On the use of TCEV and Kappa four-parameter distributions for at-site flood frequency analysis

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

Il sottoscritto Vincenzo Totaro nato a Foggia il 11/01/1987 residente a Bovino (FG) in via della Repubblica 6/a e-mail vincenzo.totaro@poliba.it iscritto al 3° anno di Corso di Dottorato di Ricerca in Rischio, Sviluppo Ambientale, Territoriale ed Edilizio, ciclo XXXII

ed essendo stato ammesso a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

On the use of TCEV and Kappa four-parameter distributions for at-site flood frequency analysis

DICHIARA

1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritto al Corso di Dottorato di ricerca in Rischio, Sviluppo Ambientale, Territoriale ed Edilizio ciclo XXXII, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari*", emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data BARI, 19/12/2019 Firma Vincenzo Totaro

Il/La sottoscritto, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data BARI, 19/12/2019 Firma Vincenzo Totaro

D.R.R.S.	POLITECNICO DI BARI	06
Doctor of Philosophy in Risk and Environmental, Territorial and Building Development		2019
Coordinator: Prof. Michele Mossa		
XXXII CYCLE ICAR/02 – Hydraulic and marine constructions and hydrology		
DICATECh Department of Civil, Environmental, Land, Building Engineering and Chemistry		
	On the use of TCEV and Kappa four-parameter distributions for at-site flood frequency analysis	
	Supervisors: Prof. Vito Iacobellis Department of Civil, Environmental, Building Engineering and Chemistry (DICATECh) Polytechnic University of Bari Prof. George Kuczera Faculty of Engineering and Built Environment The University of Newcastle, Australia	
	Ph.D. Candidate: Vincenzo Totaro	

D.R.R.S.	POLITECNICO DI BARI	06
Dottorato di Ricerca in Rischio, Sviluppo ambientale, territoriale ed edilizio		2019
Coordinatore: Prof. Michele Mossa		
XXXII CICLO ICAR/02 – Costruzioni idrauliche, marittime ed idrologia		
DICATECh Dipartimento di Ingegnerie Civile, Ambientale, del Territorio, Edile e di Chimica		
	Sull'uso delle distribuzioni a quattro parametri TCEV e Kappa nell'analisi di frequenza delle piene a scala locale	
	Supervisors: Prof. Vito Iacobellis Dipartimento di Ingegnerie Civile, Ambientale, del Territorio, Edile e di Chimica (DICATECh) Politecnico di Bari Prof. George Kuczera Faculty of Engineering and Built Environment The University of Newcastle, Australia	
	Dottorando: Vincenzo Totaro	

On the use of TCEV and Kappa four-parameter distributions for at-site flood frequency analysis

EXTENDED ABSTRACT *(eng)*

Main task of Flood Frequency Analysis (FFA) is the estimation of a design flood for a given site fitting a probability distribution to a record of peak flows. This allows to compute parameters and quantiles estimates, achievable with different approaches (e.g. frequentist and Bayesian methodologies). Although appears a conceptually simple procedure, several approaches were proposed for its implementation. As a consequence, a wide debate started up in hydrology on merits and limits of each method, making FFA an attractive topic for scientists. Furthermore, regardless of strategies adopted for achieving parameters and quantiles estimates, as well as the choice of the best fitting model, the correct evaluation of uncertainty in flood frequency estimates should be considered an important step for undertaking an aware decision strategy. At-site flood frequency analysis is one of more direct methods for making inference from data. Several probability distributions are traditionally fitted for this type of analysis, such as Gumbel, log-Normal, Generalized Extreme Value (GEV) and log-Pearson type 3 (LP3). However, these distributions are typically characterized by two or three parameters, leaving the use of distributions with more parameters only for regional applications. This is the case of Two Components Extreme Value (TCEV) and Kappa distributions, which belong to the class of four-parameter distributions. These distributions are characterized by a distinct theoretical background, which reflects on their analytical properties. With respect to at-site analysis, while some investigations for properties of Kappa were conducted, similar studies for TCEV were neglected, probably due to the supposed relevant degree of uncertainty that should affect related estimates.

