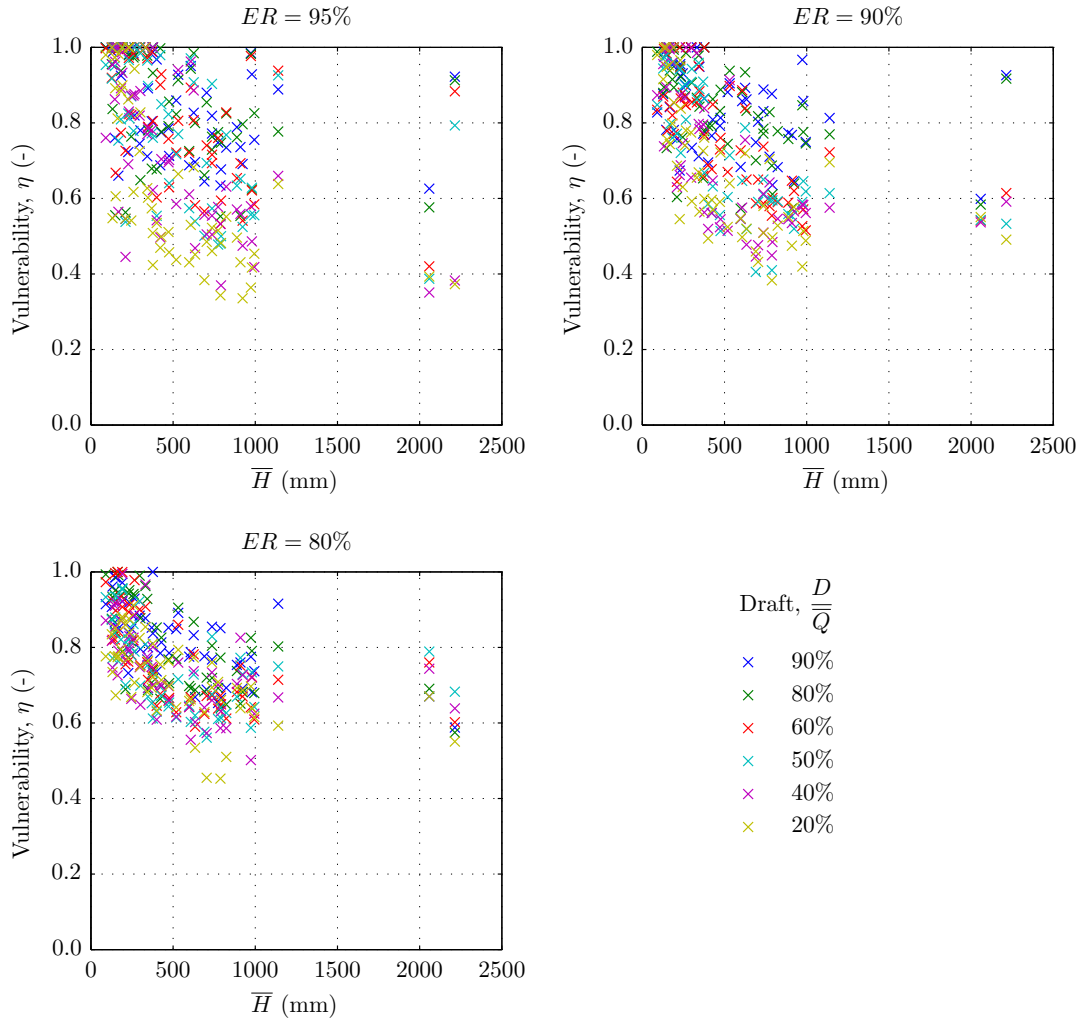
Figure 5.5: Relationship between the reservoir vulnerability estimates, η, and the mean annual flow depth, $\overline{H}$. Results based on the 54 historical streamflow samples, for fixed empirical reliabilities, ER, of 95%, 90% and 80%, and target drafts, $\frac{D}{\overline{Q}}$, from 20% to 90%.

Figure 5.6 shows the relationships between the reservoir resilience, φ, and the mean annual flow depth, $\overline{H}$, of each of the 54 samples, again for the empirical reliabilities, ER, of 95%, 90% and 80% and the target drafts of 20% to 90%. The analysis of this figure suggests that the resilience of a reservoir is more determined by the reliability of the supply than by the mean annual flow depth. The figure shows that the estimates of the resilience tend to increase as ER increases, being that this trend is more visible for low values of the draft.
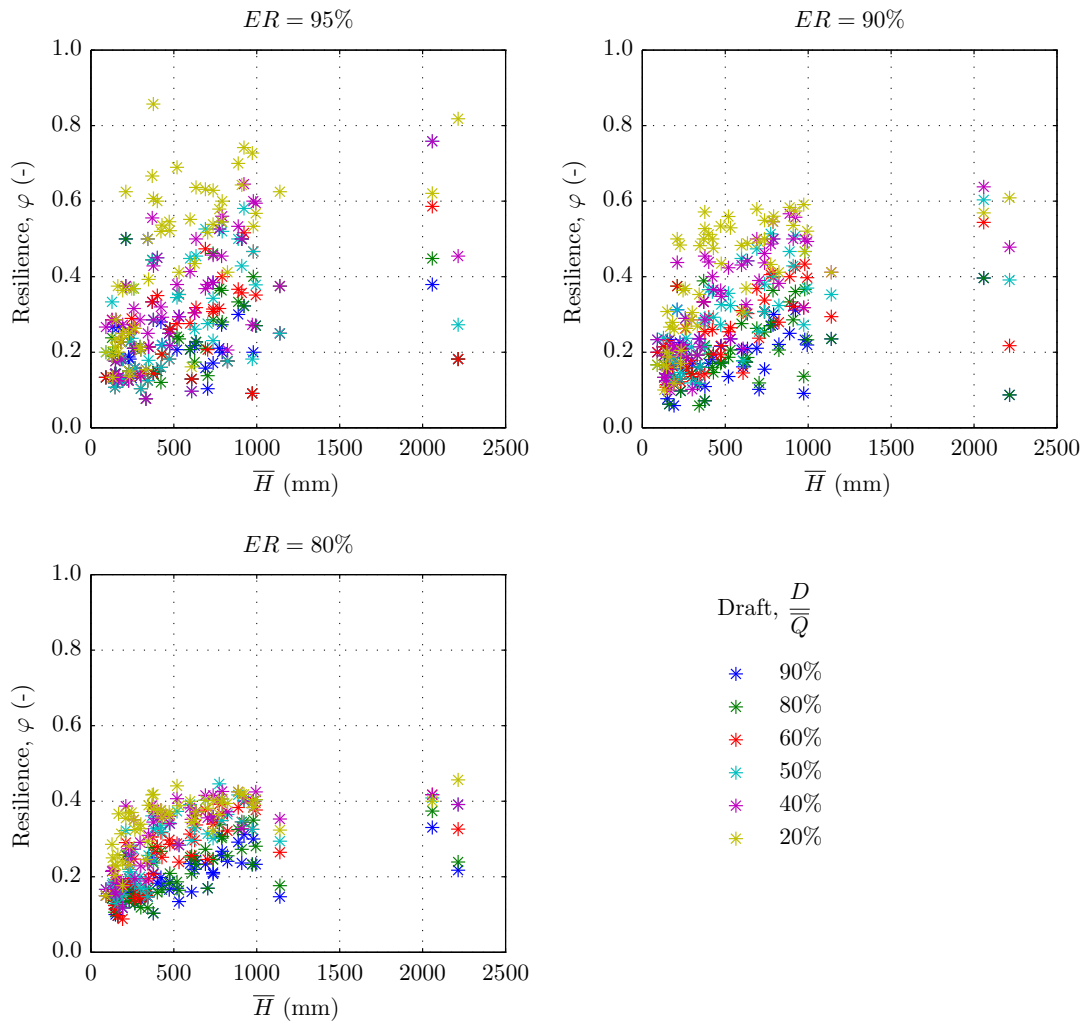

Figure 5.6: Relationship between the reservoir resilience estimates, φ, and the mean annual flow depth, $\overline{H}$. Results based on the 54 historical streamflow samples, for fixed empirical reliabilities, ER, of 95%, 90% and 80%, and target drafts, $\frac{D}{\overline{D}}$, from 20% to 90%.

In Figure 5.7 the vulnerability estimates are plotted against the resilience estimates. Linear regressions were made to establish the dashed lines that best fit the estimates, for each reliability-draft combination. The figure shows the two metrics, resilience and vulnerability, have an approximately complementary relationship.
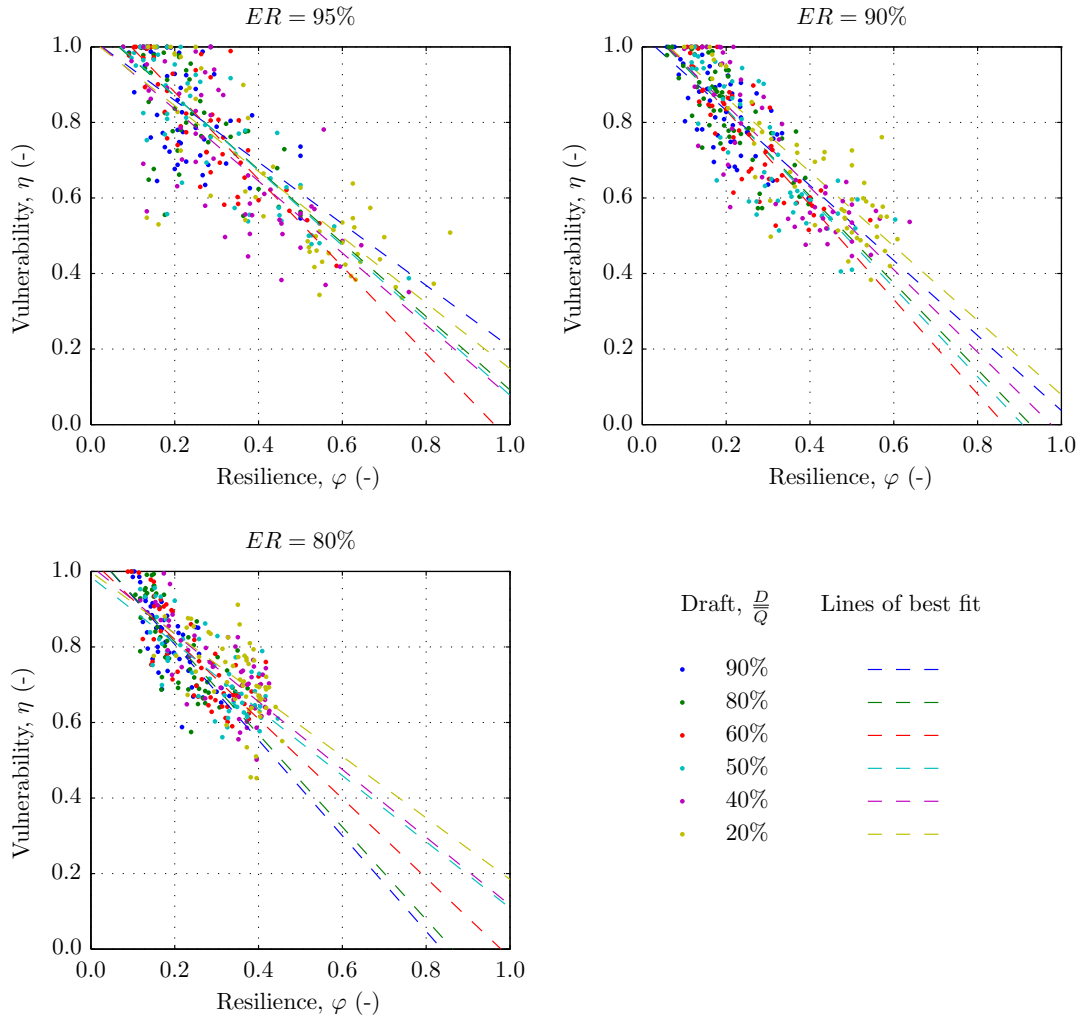

Figure 5.7: Comparison of vulnerability and resilience estimates. Results based on the 54 historical streamflow samples, for fixed empirical reliabilities, TR, of 95%, 90% and 80%, and target drafts, $\frac{D}{Q}$ from 20% to 90%.

Figure 5.8 displays the line of best fit of all the estimates shown in Figure 5.7. The line is defined by the equation $\eta = 1.03 - 0.96\varphi$ ($cc = -0.808$), which is similar to the lines obtained by McMahon *et al.* (2006).

The results of Figures 5.7 and 5.8 show that the relationship between the vulnerability and resilience performance metrics is not very linear in Portuguese rivers, therefore, both metrics should be estimated during proper storage-yield analyses.
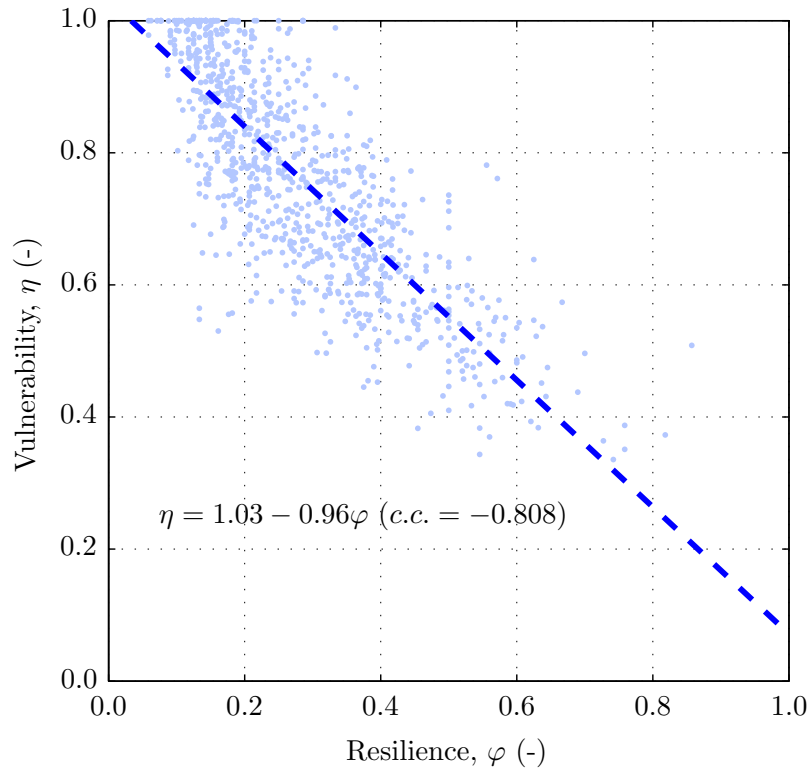

Figure 5.8: General relationship of the vulnerability and resilience estimates based on the 54 historical streamflow samples.

5.2.2.2 Results based on synthetic streamflow series

The estimates of the performance metrics resilience and vulnerability were also computed when applying the storage-yield procedure described in Item 4.2 to the generated synthetic streamflow series. In order to understand if the results concerning those two performance metrics based on historical data was consistent with the results based on the generated data, the comparisons pertained to Figures 5.5 to 5.8 were carried out using the mean estimates based on the generated streamflow series.

The relationships between those metrics and the mean annual flow depth of the streamflow samples are show in Figures 5.9 for the vulnerability and 5.10 for the resilience, on pages 57 and 58, respectively.
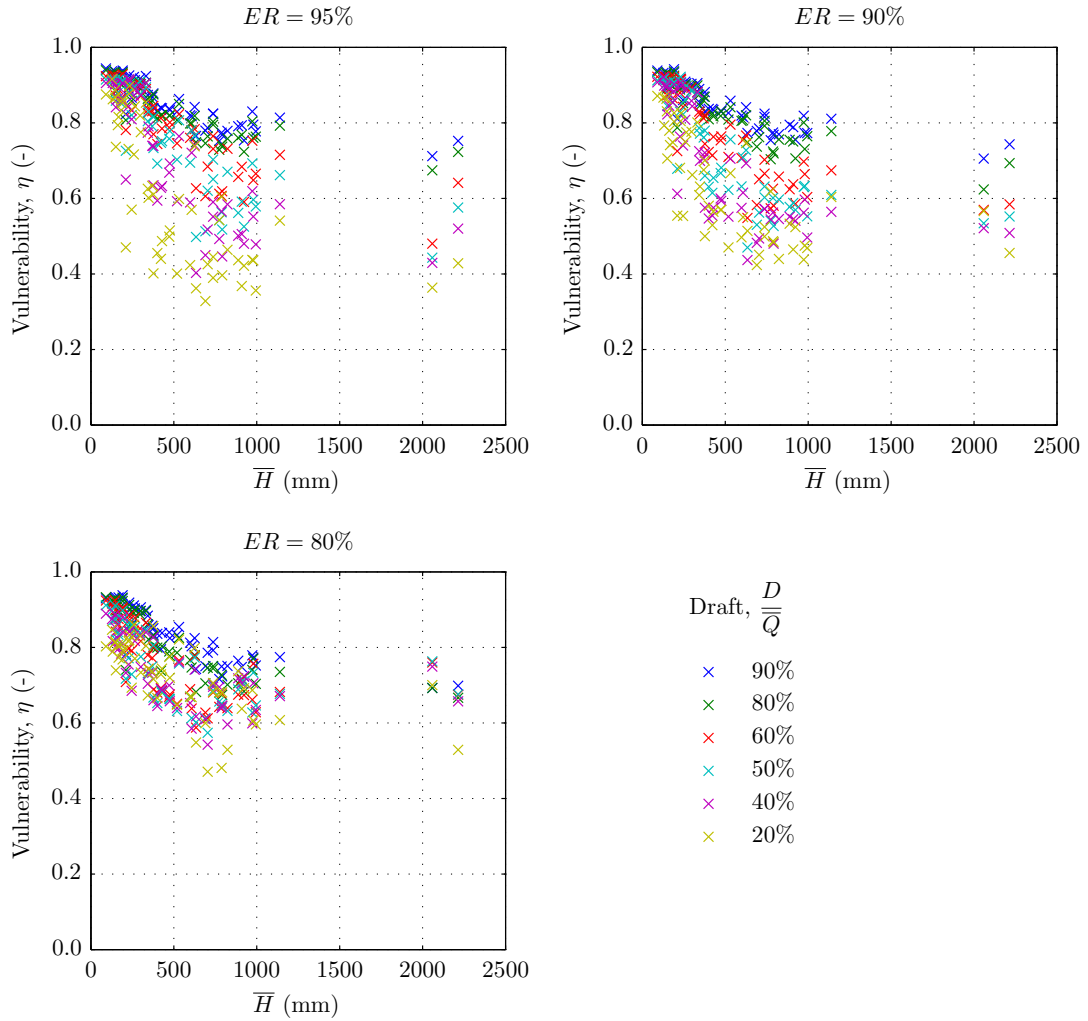
Figure 5.9: Relationship between the reservoir vulnerability estimates, η, and the mean annual flow depth, $\overline{H}$. Mean results based on the generated synthetic streamflow samples, for fixed empirical reliabilities, ER, of 95%, 90% and 80%, and target drafts, $\frac{D}{\overline{Q}}$, from 20% to 90%.

Figure 5.10: Relationship between the reservoir resilience estimates, φ, and the mean annual flow depth, $\overline{H}$. Mean results based on the generated synthetic streamflow samples, for fixed empirical reliabilities, ER, of 95%, 90% and 80%, and target drafts, $\frac{D}{\overline{Q}}$, from 20% to 90%.

Figure 5.11 presents the relationships, by means of the points, but also of the lines of best fit, between the resilience and vulnerability estimates for empirical reliabilities, ER, of 95%, 90% and 80% and target drafts of 20% to 90%.