In this thesis the applicability of a Bayesian approach for at-site estimation of parameters and quantiles of TCEV and Kappa distribution is tested. In particular,

in order to achieve a complete vision about this topic, the theoretical background of extreme value distributions was revisited, with a focus on their role in interpreting floods phenomenology.
One of the main contributions of this work can be considered the development of a Bayesian procedure for providing inferential conclusions about Kappa and TCEV, with an explicit quantification on connected uncertainty and, in the case of this latter distribution, a new measure for discerning the presence of two different populations into a sample was introduced.
As case study, Eastern and Northern Australia was selected, due to high variability of climate and floods regime in this area. In this way, several underlying mechanisms of floods generation are expected to be analyzed, and abilities of TCEV and Kappa in fitting gauged data was tested. Results of application showed that in most cases, for sites located below the latitude of 23° south, TCEV distribution provide an excellent fit to at-site data, when compared to LP3 and GEV distributions. Furthermore, better performances were detected also in terms of uncertainty for high quantiles. Results are explainable with the climate-driven floods regime that affect the region.
Finally, a single Italian case study was investigated, on the basis of several studies that documented two different mechanisms of runoff generation: as expected, in accordance with its theoretical formulation, TCEV provided a relevant ability in fitting at-site data.

key words

flood frequency analysis; at-site; TCEV; Kappa; Bayesian analysis; Australia

Sull'uso delle distribuzioni a quattro parametri TCEV e Kappa nell'analisi di frequenza delle piene a scala locale

EXTENDED ABSTRACT (ita)

Una delle principali funzioni dell'analisi di frequenza delle piene è la stima di una piena di progetto per un determinato sito applicando una certa distribuzione di probabilità ad una serie storica di eventi di piena. Questo al fine di permettere la stima dei parametri di tale distribuzione e dei relativi quantili di progetto, ottenibili con diversi approcci (e.g. metodologie frequentiste o Bayesiane). Benchè si tratti di una procedura concettualmente semplice, per la sua implementazione sono stati proposti diversi approcci. Ciò ha condotto ad un ampio dibattito nella comunità idrologica sui meriti ed i relativi limiti di ciascun metodo, rendendo così l'analisi di frequenza delle piene un argomento molto dibattuto. Inoltre, in merito alla strategia impiegata per ottenere le stime di parametri e quantili, come anche la scelta della distribuzione che meglio si adatta ai dati osservati, la corretta valutazione dell'incertezza di tali stime dovrebbe essere considerata come una fase importante per intraprendere una consapevole strategia decisionale.

L'analisi di frequenza delle piene è uno dei più diretti metodi che permette l'inferenza sui dati osservati. In quest'ambito, vengono tradizionalmente applicate diverse distribuzioni di probabilità, come la Gumbel, la Log-Normale, la Generalized Extreme Value (GEV) e la Log-Pearson di tipo 3 (LP3). Si tratta, in ogni caso, di leggi caratterizzate da due o tre parametri, riservando l'uso di distribuzioni che ne prevedono un numero maggiore solo per applicazioni di carattere regionale. È questo il caso delle distribuzioni TCEV (Two Components Extreme Value) e Kappa, che appartengono alla classe delle distribuzioni a quattro parametri. Queste distribuzioni sono caratterizzate da formulazioni teoriche diverse, che si riflettono sulle loro proprietà analitiche. Con riferimento

alle analisi a scala locale, mentre per la distribuzione Kappa sono presenti diversi studi che ne documentano le proprietà, tali investigazioni sono state trascurate per la distribuzione TCEV, molto probabilmente a causa del supposto grado di incertezza associato alle relative stime.

In questa tesi di è verificata l'applicabilità di un approccio Bayesiano per la stima di parametri e quantili per le distribuzioni TCEV e Kappa. In particolare, per ottenere un quadro completo di questo argomento, si è realizzata una rivisitazione del contesto teorico in cui queste distribuzioni sono inserite, facendo particolare attenzione sulle rispettive abilità nell'interpretare la fenomenologia delle piene.