Figure 5.11: Comparison of vulnerability and resilience estimates. Mean results based on the generated synthetic streamflow samples, for fixed empirical reliabilities, ER, of 95%, 90% and 80%, and target drafts, $\frac{D}{\overline{Q}}$, from 20% to 90%.
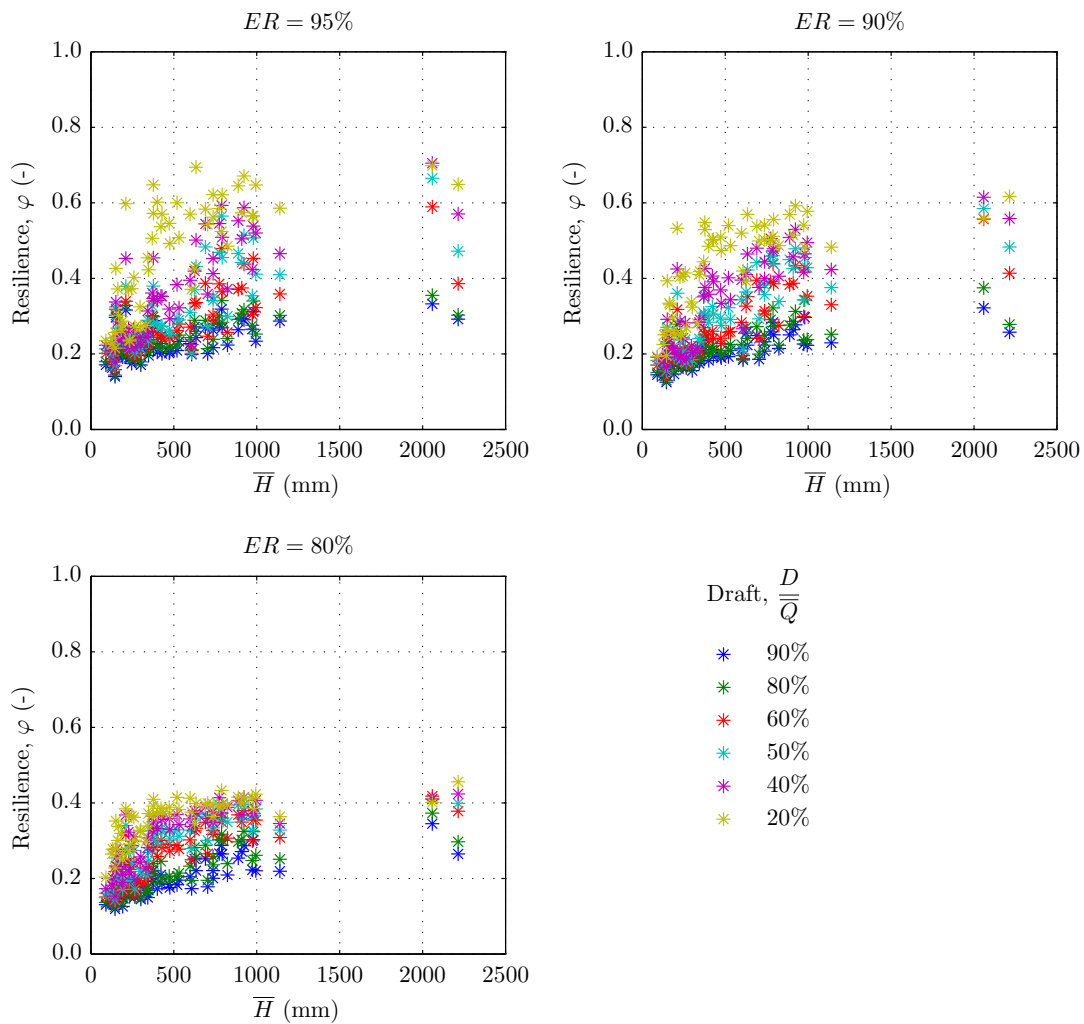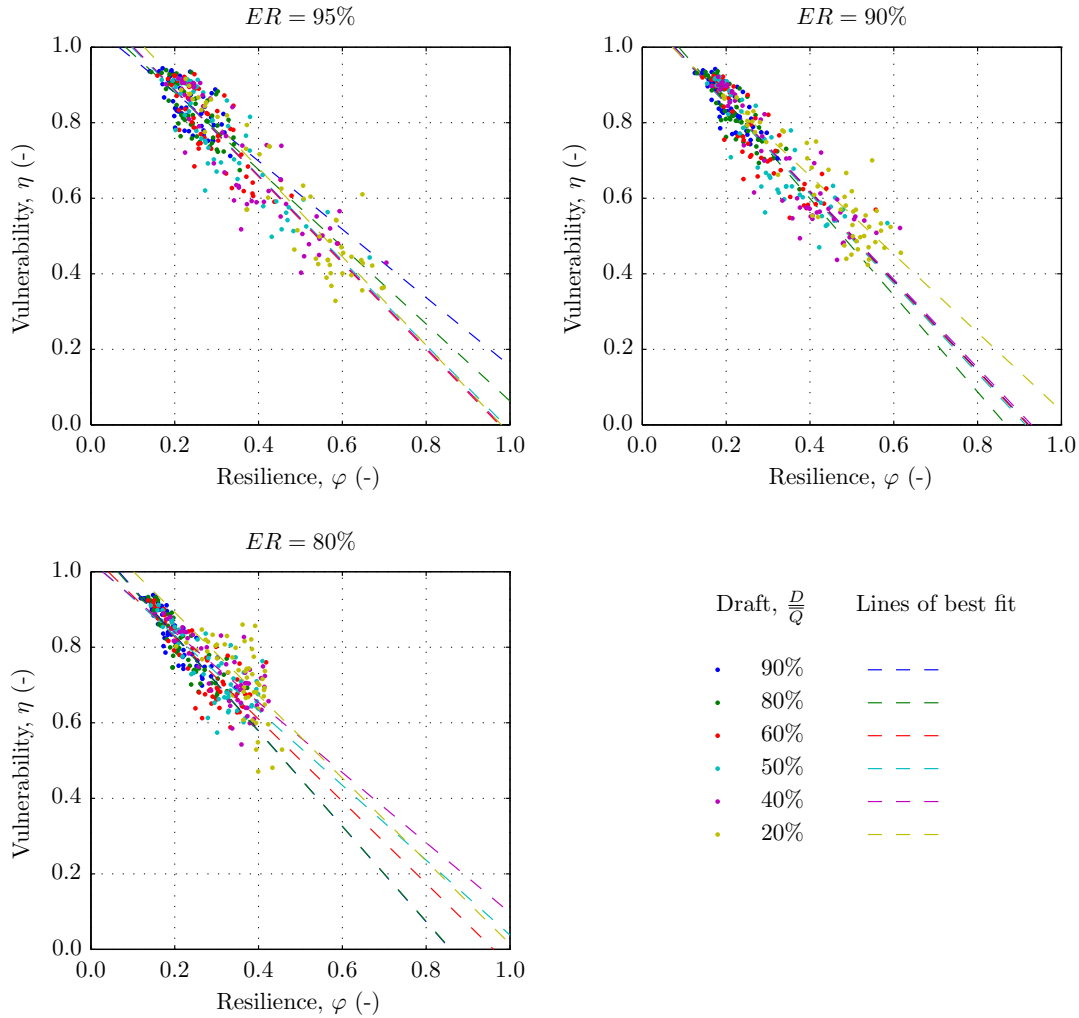
The general relationship between the mean estimates of resilience and vulnerability based on generated data is presented in Figure 5.12, which includes the equation that expresses the line of best fit to all the estimates: $\eta = 1.07 - 1.05\varphi$ ($c.c. = 0.889$). This line is consistent with the one of Figure 5.8 and with those obtained by McMahon *et al.* (2006) (Figure 2.7).

$$\eta = 1.07 - 1.05\varphi \ (c.c. = -0.889)$$
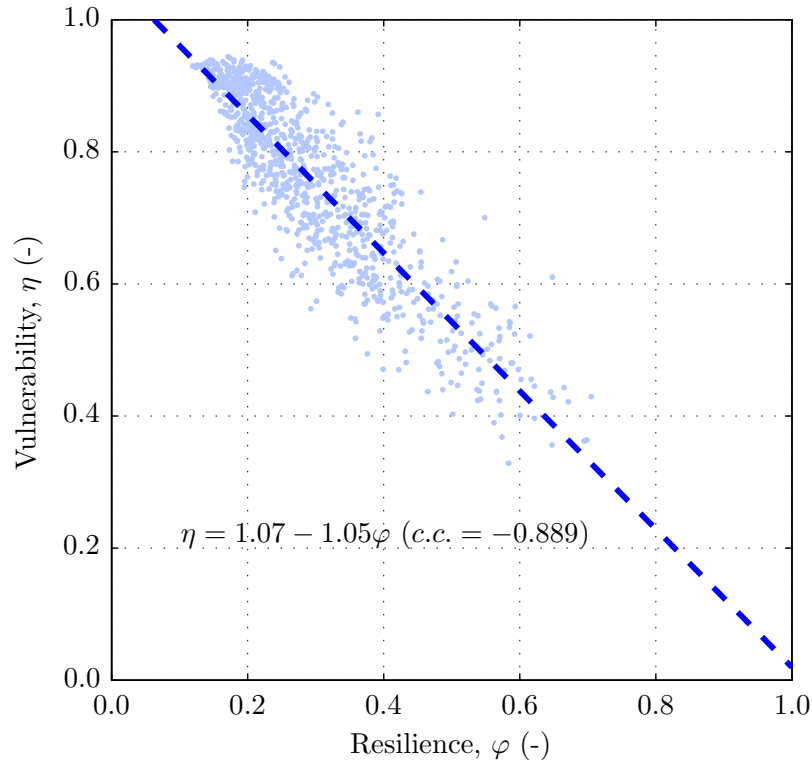
Figure 5.12: General relationship of the vulnerability and resilience estimates. Mean results based on the generated synthetic streamflow series.

It is apparent that the results based on the generated data are consistent with those based on historical data. Hence the generated synthetic streamflow series accurately reproduce the behaviour of hypothetical reservoirs, given the same operation conditions. These comparisons constitute an additional validation of the two-level streamflow generation procedure proposed in Item 4.1, in view of its application to reservoir storage-yield studies.

5.2.3 Reservoir capacity design curves

5.2.3.1 Comment on the results

Figure 5.13 presents an overview of the results achieved in this item, illustrating how they are interconnected, regarding the underlying nature of the data that supports them (be it historical or synthetic) as well as the treatments applied to such results, in accordance with the procedures described in Items 4.2 and 4.3.
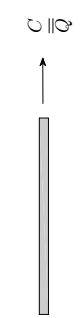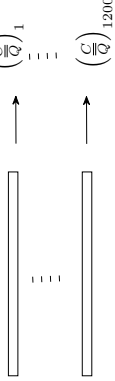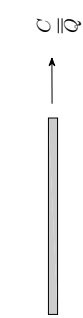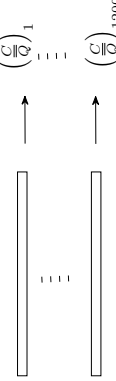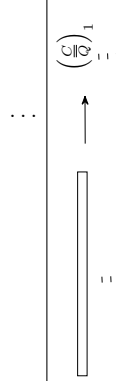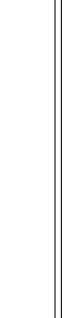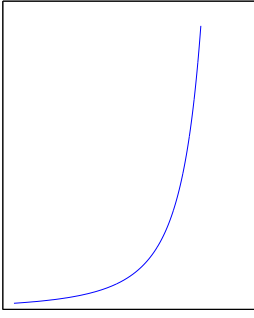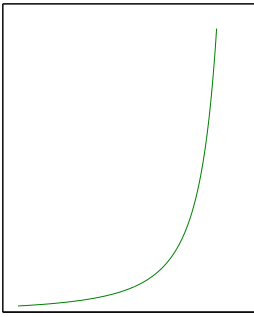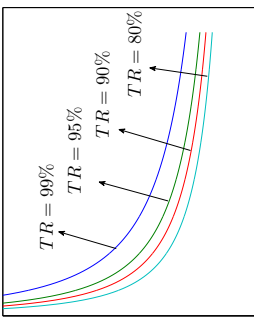
5.2.3.2 Curves based on historical streamflow samples

As mentioned in Item 4.2, the specific storage capacities, $\frac{C}{Q}$, of hypothetical reservoirs at the gauging stations of Table 3.1, were estimated, based on historical data, considering empirical reliabilities, ER, of 100%, 95%, 90%, and 80%, and target drafts, $\frac{D}{Q}$, of 20% to 90% of the mean annual inflow. The estimated specific storage capacities are related to the mean annual flow depth of the streamflow samples, $\overline{H}$, with significant correlations, which is consistent with the results obtained by Portela & Quintela (2002a,b, 2006a,b). Accordingly, curves defined by Equation (2.15) were fitted to the estimates of the specific storage capacities for several different values of ER and $\frac{D}{Q}$.

For the historical samples, Figures 5.14 (for $\frac{D}{Q} = 90\%$, 80%, and 60%) and 5.15 (for $\frac{D}{Q} = 50\%$, 40%, and 20%), on pages 63 and 64, respectively, show the points that express the relationship between $\overline{H}$ and $\frac{C}{Q}$ as well as the curves provided by linear regression analysis based on Equation (2.15). It should be stressed that the regression analysis was carried out not in the domain of the variables themselves but instead of their logarithmic transforms, as stated in Item 5.2.1.

Table 5.4, on page 65, contains the parameters of the fitted curves, α and β (see equation 2.15), as well as the correlation coefficients, $c.c.$, for each empirical reliability-draft combination, based on the storage-yield procedure applied to the historical samples.

The correlation coefficients obtained in Table 5.4 demonstrate that the relationship between the specific storage, $\frac{C}{Q}$, and the mean annual flow depth, $\overline{H}$, becomes more evident when the operation conditions of the reservoirs become more tolerant regarding failures in the supply (lower values of the empirical re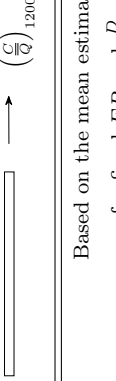liability) and less demanding on the natural inflow regime (lower values of the target draft, $\frac{D}{Q}$); consequently, when the operation conditions become more demanding and stringent (higher values of the empirical reliabilities and of the draft), the correlation coefficients become less significant.

Results for each combination of the empirical reliability, ER, (from 80% to 100%), and draft, $\frac{D}{Q}$, (from 20% to 90%)

Sample number (length, N, and mean annual flow depth, $\overline{H}$)	Based on historical data	Based on synthetic data	Statistical analysis
		1200 estimates for each sample	
$1\ (N_1, \overline{H}_1)$	$\frac{C}{Q}$	$\left(\frac{C}{Q}\right)_1 \cdots \left(\frac{C}{Q}\right)_{1200} \Rightarrow \text{Mean}\left(\frac{C}{Q}\right)$	$\frac{C}{Q}$ for the theoretical reliabilities, TR, of 99%, 95%, 90% and 80%
$2\ (N_2, \overline{H}_2)$	$\frac{C}{Q}$	$\left(\frac{C}{Q}\right)_1 \cdots \left(\frac{C}{Q}\right)_{1200} \Rightarrow \text{Mean}\left(\frac{C}{Q}\right)$	$\frac{C}{Q}$ for the theoretical reliabilities, TR, of 99%, 95%, 90% and 80%
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$54\ (N_{54}, \overline{H}_{54})$	$\frac{C}{Q}$	$\left(\frac{C}{Q}\right)_1 \cdots \left(\frac{C}{Q}\right)_{1200} \Rightarrow \text{Mean}\left(\frac{C}{Q}\right)$	$\frac{C}{Q}$ for the theoretical reliabilities, TR, of 99%, 95%, 90% and 80%
Based on the 54 samples	For fixed ER and $\frac{D}{Q}$	Based on the mean estimates for fixed ER and $\frac{D}{Q}$	Based on statistical analysis for fixed ER and $\frac{D}{Q}$

Legend: ▬▬ Historical streamflow sample ▬▬ Synthetic streamflow series

Figure 5.13: Schematic representation of the implementation of the storage-yield analyses.
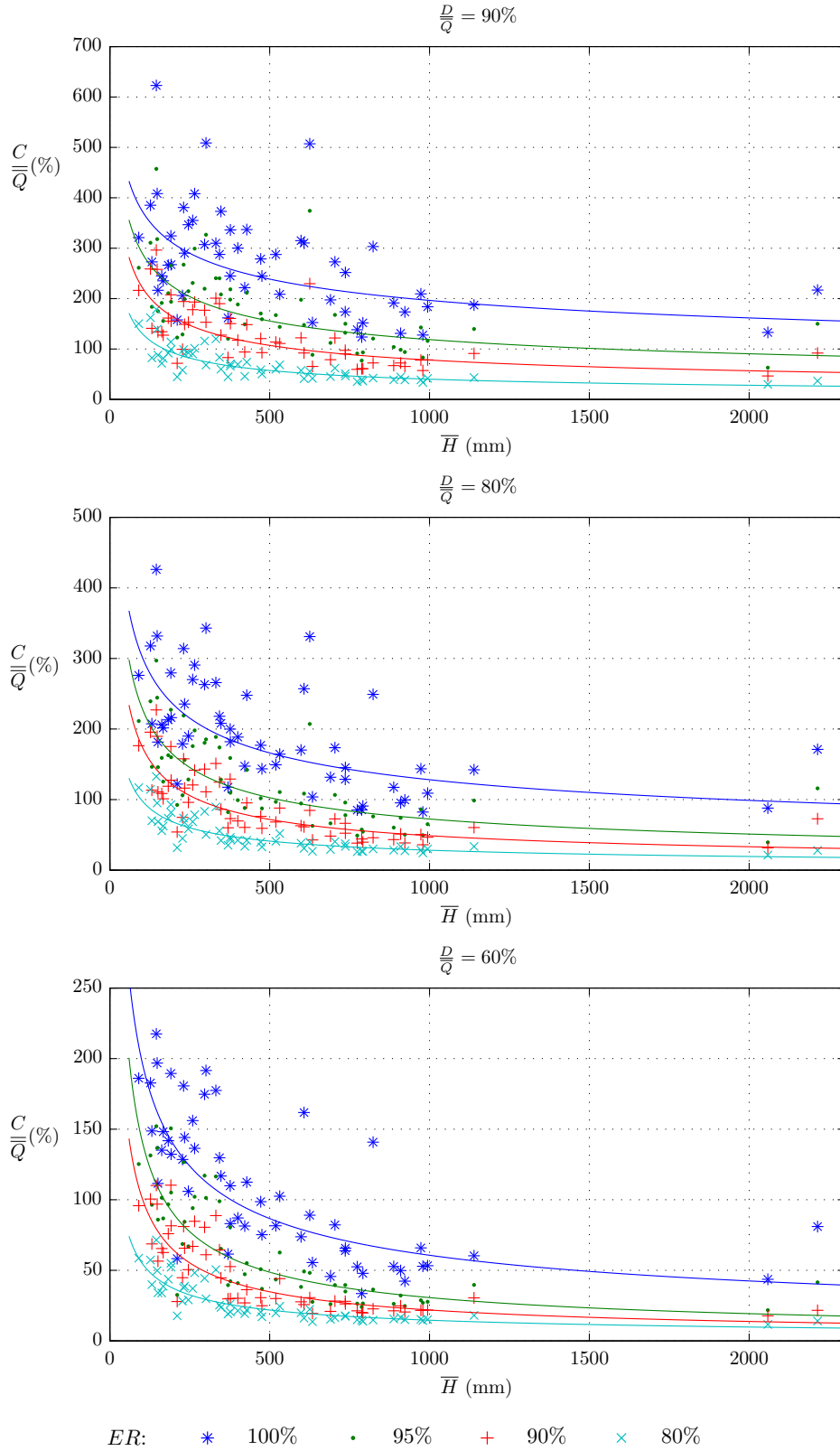
Figure 5.14: Specific storage estimates, $\frac{C}{\overline{Q}}$, based on the 54 streamflow samples for the empirical reliabilities, ER, of 80% to 100%, and target drafts, $\frac{D}{\overline{Q}}$, of 90%, 80% and 60%. Curves defined by linear regression applied to the logarithms of Equation (2.15).
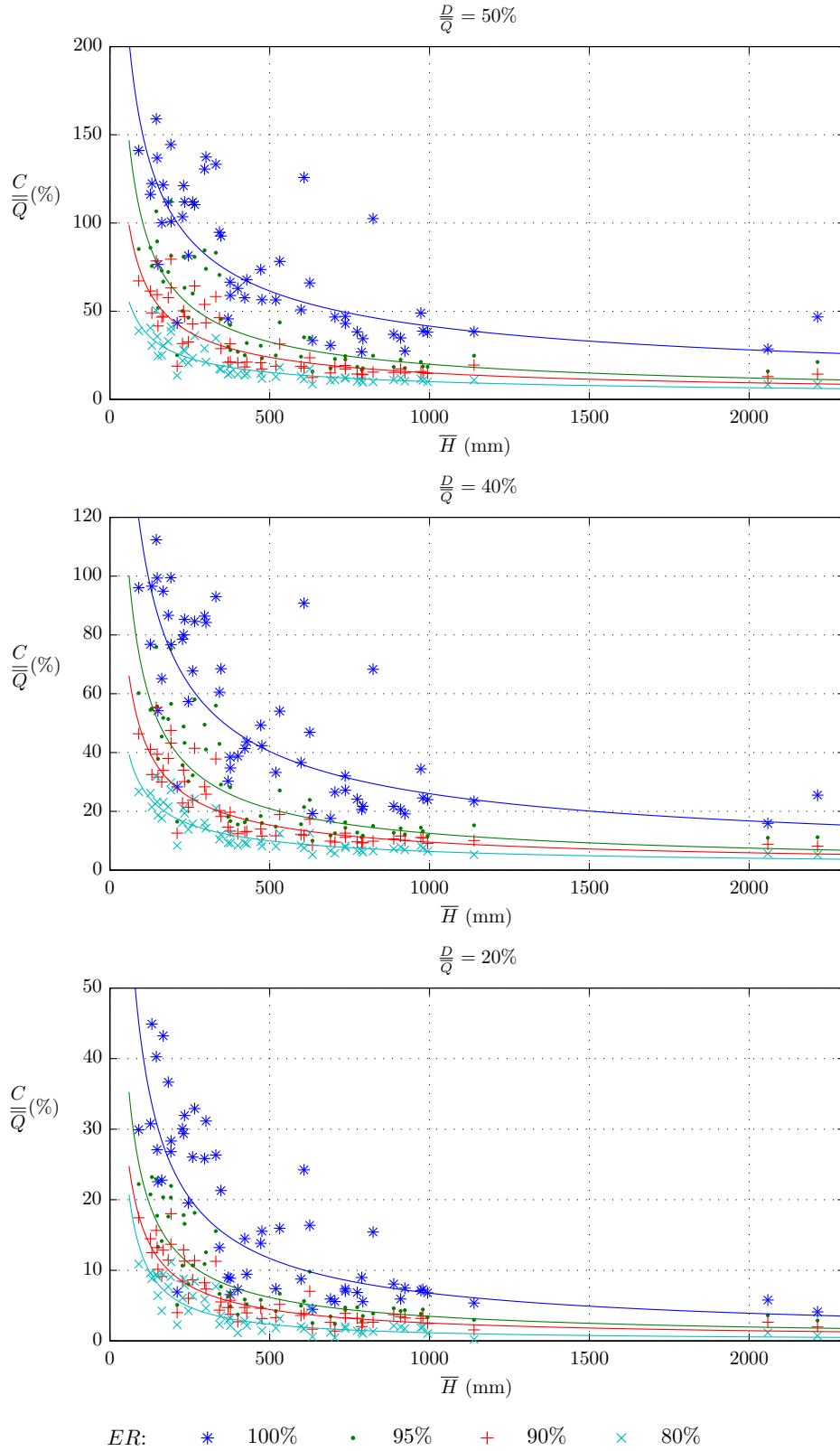
Figure 5.15: Specific storage estimates, $\frac{\overline{\overline{C}}}{\overline{Q}}$, based on the 54 streamflow samples for the empirical reliabilities, ER, of 80% to 100%, and target drafts, $\frac{\overline{\overline{D}}}{\overline{Q}}$, of 50%, 40% and 20%. Curves defined by linear regression applied to the logarithms of Equation (2.15).

Table 5.4: Results of the storage-yield analyses applied to the historical streamflow samples. Parameters α and β and correlation coefficients, $c.c.$, of the curves defined by Equation (2.15) for different values of the empirical reliability, ER, and the target draft, $\frac{D}{\overline{\overline{Q}}}$.