Uno dei principali contributi di questo lavoro di tesi può essere considerato lo sviluppo di una procedura Bayesiana per realizzare analisi di natura inferenziale sulle distribuzioni TCEV e Kappa, con una quantificazione esplicita della relativa incertezza. Inoltre, nel caso della TCEV, è stata introdotta una nuova misura per distinguere la presenza delle due componenti in un campione.

L'Australia orientale e settentrionale è stata selezionata come caso di studio, motivando l'opportunità di questa scelta con l'elevata variabilità del clima e del regime delle piene che si verifica in quest'area. In questa maniera, si sono potuti confrontare diversi meccanismi di generazione delle piene, verificando l'abilità delle distribuzioni TCEV e Kappa nell'interpretare i dati delle piene. In termini di risultati si è osservato che, in molti casi, per siti posizionati al di sotto della latitudine 23° sud, la TCEV presenta un eccellente adattamento sulla scala locale, rispetto alla LP3 ed alla GEV. Inoltre, performances migliori sono state notate anche in termini di incertezza per quantili elevati. Questi risultati sono spiegabili con l'influenza del clima sul regime delle piene nella regione così determinata.

Infine, è stato analizzato un singolo caso studio italiano, scelto sulla base di studi che documentano la presenza di due differenti meccanismi di generazione dei deflussi: come atteso, in accordo con la sua formulazione teorica, la TCEV ha un'elevata affidabilità nell'interpretare i dati osservati.

key words

analisi di frequenza delle piene; scala locale; TCEV; Kappa; analisi Bayesiana; Australia

INDEX

CHAPTER 1 - INTRODUCTION

1.1 - Overview

The topic of Flood Frequency Analysis (FFA) is usually introduced reporting that its goal is to estimate a design flood for a given site or location (Stedinger et al., 1993; Laio et al., 2009; Castellarin et al., 2012). This can be obtained by fitting a probability distribution to a series of recorded maxima of peak flows and estimating their parameters and quantiles (Cunnane, 1989). At the expense of this apparent simplicity, each of these steps involves number of questions and problems, whose discussion is crucial for a proper interpretation of results, aimed at providing the practitioner of affordable design rules. Furthermore, regardless of strategies adopted for achieving parameters and quantiles estimates, as well as the choice of the best fitting model, the correct evaluation of uncertainty in flood frequency estimates is a necessary condition for undertake an aware decision strategy (Merz and Thieken, 2005; Parkes and Demeritt, 2016).

Quae cum ita sint, is fully comprehensible how flood frequency analysis has always been an attractive topic for hydrologists, subject of numerous studies and reports (e.g. Cunnane, 1985, 1989; Bobée and Rasmussen, 1995; Hosking and Wallis, 1997; Ramachandra Rao and Hamed, 2000; Singh and Strupczewski, 2002).

Different probability distributions are classically proposed for flood frequency analysis: the most used can be considered log-Normal, Gumbel, Generalized Pareto (GP), Generalized Extreme Value (GEV) and log-Pearson type 3 (LP3). Typically, they rely in the space of two (log-Normal, Gumbel) or three (GEV, LP3) parameters. Use of four- or five-parameter distributions is typically diffused in Regional Flood Frequency Analysis (RFFA; Cunnane, 1989). Such in the case of two famous distributions, the Two Component Extreme Value (TCEV; Rossi et al., 1984) and the Kappa distributions (Hosking, 1994). While Kappa distribution is widely employed in the framework of L-Moments RFFA (Hosking and Wallis,

1997), TCEV was employed in the same framework almost exclusively in Italy (Castellarin et al., 2012). Furthermore, it should be mentioned that some multi-parameter distributions (such as TCEV and Wakeby; Cunnane, 1989) are able to justify the condition of separation introduced by Matalas et al. (1975).

Moreover, as is the TCEV case, the use of a four-parameter distribution can be theoretically justified by the presence of different sub-populations. As noted by Cunnane (1985), floods can be generated by some different contributions, such as frontal and cyclonic rains, hurricanes and snowmelt. In the field of theoretically derived distributions, in the spite of the original work of Rossi et al., (1984), Gioia et al. (2008) exploited the use of a two-component threshold driven mechanisms of runoff generation, which justify the presence of ordinary (with frequent occurrences) and rare flood events. A such made distribution, furthermore, is able to interpret the presence of high outliers in a time series, which can affect the output of inferential procedures with two- or three-parameter distributions.