ER (%)	$\frac{D}{\overline{\overline{Q}}}$ (%)	α	β	$c.c.$
100	90	1367.4	-0.281	-0.548
	80	1702.3	-0.374	-0.658
	60	2124.0	-0.515	-0.760
	50	2054.7	-0.565	-0.785
	40	2086.2	-0.635	-0.800
	20	1576.0	-0.789	-0.815
95	90	1759.8	-0.390	-0.672
	80	2333.9	-0.503	-0.761
	60	3084.7	-0.668	-0.838
	50	2687.8	-0.710	-0.851
	40	2065.2	-0.739	-0.863
	20	1024.3	-0.823	-0.867
90	90	1831.0	-0.457	-0.744
	80	2286.6	-0.557	-0.811
	60	2207.0	-0.668	-0.860
	50	1542.6	-0.671	-0.860
	40	1109.1	-0.689	-0.882
	20	685.6	-0.811	-0.861
80	90	1409.4	-0.515	-0.848
	80	1208.9	-0.544	-0.861
	60	787.1	-0.577	-0.874
	50	663.0	-0.606	-0.888
	40	557.3	-0.647	-0.887
	20	1387.2	-1.026	-0.830

In item 3 it was mentioned that the data set utilized by Portela & Quintela (2006a,b) consisted of streamflow records in most of the same gauging stations utilized in the data set of the current research, albeit with slightly shorter recording periods. Many of the curves obtained in Figures 5.14 and 5.15 are directly comparable with the ones obtained by those authors. Figure 5.16 illustrates the aforementioned comparison of fitted curves, for empirical reliabilities, ER, of 95%, 90%, and 80% and target drafts, $\frac{D}{\overline{\overline{Q}}}$, of 90%, 50% and 20%.

The two sets of curves in Figure 5.16 are almost coincident, although some differences are visible for higher values of empirical reliability and draft, these differences should be explained by the shorter length of some of the observed streamflow series utilized by Portela & Quintela (2006a,b). As those authors noted, under more stringent operation conditions, a higher number of years would be necessary to fully characterize the restrictions on water supply. The samples utilized in the present research partially overcome this situation as they are longer. Despite the previous remarks, the results now obtained clearly reinforces the conclusions stated by those authors and also the hypothesis that the mean annual flow depth is capable of describing the storage needs of flow regulating reservoirs with significant correlation in Portuguese rivers.

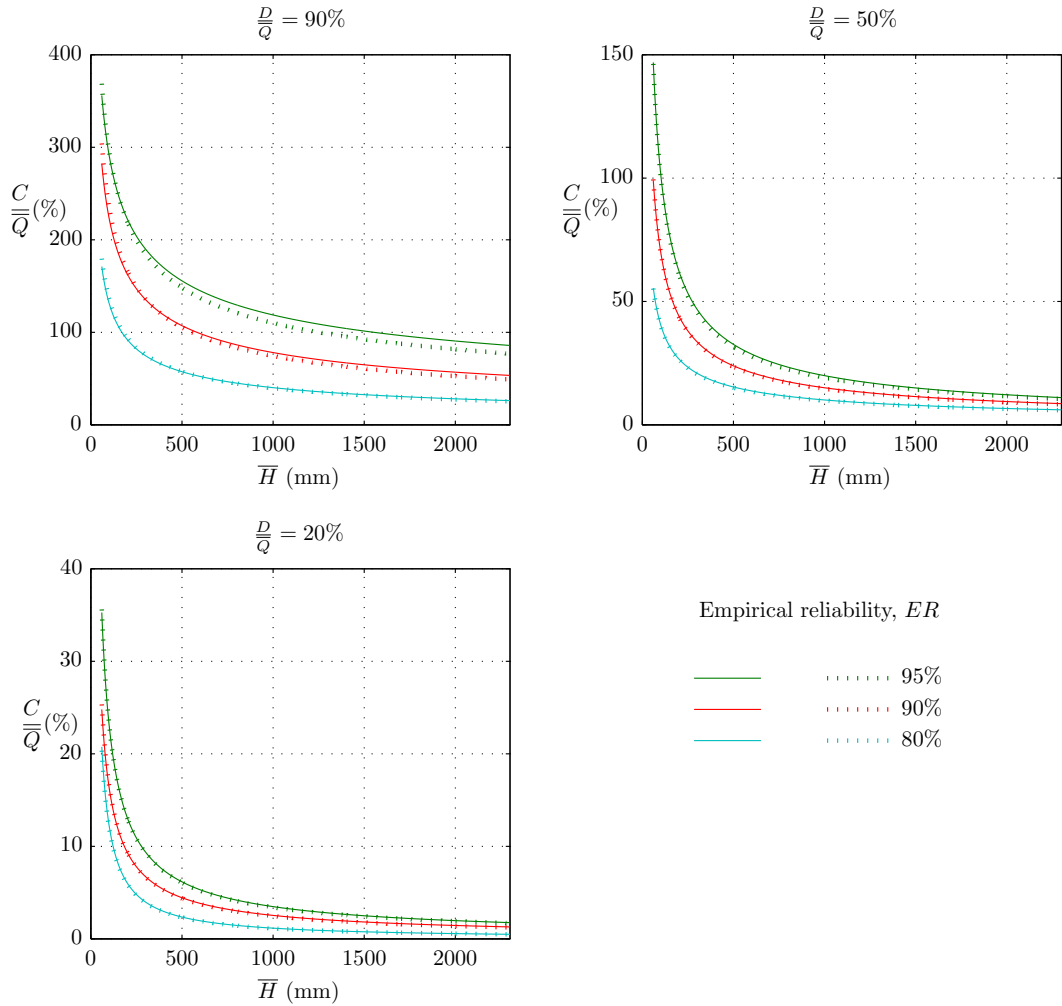Figure 5.16: Target drafts, $\frac{D}{Q}$, of 90%, 50% and 20%: curves defined by Equation (2.15) with the parameters α and β contained in Table 5.4 (solid curves) and equivalent curves obtained by Portela & Quintela (2006a,b) (dotted curves).

5.2.3.3 Curves based on synthetic streamflow series

Although the curves of Figures 5.14 and 5.15, or Equation (2.15) with the parameters of Table 5.4, constitute justifiable criteria for the preliminary design of the storage capacities of reservoirs in Portuguese rivers, it should be stressed that, for each set of values of $\frac{C}{Q}$, $\frac{D}{Q}$ and ER, they were obtained on the basis of a **single estimate of the specific storage for each gauging station**, the one provided by the historical sample of streamflows. As discussed in Item 2.3.4, when evaluating the storage requirements of a reservoir, a historical sample of monthly streamflows constitutes a single event, therefore it cannot result in more than one estimate of the reservoir storage capacity, and, accordingly, it should not be adopted as design criteria. Under this understanding, the reservoir storage-yield analyses carried out utilizing the 1200 synthetic streamflow series generated for each streamflow sample, as described in Item 4.2, resulted in the same number of specific storage estimates, which, in turn, could be assessed by means of statistical analysis tools.

To evaluate how accurately the generated streamflow series reproduce storage needs of reservoirs in Portuguese rivers, the results of the storage-yield analyses based on the generated data were compared with those of the analyses based on the observed data. That comparison was made by plotting the curves obtained by applying Equation (2.15) to the means of the 1200 specific storage estimates, obtained for the 54 samples, over the curves of Figures 5.14 and 5.15. The comparison of the two sets of curves is exemplified in Figure 5.17 for the target drafts, $\frac{D}{Q}$, of 90%, 50% and 20%. It should be noted that the mean results correspond to a theoretical reliability, TR, of nearly 57%, because the probability factor of the Gumbel distribution (Equation 4.13) is null for $F \approx 0.57$.

Figure 5.17 shows that for the same operation conditions the two sets of curves are almost coincident, although for higher values of the draft and reliability they may differ slightly. The comparison between the two sets of curves shows that the results based on generated data are consistent with the mean results based on historical data. Hence, that figure also provides a further confirmation of the adequacy of the proposed streamflow series generation model, in view of its inclusion in studies regarding reservoir storage design.

Table 5.5, on page 69, contains the parameters α and β and the correlation coefficients $c.c.$ of the curves defined by Equation (2.15) fitted to the mean estimates based on generated data. For each empirical reliability-draft combination, the correlation coefficients shown in that table are generally greater than the ones of Table 5.4 (using the historical data), which suggests that the effect of averaging such a large number of synthetic estimates accentuates the relationship between the specific storage of the reservoirs and the mean annual flow depth.
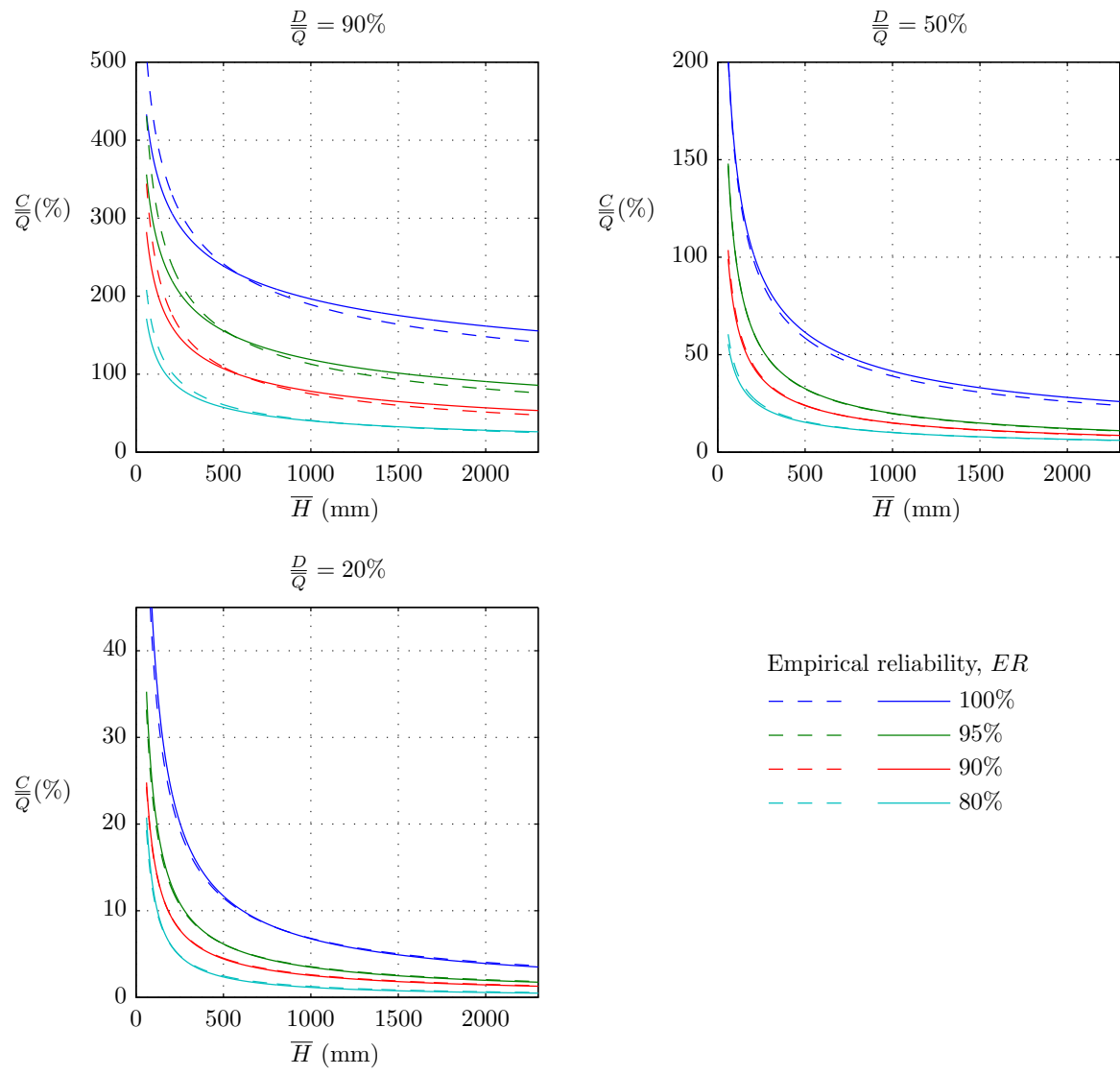
Figure 5.17: Curves defined by applying Equation (2.15) to: the mean results based on generated streamflows (dashed curves with the parameters shown in Table 5.5; the streamflow samples (solid curves with the parameters shown in Table 5.4).