As previously reported, Kappa distribution is widely employed in RFFA with L-moments. Especially, an extensive use of this distribution is made during the evaluation of the heterogeneity measure proposed by Hosking and Wallis (1993). As the matter of the fact, at-site uses of Kappa and TCEV was investigated only by some authors. For example, Parida (1999) proposed an at-site use of Kappa distribution for modelling Indian summer monsoon rainfall using L-Moments Estimation (LME). A comparison between statistical properties of Maximum Likelihood (MLE) and LME was performed by Winchester (2000) and Dupuis and Winchester (2001). At-site use of TCEV distribution, instead, is very difficult to be found. A technique based on the use of a Newton-Raphson algorithm for obtaining MLE of the four parameters is reported in Rossi et al. (1984). However, the same authors reported that "when the four parameters of the TCEV distribution are estimated from a single AFS[1], the uncertainty is great, particularly as regards the parameters of the outlying component. The uncertainty becomes

[1] Annual Flood Series

extremely high for AFS's without outliers…". This statement is very important for this thesis, because explicitly refers to the concept of uncertainty. Another use of an at-site TCEV can be found in Connell and Pearson (2001), that applied to annual maximum floods for the Canterbury region in New Zealand using least squared method, finding good results for several sites.

Interest in four-parameter distribution can arise also by the necessity of looking for different distribution able to model changes in modifying environment.

For example, perception of living in a changing environment, and the need of mitigate negative impacts on people and assets, led European Commission to enact the so-called "Flood directive" 2007/60/CE, that bound member countries to promulgate acts finalized to the reduction of flood risk. One of the most important statements in this directive, is "… some human activities (such as increasing human settlements and economic assets in floodplains and the reduction of the natural water retention by land use) and climate change contribute to an increase in the likelihood and adverse impacts of flood events". It is important to note that there is an explicit reference to climate change in relationship with a consequential increasing in flood risk. If, from one side, is symptomatic of an awareness of the importance of managing and reduction of flood risk, from the other side hands the responsibility to the practitioner to understand links between climate variables and flows regime. This is a very difficult task, because of natural variability can mask anthropogenically inducted changes in flooding (McCabe and Wolock, 1997). Furthermore, flood regime can be also influenced on multidecadal time scales, linked with climatic phenomena such as El Nino Southern Oscillation (ENSO). This is the case of New South Wales (Australia), where evidences of this phenomena were documented by Erskine and Warner (1988), Franks and Kuczera (2002), Micevski et al. (2006). A practical implication of this climate driven floods regime was documented by Franks and Kuczera (2002), which found inconsistent the hypothesis that annual maximum peak flows could be retained identically distributed.

In this framework, potential of TCEV and Kappa for at-site frequency analysis deserves a proper assessment. Moreover, the need of giving ad adequate quantitative estimation of the related uncertainty in parameters and quantiles estimates make the Bayesian theory the most suitable method for performing a flood frequency analysis. Such investigations motivated this study.

In this spite, first goal of this thesis is to provide a common theoretical framework able to describe the role of TCEV and Kappa in the world of flood frequency distributions; this can be a first and helpful tool for interpreting results of analysis.
Secondly, a unique description of all the properties of these two distributions into the frequentist theory is essential. The need of these discussion arises mostly for TCEV distribution, where L-Moments theory was not described in detail in last decades, also because analytical complexity that arise in the LME procedure. This is reflected in the neglection of TCEV distribution in the L-Moments ratio diagram, that is here illustrated in a complete version.
Implementation of an appropriate procedure able to model TCEV and Kappa in a Bayesian analysis is another cornerstone of this study. Evaluation of uncertainty is an important step in a flood frequency procedure, and its relevance will be decisive when two or more distributions will have similar fits to observed data.
Lastly, application of this new tool for an at-site flood frequency analysis is the crucial point in this thesis. Analysis of the goodness of fit of TCEV and Kappa to a real dataset will be a helpful step for understanding the increase in the degree of knowledge from an at-site data that is possible to obtain.