Table 5.5: Mean results ($TR \approx 57\%$) of the storage-yield analyses applied to the synthetic streamflow series. Parameters α and β and correlation coefficients, *c.c.* of the curves defined by Equation (2.15) for different values of empirical reliability, ER, and target draft, $\frac{D}{Q}$.

ER (%)	$\frac{D}{Q}$ (%)	α	β	*c.c.*
100	90	2162.1	-0.353	-0.713
	80	2454.9	-0.430	-0.800
	60	2399.0	-0.537	-0.862
	50	2224.0	-0.585	-0.872
	40	2004.5	-0.637	-0.876
	20	1239.6	-0.753	-0.863
95	90	3024.0	-0.476	-0.841
	80	3557.6	-0.567	-0.882
	60	3385.2	-0.683	-0.901
	50	2774.6	-0.716	-0.898
	40	2023.3	-0.736	-0.891
	20	860.2	-0.795	-0.870
90	90	3169.1	-0.542	-0.878
	80	3304.2	-0.612	-0.897
	60	2442.0	-0.682	-0.902
	50	1740.1	-0.689	-0.894
	40	1161.2	-0.692	-0.887
	20	624.8	-0.793	-0.862
80	90	2232.4	-0.580	-0.891
	80	1867.5	-0.604	-0.895
	60	1099.3	-0.625	-0.895
	50	810.9	-0.634	-0.893
	40	637.1	-0.664	-0.892
	20	1022.1	-0.969	-0.841

Table 5.6, shows the parameters α and β, and the correlation coefficients, *c.c.*, of the curves defined by Equation (2.15) fitted to the results of the statistical analysis applied to the large number of estimates of the specific storage, for the theoretical reliabilities, TR, of 99%, 95%, 90% and 80%. As previously presented (Item 4.3), for each draft, $\frac{D}{Q}$ and empirical reliability, ER, the computation of the specific storage capacities, $\frac{C}{Q}$, for a given theoretical reliability, TR, requires the adjustment of a statistical law. For that purpose the Gumbel law was adopted. In such understanding, each theoretical reliability represents a theoretical non-exceedance probability, the corresponding value of the specific storage capacity being provided by the inversion of the probability distribution function of the above mentioned law.

Suppose that by means of the previous analysis the following results were achieved: $\frac{C}{Q} = 150\%$, for $\frac{D}{Q} = 80\%$ and ER of 95%, with a TR of 90%. This would mean that to fulfill a draft of 80% of the mean annual inflow volume, with an operation rule that only tolerates restrictions to the supply in $100 - 95 = 5\%$ of the months of the service life, a storage capacity of 150% of the same mean annual inflow volume would be needed. The probability of that storage requirement not being exceeded is 90%, which means that there is an underlying probability of 10% of non-fulfillment of the envisaged performance.

Table 5.6: Results of the storage-yield analyses applied to the synthetic streamflow series. Parameters α and β and correlation coefficients, $c.c.$ of the curves defined by Equation (2.15) for different values of theoretical reliabiliy, TR, empirical reliability, ER, and target draft, $\frac{D}{\overline{Q}}$.

ER (%)	$\frac{D}{\overline{Q}}$ (%)	$TR = 99\%$			ER (%)	$\frac{D}{\overline{Q}}$ (%)	$TR = 95\%$		
		α	β	$c.c.$			α	β	$c.c.$
100	90	4223.6	-0.358	-0.725	100	90	3388.1	-0.357	-0.722
	80	5151.9	-0.446	-0.821		80	4055.5	-0.442	-0.816
	60	5601.7	-0.572	-0.882		60	4285.4	-0.563	-0.878
	50	5112.2	-0.618	-0.887		50	3925.3	-0.609	-0.884
	40	4586.1	-0.669	-0.885		40	3518.9	-0.660	-0.885
	20	3692.7	-0.830	-0.879		20	2617.1	-0.807	-0.879
95	90	6944.7	-0.495	-0.845	95	90	5344.1	-0.490	-0.844
	80	10064.7	-0.625	-0.887		80	7326.1	-0.609	-0.887
	60	13872.6	-0.813	-0.898		60	9055.9	-0.777	-0.901
	50	13688.7	-0.884	-0.893		50	8419.0	-0.835	-0.898
	40	11760.8	-0.940	-0.891		40	6845.3	-0.879	-0.894
	20	4813.6	-1.004	-0.875		20	2821.4	-0.942	-0.877
90	90	9705.9	-0.614	-0.879	90	90	6917.6	-0.595	-0.881
	80	12391.6	-0.725	-0.896		80	8329.4	-0.694	-0.899
	60	13376.6	-0.874	-0.901		60	7969.8	-0.818	-0.904
	50	10841.6	-0.911	-0.898		50	6166.1	-0.845	-0.900
	40	7280.8	-0.922	-0.892		40	4106.2	-0.853	-0.894
	20	2384.8	-0.948	-0.873		20	1559.6	-0.901	-0.875
80	90	9302.5	-0.721	-0.885	80	90	6039.8	-0.681	-0.890
	80	9018.7	-0.780	-0.893		80	5575.8	-0.729	-0.897
	60	5580.6	-0.823	-0.894		60	3356.1	-0.763	-0.897
	50	3744.0	-0.821	-0.890		50	2305.9	-0.764	-0.894
	40	2360.4	-0.818	-0.890		40	1551.0	-0.770	-0.894
	20	1528.1	-0.966	-0.875		20	1318.6	-0.966	-0.870

ER (%)	$\frac{D}{\overline{Q}}$ (%)	$TR = 90\%$			ER (%)	$\frac{D}{\overline{Q}}$ (%)	$TR = 80\%$		
		α	β	$c.c.$			α	β	$c.c.$
100	90	3019.3	-0.356	-0.720	100	90	2634.8	-0.355	-0.717
	80	3572.5	-0.439	-0.813		80	3070.2	-0.436	-0.808
	60	3710.4	-0.557	-0.875		60	3116.8	-0.550	-0.871
	50	3406.8	-0.604	-0.882		50	2871.6	-0.597	-0.879
	40	3055.0	-0.655	-0.884		40	2578.2	-0.648	-0.882
	20	2173.4	-0.794	-0.878		20	1736.0	-0.778	-0.874
95	90	4641.1	-0.487	-0.844	95	90	3912.1	-0.483	-0.843
	80	6151.8	-0.600	-0.887		80	4960.2	-0.588	-0.886
	60	7148.4	-0.755	-0.902		60	5333.1	-0.728	-0.903
	50	6436.7	-0.808	-0.899		50	4622.2	-0.772	-0.900
	40	5085.9	-0.845	-0.895		40	3532.5	-0.803	-0.895
	20	2107.9	-0.907	-0.877		20	1476.7	-0.863	-0.876
90	90	5735.1	-0.583	-0.881	90	90	4546.8	-0.568	-0.881
	80	6677.3	-0.675	-0.900		80	5073.1	-0.651	-0.900
	60	5986.7	-0.786	-0.905		60	4206.5	-0.746	-0.905
	50	4526.1	-0.808	-0.900		50	3099.4	-0.761	-0.899
	40	3007.2	-0.814	-0.894		40	2058.5	-0.766	-0.893
	20	1241.1	-0.875	-0.874		20	942.3	-0.843	-0.871
80	90	4753.7	-0.658	-0.892	80	90	3534.3	-0.628	-0.893
	80	4276.8	-0.700	-0.898		80	3085.2	-0.663	-0.899
	60	2546.8	-0.730	-0.898		60	1821.5	-0.689	-0.899
	50	1776.5	-0.732	-0.896		50	1296.9	-0.694	-0.896
	40	1239.6	-0.744	-0.895		40	947.7	-0.712	-0.896
	20	1226.9	-0.966	-0.865		20	1132.8	-0.967	-0.857

The correlation coefficients in Table 5.6 show that the dependency between the specific storage estimates associated with a given probability of non-exceedance, or theoretical reliability, TR, and the mean annual flow depth of the historical sample are significant. Furthermore, these results validate the use of a large number of synthetic streamflow series to design the storage capacities of artificial reservoirs.

In order to illustrate how the curves defined by the parameters in Table 5.6 relate to the curves previously presented in the document, Figures 5.18 (for $ER = 80\%$ and $\frac{D}{\overline{Q}} = 40\%$) and 5.19 (for $ER = 95\%$ and $\frac{D}{\overline{Q}} = 80\%$), included in pages 72 and 73 respectively, exemplify the relative position of the specific storage estimates, $\frac{C}{\overline{Q}}$, using the observed streamflow records, and the curves defined by the parameters showed in Tables 2.2, 5.4, 5.5, and 5.6, in this last case for $TR = 80\%$ and 95% only.

The juxtaposition of the different sets of curves of each Figure 5.18 and 5.19 illustrate that **the curves established on the basis of historical streamflow data are concurrent, as a design criteria, with the curves obtained previously by Portela & Quintela (2006a,b) and with those supported by the mean results based on generated data, as was previously discussed**. The curves established on the basis of a statistical analysis applied to the large number of specific storage estimates resulting from the synthetic streamflow series (defined by the parameters in Table 5.6) have an additional underlying principle - the possibility of the non-fulfillment of the demanded supply associated with a given empirical reliability. Therefore, the application of the latter curves as design criterion produces more conservative, *i.e.* higher, storage capacity estimates, for the same operating condition and desired performance.

When applying the curves defined by the parameters in Table 5.6, the considered theoretical reliability, TR, should be proportionate to the relative importance of the reservoir system being designed. Evidently, if one is carrying out the preliminary design of a reservoir system of critical importance, one should adopt a higher standard regarding the theoretical reliability.

It should be pointed out that as the mean annual flow depth increases, and consequently, as the streamflow regime becomes more regular, the storage capacities provided by statistical approaches tend to be closer to those resulting from the historical data.