1.2 – Outline

With the purpose of describing logical steps between the state of the art and the new approach proposed, thesis is articulated with the following structure.

In chapter 2 Extreme Value Theory (EVT) is illustrated in respect of both its asymptotic and exact developments. Both frameworks are described in their main features, in order to highlight properties of the traditional Extreme Values Distributions (EVD). In particular, special attention is given to the exact formulation of theoretically probability distributions of floods, with a detailed description of the theoretical framework, formulated firstly by Zelenhasic (1969) and Todorovic and Zelenhasic (1970). This is the fundamental background for understanding properties and physical nature of Two Component Extreme Value (TCEV) and Kappa distributions. The opportunity of reporting these concepts will be evident during the discussion of results, where the physical nature of floods phenomena has to be taken into account.

In chapter 3 a summary on the frequentist approach to statistical inference for EVDs is provided. Goal of this chapter is to illustrate the main concepts of this field, in order to provide a basis for framing the literature review of TCEV and Kappa.

This review is reported in chapter 4, where a full description of all the frequentist approach to these two distributions is illustrated. The wide use of Kappa distribution for Regional Flood Frequency Analysis (RFFA) led to a moderate diffusion of its properties for at-site analysis, also because of the availability of the code provided by Hosking (2000) for the estimation of L-Moments. On the contrary, TCEV distribution has been used for FFA only in Italy and Spain into a regional framework. Properties of at-site estimation were investigated only in some cases (Rossi et al., 1984; Connell and Pearson, 2001). A lack of insights is missing on the L-moments estimation, due to its complexity. An investigation on this topic is reported, along with the plots of this distribution on L-Moments Ratio Diagram (LMRD).

The problem of parameters and quantiles estimation coupled with a quantification of their uncertainty is investigated with a Bayesian approach, illustrated in Chapter 5. This inferential framework is described in its main aspects, together with the explanation of its practical implementation. Finally, a

new measure based on the sampling of the posterior distribution, TC_{ratio}, is introduced, with the aim to provide a graphical tool for a discrimination of the presence of two different populations into a single series.

In Chapter 6 an application to at-site flood frequency analysis using TCEV and Kappa is proposed for eastern and northern Australia. Performances of these distributions are investigated and compared to those provided by traditional distributions. The role of uncertainty for discriminating between candidate distributions is investigated.

A summary of conclusions and a proposal for some future research directions is provided in Chapter 7.

CHAPTER 2 – AN OVERVIEW ON EXTREME VALUE THEORY

2.1 – Introduction

Consistently with the scope of this thesis, in this chapter the main outlines of Extreme Value Theory (EVT) will be illustrated. The goal is the description of a common background in which framing the two four-parameter distributions analyzed in this work, TCEV and Kappa. Indeed, when fitting these distributions to real data, is fundamental to understand their theoretical basis for interpreting results.

In Par. 2.2, an historical overview on the use of extremes in hydrology is illustrated, while in Par. 2.3 a short introduction to FFA is reported. Then, in Par. 2.4, the asymptotic approach to extreme value theory is described and in Par. 2.5 will be discussed in detail the theoretical approach for deriving probability distributions of floods. Focus on this latter model is justified by its specificity, that can allow to a complete description of underlying processes. As will be shown in Chapter 6, this theory can be very useful also for discriminating between distributions.

Finally, Annual Maximum (AM) and Peak-Over-Threshold (POT) approaches will be compared with a focus in their main characters.

2.2 – Some historical notes on Extreme Value Theory

What is the origin of the term “extreme”? Stopping at the Roman period, this term comes from the Latin *extremus*, superlative of the adjective *extĕr*, which means «that stands out».

The importance of dealing with extremes has been recognized since the fourth century B. C., the age of the Greek philosopher Plato. A trace can be found in Phaedo 90α:

ὥσπερ, ἦ δ᾽ ὅς, περὶ τῶν σφόδρα σμικρῶν καὶ μεγάλων: οἴει τι σπανιώτερον εἶναι ἢ σφόδρα μέγαν ἢ σφόδρα σμικρὸν ἐξευρεῖν ἄνθρωπον ἢ κύνα ἢ ἄλλο ὁτιοῦν; ἢ αὖ ταχὺν ἢ βραδὺν ἢ αἰσχρὸν ἢ καλὸν ἢ λευκὸν ἢ μέλανα; ἢ οὐχὶ ᾔσθησαι ὅτι πάντων τῶν τοιούτων τὰ μὲν ἄκρα τῶν ἐσχάτων σπάνια καὶ ὀλίγα, τὰ δὲ μεταξὺ ἄφθονα καὶ πολλά;

Just what is true of extremely large and extremely small things, he replied. What is rarer than to find a man, or a dog, or anything else which is either extremely large or extremely small? Or again, what is rarer than to find a man who is extremely swift or slow, or extremely base or honourable, or extremely black or white? Have you not noticed that in all these cases the extremes are rare and few, and that the average specimens are abundant and many?

It is very interesting remarking that Plato used the two words *extremes* (*ἄκρα*) and *average* (*μεταξὺ*). This can be considered a primitive description of a statistical distribution, which highlight how since ancient times the analysis of exceptional events was recognized as crucial. A strong link between extremes and social life can be traced as a consequence of the decisioning process on the location of urban and rural settlements, with regard to the availability of water resources. Although since the Bronze Age is documented the presence of technological urban hydraulic systems for water supply (Mays et al. 2007), rivers have always considered as a preferential component for civil settlements. But, if living in flood-prone areas from one hand means having flat and fertile soils (Tingsanchali, 2012), from the other hand exposes people to flood risk. An exception to this urban planning strategy can be considered the ancient Greece, where a preventive approach for managing flood risk was applied, and cities were mostly built far from water courses (Koutsoyiannis et al., 2012). The unavoidable need of people to coping with flood risk, led them to find some measures for its mitigation: for example, Alyzia dam (Western Greece) was built reasonably during sixth century BC probably for flood mitigation (Koustoyiannis et al., 2008).

This short *excursus* has been reported just for illustrating to the reader how analysis of extremes has always been an important issue during centuries.

Quae cum ita sint, is evident how the necessity of finding some theoretical tools able to manage with water-linked calamities has been a key question in engineering. The necessity of monitoring hydrological variables followed from this problem, and the contemporary availability of recorded measures and new statistical tools pushed the attention to extreme value analysis.

Framed in a historical context, the statistics of extreme values represents a relatively modern field of research. Describing its history is not only a literary exercise, but can help us in a better comprehension of reasons and developments of this theory.

As reported in Gumbel (1958), the first example of analysis of extreme values can be traced back to Nicolaus Bernoulli[2] that, in 1709 in Basel, published his work *Specimina Artis Conjectandi, ad quaestiones Juris applicatae*. Among the questions addressed in this opera, there is that of considering the age of the last survivor of a certain number of men of the same age who die in a certain number of years. Todhunter (1865) reported how from this essay it is not possible to appreciate the mathematical power of its author, but still manages to emerge his faith in the theory of probability.

The importance of dealing with extreme values emerged subsequently in astronomy, with the need of finding objective criteria for accepting or rejecting suspect measures (such as those that may concern diameters or distances between stars). The research in this topic, consistent with the nature of the problem, was carried out within the gaussian distribution framework. Literature in this sense was produced by Peirce (1852), Chauvenet (1868) and Rider (1933).

[2] Nicolaus Bernoulli (1687-1759) was the nephew of Jacob (1654-1705, who introduced the famous distribution of B.) and of Johann (1667-1748, who contributed to the development of differential calculus in Europe), and cousin of Daniel (1700-1782) known for the fluid dynamics theorem bearing his surname.