Figure 5.18: Empirical reliability, ER, of 80% and target draft, $\frac{D}{\overline{Q}}$, of 40%. Relative position of estimates of the specific storage of reservoirs and different design curves defined by liner regression on the logarithms Equation (2.15)
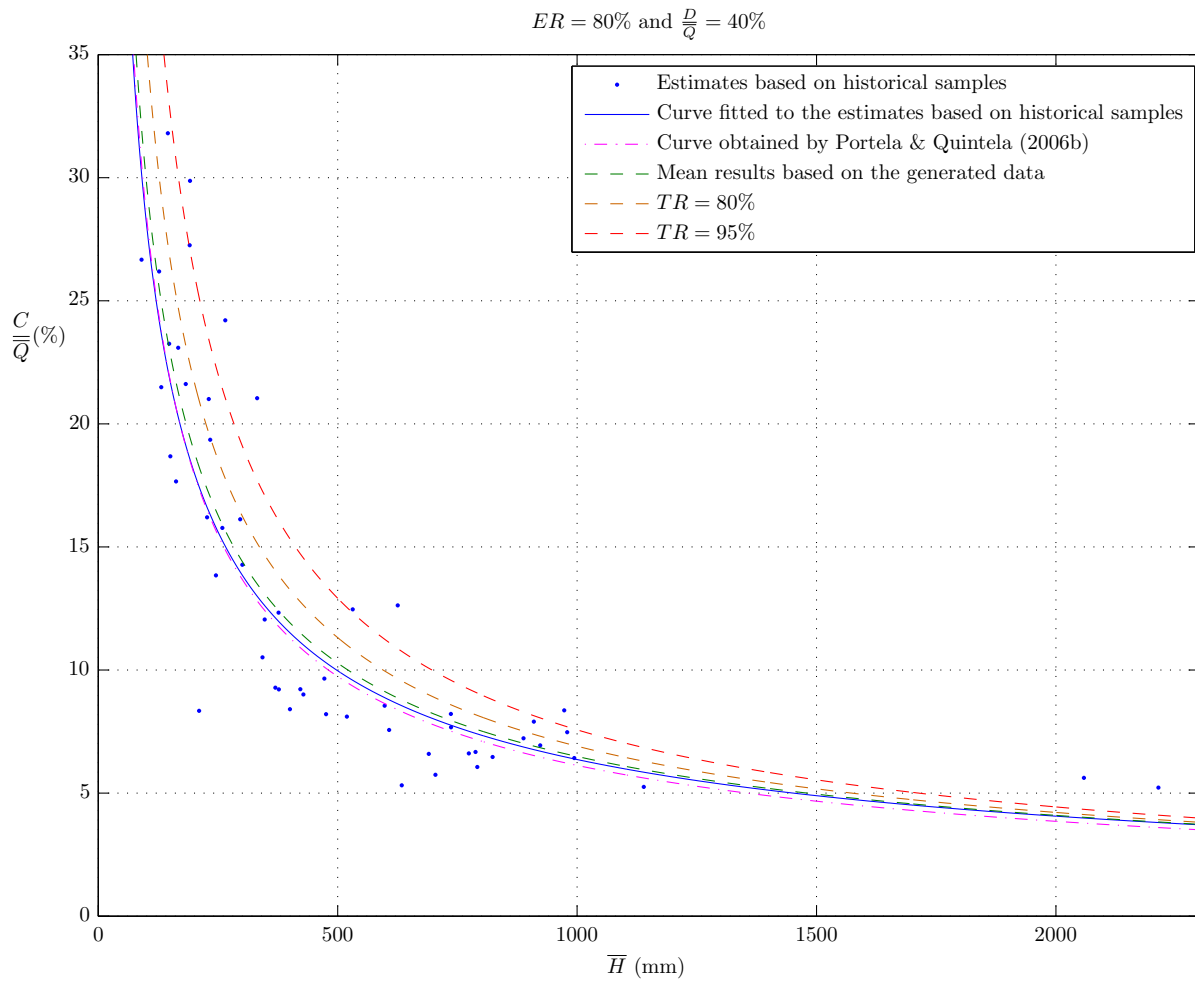
Figure 5.19: Empirical reliability, ER, of 95% and target draft, $\frac{D}{\overline{Q}}$, of 80%. Relative position of estimates of the specific storage of reservoirs and different design curves defined by liner regression on the logarithms Equation (2.15)
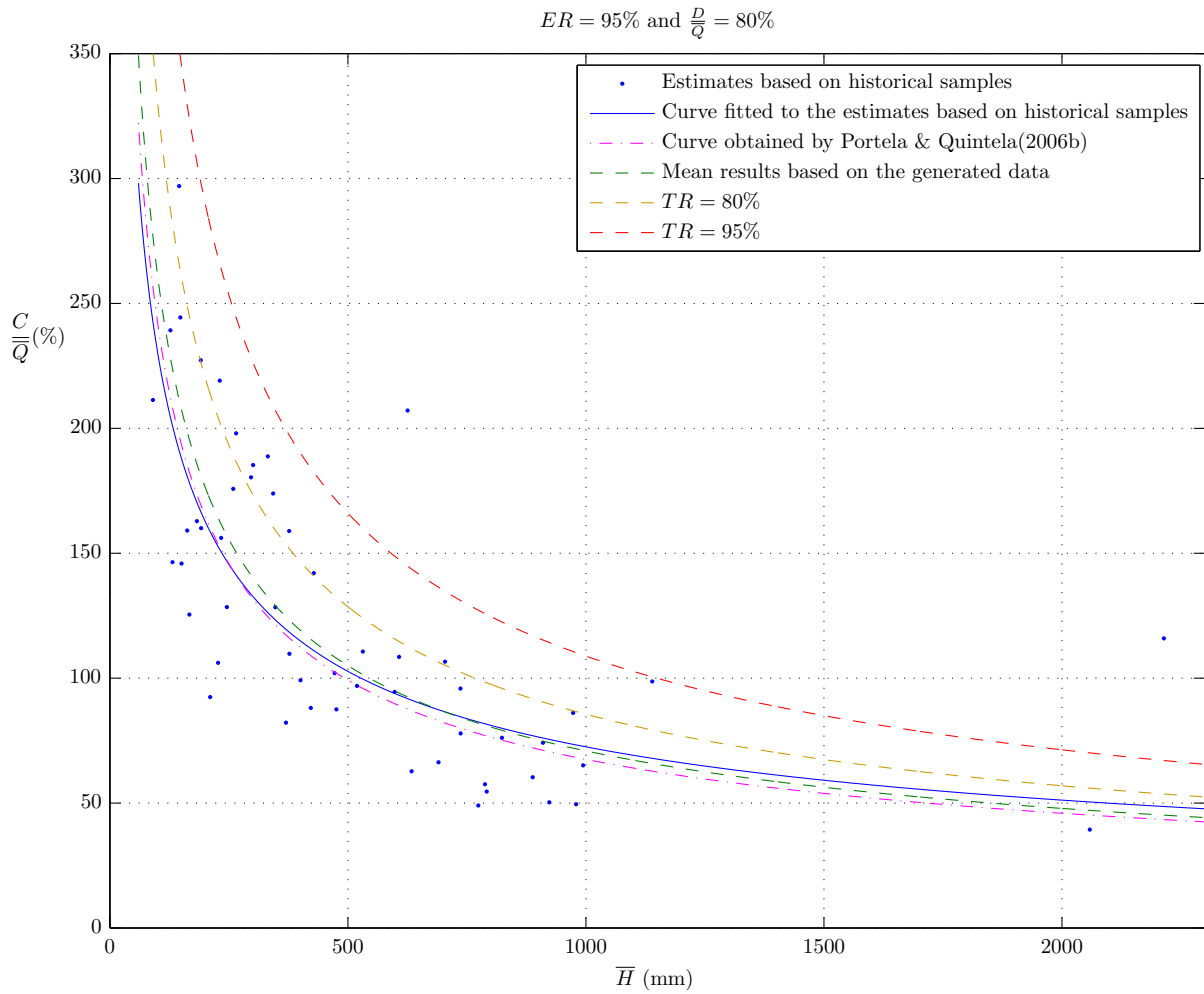
It is believed that the results contained in Table 5.6 may be directly applied to the preliminary design of a reservoir created by a dam in a Portuguese river, to evaluate:

i the storage capacity that insures the steady supply of a given annual volume of water, with a desired performance and a given reliability[1];

ii the annual volume that, for a given storage capacity and reliability, is possible to supply with a desired performance;

iii the performance of a steady supply of a desired annual volume, from a reservoir with a given storage capacity, and a given reliability;

iv the reliability associated with the non-exceedence of the storage capacity of a reservoir that provides a steady supply of a given annual volume with a desired performance.

The curves defined by the parameters presented in Table 5.6 constitute, therefore, a quick and useful tool that may be applied, at a preliminary level, to the planning and management of water resources systems involving the supply of water from a reservoir in Mainland Portugal.

---

[1]Here, the term "performance" refers to *empirical reliability*, and the term "reliability" refers to *theoretical reliability*.

6 Conclusions and future developments

In this dissertation a regional study was carried out on the design of the storage capacities of artificial reservoirs in Mainland Portugal.

The study follows previous research carried out on the design of the storage capacities of reservoirs in Portuguese rivers that concluded that the mean annual flow depth, $\overline{H}$, is a powerful parameter for the regionalization of hydrometric data, capable of characterizing the irregularity of the natural streamflow regimes and the storage requirements of reservoirs. However, the previous available results did not account for the intrinsic stochastic nature of the stremaflows. For this reason, the research carried out in this dissertation encompassed an extensive generation and utilization of synthetic streamflow data.

A procedure for generating synthetic annual and monthly streamflow series was developed and tested, integrating a probabilistic generation model at the annual level, and a disaggregation model - the method of fragments - at the monthly level. The procedure was applied to 54 streamflow samples from 53 gauging stations geographically spread over Mainland Portugal. A total of 22440 monthly flow records were collected, analyzed and utilized to generate nearly 27 million monthly streamflows.

The probabilistic model used for generating annual streamflows, which is based on a random sampling of the log-Pearson III law of probability, proved to be robust. The application of the model resulted in the preservation of the statistics of the historical samples, although those samples exhibit quite different statistical characteristics among them. This is due to the versatility of the log-Pearson III law, conferred by its three parameters. Since the statistical model generates logarithms of annual flows, the occurrence of synthetic negative streamflows is naturally avoided.

The method of fragments also proved to be appropriate to disaggregate annual streamflows into monthly streamflows. The procedure developed to automatically define the classes of fragments, which proved to lead to good results, completely eliminates any uncertainty in the establishment of those classes, thus increasing the robustness and generality of the disaggregation model.

The verification that, in general, the proposed two-level generation model accurately preserves the main statistical characteristics of the historical streamflow samples, confirms the adequacy of the procedure for generating synthetic series of annual and monthly streamflows.

A reservoir storage-yield procedure, based on the simulation algorithm, was applied to both the historical streamflow samples, and the synthetic streamflow series, for different empirical reliabilities and uniform water demands. The following is a summary of the conclusions achieved:

- By extensively using synthetic flows series the studies carried out are much more general than the previous ones on the subject as they account for the stochastic nature of the flow regime in the Portuguese rivers.

- Also, additional performance measures besides the empirical reliability, namely vulnerability and resilience, were computed and compared regarding the evaluation of the reliability of the water supplies based on artificial reservoirs in Portuguese rivers.

- Given certain operation conditions (fixed draft and empirical reliability), as the mean annual flow depth decreases, the reservoirs become increasingly more vulnerable to failure, *i.e.*, the magnitude of the shortage in the water supply increases as the inflows become more irregular.