Kotz and Nadarajah (2000) reported that Von Bortkiewicz (1922) introduced the concept of extreme value distribution for the first time. In the following years the developments in this field were linked exclusively to the normal distribution, and the first use of a non-Gaussian distribution is due to Fréchet (1927), who was the first to derive an asymptotic distribution of extreme values. Moreover, in this work the *stability postulate* was introduced, according to which the distribution of the extreme value is equal to the initial one, except for a linear transformation. It is important to note that Gumbel (1958) recognized how this work was underestimated, so much so that it was published in a second-order journal.

Fisher and Tippett (1928) showed how, using the same stability postulate, limit distributions can be of three types. Von Mises (1936) formulated the sufficient conditions for the validity of these three asymptotic distributions. Gnedenko (1943) completed this work by formulating the necessary and sufficient conditions. Gumbel (1941) applied Fisher and Tippett theory to Annual Maximum (AM) values under the hypothesis that maximum daily flow in one year should be distributed as an extreme value variate (Adams and McMahon, 1985). This work of Gumbel *("The return period of flood flows")* is fundamental, because in can be retained as the first explicit reference to the application of statistic of extremes to hydrology (Katz et al., 2002). In 1955 Jenkinson provided a complete interpretation of the Asymptotic distributions formulating the Generalized Extreme Value (GEV) distribution, while Gumbel's first fundamental text on the theory of extreme values dates back to 1958. Focusing the discussion only in the hydrological framework, is has to be remarked that Todorovic and Zelenhasic (1970) introduced the Peak-Over-Threshold (POT) model in hydrology, showing how results similar to those of the asymptotic theory of extreme values can be obtained. Picklands (1975) proposed the use of above threshold models and introduced the Generalized Pareto (GP) distribution. Landwehr et al. (1979) introduced Probability Weighted Moments (PWM) and in 1990 Hosking developed their linear combinations, L-Moments. In the perspective of improving the distributions of extreme values in the adaptation to physical phenomena,

Rossi et al. (1984) introduced TCEV distribution, which had the remarkable merit of contemplating double-component processes.
This is only a brief historical overview on the extreme value theory, which goal is to illustrate how dynamic and relatively young is this science. Its role is remarkable also in other fields, such as ecology (Katz et al., 2005) and climatology (Buishand, 1989).

2.3 – An overview on flood frequency analysis (FFA)

Introductory chapters of papers and books about Flood Frequency Analysis (FFA) report, even if with different sentences, that its goal is to estimate a biunivocal relationship between flood magnitude and a certain probability of occurrence (Stedinger et al., 1993; Laio et al., 2009; Ramachandra Rao and Hamed, 2000). This inferential procedure requires the analysis of an adequate flow record, from which formulate basic hypothesis and derive all the information needed.
For this purpose, three main models can be used (Cunnane, 1989):

a) Time Series (TS): represents the flow hydrograph by a series of ordinates at equally spaced time intervals (one day for flood frequency analysis) called a *time series*;
b) *Block Maxima (BM)*: this method provides the division of the record into a series of equal nonoverlapping periods, taking into account only their maximum value (Ferreira and de Haan, 2015). If the interval is one year, then the method is called *Annual Maximum (AM)*;
c) Peak Over Threshold (POT): in this approach only the values that exceed a certain threshold are selected.

AM approach can be regarded as a particular case of TS modelling, where the unit of time is one year and the flow representing that time is the highest flow during the year (Adams and McMahon, 1985).

This thesis will deal only with AM approach; however, POT will be discussed because its formulation contains important concepts that can be helpful for analyze and comprehend results. TS method is here cited only for completeness. Traditional FFA can be modeled in 3 main steps:

1. Choice of the model type, between a), b) or c);
2. Choice of the probability distribution;
3. Definition of the method for parameter and quantile estimation.

Each of those three points requires particular care in FFA modeling. An adequate analysis of floods frequency cannot disregard a critical analysis of these steps.

2.4 – Asymptotic theory of extremes

The largest of a number of independent and identically distributed random variables can be described following both an asymptotic and an exact theoretical derivation (Koutsoyiannis, 2003). Due to their simplicity and theoretical basis, asymptotic distributions are widely used in hydrological applications. However, because of the conceptual adequacy of theoretical approach to extremes to investigate flood dynamics, this topic will be illustrated in detail in paragraph 2.5.