- Given certain operation conditions, as the mean annual flow depth increases, the reservoirs become increasingly more resilient, *i.e.*, they recover from a failure more rapidly as the inflows become less irregular.

- Regardless of the operation conditions, the vulnerability and resilience of a reservoir tend to display a complementary relationship, which suggests that more resilient reservoirs tend to be less vulnerable, and *vice-versa*.

- As the mean annual flow depth increases, for a given operation condition, the storage requirements of a reservoir decreases, and *vice-versa*.

- A comparison of the results obtained using, primarily, historical data, and, subsequently, synthetic data, provided a proof of the adequacy of the proposed streamflow generation model, in view of its inclusion in studies regarding reservoir storage design.

- A large number of storage estimates may be subjected to statistical analysis with the objective of estimating the reliability (expressed as a theoretical non-exceedance probability) associated with a specific storage capacity.

The previous conclusions are consistent with the ability of the mean annual flow depth to characterize the streamflow variability in Portuguese rivers.

The obtained results made possible the establishment of curves that enable the estimation of the specific storage of reservoirs in Portuguese rivers, as a function of the mean annual flow depth, with significant correlations, for different reliabilities and performance levels. The correlation coefficients obtained for these

curves constitute an additional evidence of the aptitude of the mean annual flow depth also to assess the storage requirements of reservoirs.

It is believed that the obtained curves may be applied directly, at preliminary stages of studies involving water resources systems planning and evaluation, to the design of the storage capacities of artificial reservoirs. Hence, they constitute a useful and expeditious tool for the management of water resources in Mainland Portugal.

Although the proposed methodology was applied to hydrometric data from gauging stations in Portuguese rivers, it is believed that in other European regions, specifically in the South of Europe, analogous procedures may by developed with the objective of obtaining similar curves that are susceptible of being applied in those countries.

In terms of future developments a challenge remains concerning the assessment of non-uniform water demands, such as the seasonal demands required by agriculture. Furthermore, as the available streamflow samples increase, be it in number as well as in length, as long as the rivers remain unregulated, the procedures presented in this dissertation should be periodically implemented and their results updated. Evidently, when carrying out this updating, one should repeat the tests of independence in time and test the additional streamflow samples in order to detect possible trends or other signs of climate change, and to re-evaluate the effect of such trends on the overall performance of reservoirs.

Bibliography

ANDERSON, R. L. (1942). 'Distribution of the serial correlation coefficient'*. *The Annals of Mathematical Statistics* **13**(1), 1–13.

ARSÉNIO, J. J. G. (2003). *Gestão de Albufeiras: avaliação do potencial dos sistemas com recurso a séires sintéticas de escoamentos.* Master's thesis, Instituto Superior Técnico, Lisbon.

BOX, G., JENKINS, G. & REINSEL, G. (1994). *Time series analysis: forecasting and control.* Prentice Hall, third ed.

CHOW, V. T., MAIDMENT, D. & MAYS, L. (1988). *Applied hydrology.* Singapore: McGraw-Hill International Publications.

GUIMARÃES, R. C. P. D. C. (2005). *Simulação no Dimensionamento e Gestão de Sistemas de Recursos Hídricos. Geração de Séries Sintéticas de Escoamento.* Ph.D. thesis, Universidade de Évora, Évora.

HASHIMOTO, T., STEDINGER, J. & LOUCKS, D. (1982). Reliability, Resiliency, and Vulnerability Criteria for Water Resource System Performance Evaluation.*. *Water resources research* **18**(1), 14–20.

KLEMEŠ, V. (1987). 'One hundred years of applied storage reservoir theory'. *Water Resources Management* **1**(3), 159–175.

MATHWORKS, I. (2008). *MATLAB: the language of technical computing. Desktop tools and development environment, version 7.* Mathworks.

MCMAHON, T. & ADELOYE, A. (2005). *Water resources yield.* Colorado: Water Resources Publication.

MCMAHON, T., ADELOYE, A. & ZHOU, S. (2006). 'Understanding performance measures of reservoirs'. *Journal of Hydrology* **324**(1-4), 359–382.

MCMAHON, T., PEGRAM, G., VOGEL, R. & PEEL, M. (2007a). 'Revisiting reservoir storage-yield relationships using a global streamflow database'. *Advances in Water Resources* **30**(8), 1858–1872.

MCMAHON, T., VOGEL, R., PEGRAM, G., PEEL, M. & ETKIN, D. (2007b). 'Global streamflows–Part 2. Reservoir storage–yield performance'. *Journal of Hydrology* **347**, 260–271.

McMahon, T. A. & Mein, R. G. (1978). *Reservoir Capacity and Yield.* Amsterdam: Elsevier Scientific Publishing Company.

Moran, P. (1959). *The theory of storage*.* London: Methuen.

Naghettini, M. & Pinto, E. J. A. (2007). *Hidrologia estatística.* Belo Horizonte: CPRM Serviço Geológico do Brasil.

Nagy, I., Asante-Duah, D. & Zsuffa, I. (2002). *Hydrological dimensioning and operation of reservoirs: practical design concepts and principles.* Dordrecht: Kluwer Academic Publishers.

Portela, M. M. & Quintela, A. C. (2000). 'A altura do escoamento anual médio numa bacia hidrográfica como parâmetro de regionalização de informação hidrométrica'. In: *1º Congresso sobre Aproveitamentos e Gestão de Recursos Hídricos em Países de Idioma Português.* Rio de Janeiro.

Portela, M. M. & Quintela, A. C. (2001). 'A altura do escoamento anual médio como parâmetro de regionalização de informação hidrométrica'. *Tecnologia da Água* **4 (2)**, 16–22. Ed.: Elsevier Información Professional, S.A., ISSN 0211 8173, Portugal (invited paper).

Portela, M. M. & Quintela, A. C. (2002a). 'Assessment of The Streamflow Characteristics Under Unavailability of Discharge Data: The Mean Annual Flow Depth Over The Watershed As A Regionalization Parameter. The Portuguese Case.'. In: *EGS XXVII General Assembly, Nice, 21-26 April 2002*, vol. 27.

Portela, M. M. & Quintela, A. C. (2002b). 'Evaluation of the water resources in Portuguese watersheds without stream flow data'. In: *International Conference of Basin Organizations (Conferencia Internacional de organismos de Cuenca)*, vol. 6. Madrid.

Portela, M. M. & Quintela, A. C. (2005a). 'Estimação de séries de caudais médios diários na ausência de informação hidrométrica'. *VII Simpósio de Hidráulica e Recursos Hídricos dos Países de Língua Oficial Portuguesa* ,Évora.

Portela, M. M. & Quintela, A. C. (2005b). 'Regionalization of hydrologic information: establishment of flow series at ungauged watersheds'. *Water Resources Management III* , 11–20,Wessex Institute, WITPress, ISBN 1-84564-007-1, Southampton, Boston.

Portela, M. M. & Quintela, A. C. (2006a). 'Estimação em Portugal Continental de escoamentos e de capacidades úteis de albufeiras de regularização na ausência de informação'. *Recursos Hídricos* **27**(2), 7–18. Número Temático: Hidrologia e Modelação Hidrológica.

Portela, M. M. & Quintela, A. C. (2006b). 'Preliminary design of the storage capacity of reservoirs based on a flow regionalization parameter'. In: *BALWOIS 2006, International Conference on Water Observation and Information Systems for Decision Support.* Ohrid.

Quintela, A. (1967). *Recursos de águas superficiais em Portugal Continental**. Ph.D. thesis, Instituto Superior Técnico, Lisbon.

Quintela, A. C. & Portela, M. M. (2002). 'A modelação hidrológica em Portugal nos últimos 25 anos do século XX nas perspectivas determinística e estocástica.'. *Recursos Hídricos* **23**(2), 7–22.

Ribeiro, P. J. d. S. F. (1996). *Dimensionamento do Volume Útil de Albufeiras.* Master's thesis, Instituto Superior Técnico, Lisbon.

Rippl, W. (1883). 'The capacity of storage reservoirs for water supply'*. *Minutes of proc.* **71**, 270–278.

Salas, J. D., Delleur, J. W., Yevjevich, V. & Lane, W. L. (1980). *Applied Modeling of Hydrologic Time Series.* Colorado: Water Resources Publications, second ed.

Santos, E. G. (1983). *Disaggregation modeling of hydrologic time series.* Ph.D. thesis, Colorado State University, Colorado.

Silva, M. M. P. R. (1989). *Desagregação diária por etapas. Aplicação à geração de séries sintéticas de escoamentos diários.* Master's thesis, Instituto Superior Técnico, Lisbon.

Sudler, C. (1927). 'Storage required for the regulation of streamflow'*. *Transactions of the American Society of Civil Engineers* **91**.

Svanidze, G. G. (1980). 'Mathematical modeling of hydrologic series; for hydroelectric and water resources computations.'*. *Water Resources Publications* .

Thomas, H. & Burden, R. (1963). 'Operations research in water quality management'*. Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachussets.

Valencia, R. D. & Schaake Jr, J. (1972). 'A disaggregation model for time series analysis and synthesis'*. *Ralph M. Parson Laboratory, Report no. 149* .

Vaz, A. C. (1984). *Modelos de Planeamento de sistemas de albufeiras em condições de incerteza.* Ph.D. thesis, Instituto Superior Técnico.

(*) Reference was not consulted directly. The source of the citation is indicated in the text.

Appendix A

Basic statistical parameters

The statistical representation of a univariate series of N random variables, x_t, is based on a set of *statistics*, or *statistical parameters*. The following is a presentation of the definition and formulae adopted in the dissertation of the most relevant statistics.

1 The *mean*, $\overline{x}$, is a measure of central tendency:

$$\overline{x} = \frac{1}{N} \sum_{t=1}^{N} x_t \tag{A.1}$$

2 The *standard deviation*, s, is a measure of the degree to which the data spreads about the mean value. The unbiased estimation of this parameter is obtained by:

$$s_x = \sqrt{\frac{1}{N-1} \sum_{t=1}^{N} (x_t - \overline{x})^2} \tag{A.2}$$

3 The *coefficient of variation*, C_V, is a dimensionless measure of relative variability:

$$C_V = \frac{s_x}{\overline{x}} \tag{A.3}$$

4 The *skewness coefficient*, g, is a measure of the asymmetry of the distribution of the values of x_t. The unbiased estimation of this parameter is obtained by:

$$g(x) = \frac{N}{(N-1)(N-2)s_x^3} \sum_{t=1}^{N} (x_t - \overline{x})^3 \tag{A.4}$$

5 The *lag-k autocovariance*, c_k, is a measure of the degree of self-dependence of a time-series. The autocovariance c_k between x_t and x_{t+k} is obtained by:

$$c_k = \frac{2}{N} \sum_{t=1}^{N-k} (x_t - \overline{x})(x_{t+k} - \overline{x}), 0 \leq k < N \tag{A.5}$$

6 The *lag-k autocorrelation coefficient* is a dimensionless measure of linear dependence. this statistic is obtained by dividing c_k of Equation (A.5) by c_0:

$$r_k = \frac{c_k}{c_0} = \frac{\sum\limits_{t=1}^{N-k} (x_t - \overline{x})(x_{t+k} - \overline{x})}{\sum\limits_{t=1}^{N} (x_t - \overline{x})^2} \tag{A.6}$$