2.4.1 – Annual Maximum models

Denote by $X_1, X_2, \ldots, X_n$ a sequence of independent random variables having a common distribution function *F*. Then, variables $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are *order statistic*[3] related to given sample. In the following the statistical behavior of this quantity will be studied:

[3] It is possible to refer to orders statistic using also this notation: $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$.

$$M_n = max\{X_1, \dots, X_n\} = X_{n:n} \tag{2.1}$$

Eq. (2.1) represents the maximum over *n* time units of observations; then, if *n* is equal to the number of observations in one year, then M_n is the annual maximum.

The independence hypothesis about variable X_i allows to define the probability distribution of the maximum as

$$H_n(x) = Pr\{M_n \leq x\} = Pr\{X_{1:n} \leq x, \dots, X_{n:n} \leq x\} = Pr\{X_{1:n} \leq x\} \cdot \dots \cdot Pr\{X_{n:n} \leq x\} = \{F(x)\}^n$$

However, in the general case is F is unknown and is necessary to study the asymptotic behavior of F^n, for *n* that tends to infinity.

According to Castillo (2012), is important the investigation of the behavior of the function *F* in its domain borders. Defining

$$\alpha(F) = inf\{x: F(x) > 0\}$$

as the lower end point of $F(x)$, and

$$\omega(F) = sup\{x: F(x) < 1\}$$

as the upper end point, for *n* that goes to infinity is

$$\lim_{n\to\infty} F^n(x) = \begin{cases} 1 & if\ F(x) = 1 \\ 0 & if\ F(x) < 1 \end{cases}$$

This is symptomatic of a degeneration of the limit distribution in the extremes. According to a classical approach in statistic, this problem is usually faced verifying the existence of two successions of variables $\{a_n\}$ and $\{b_n > 0\}$, $n \in \mathbb{N}$, such that:

$$G(x) = \lim_{n\to\infty} F^n(a_n + b_n x), \qquad x \in \mathbb{R} \tag{2.2}$$

i.e., this is equivalent of the study of the asymptotic behavior of the rescaled variable M_n:

$$M_n^* = \frac{M_n - a_n}{b_n} \tag{2.3}$$

F is said to be concerned to the domain of attraction for the maxima of G if satisfies (2.2) for good $\{a_n\}$ and $\{b_n > 0\}$.

In this framework, the following key theorem was introduced.

Fisher and Tippett Theorem: if exist series $\{a_n\}$ and $\{b_n > 0\}$ such that is, for $z \in \mathbb{R}$,

$$\lim_{n\to\infty} Pr\left\{\frac{M_n - a_n}{b_n} \le z\right\} = G(z) \tag{2.4}$$

Where *G* is a non-degenerate function. Then, *G* belongs to one of the following families of extreme value for maxima:

$$G(z) = exp\left[-exp\left(-\tfrac{z-a}{b}\right)\right] \qquad -\infty < z < +\infty$$

$$G(z) = \begin{cases} 0 & z \le a \\ exp\left[-\left(\frac{z-a}{b}\right)^{-\gamma}\right] & z > a \end{cases}$$

$$G(z) = \begin{cases} exp\left\{-\left[\left(-\frac{z-a}{b}\right)^{\gamma}\right]\right\} & z < a \\ 1 & z \ge a \end{cases}$$

□

Jenkinson (1955) combined these distributions into the Generalized Extreme Value (GEV) distribution, that can be expressed using the following form:

$$F(x) = \begin{cases} exp\left\{-\left[1-\varepsilon\left(\frac{x-\zeta}{\sigma}\right)\right]^{1/\varepsilon}\right\} & \varepsilon \neq 0 \\ exp\left\{-exp\left[-\left(\frac{x-\zeta}{\sigma}\right)\right]\right\} & \varepsilon = 0 \end{cases} \quad (2.5)$$

where $\sigma > 0$ and $1-\varepsilon\left(\frac{x-\zeta}{\sigma}\right) > 0$. Shape parameter ε, instead, is defined in the whole real domain. However, from a practical perspective it can be expected that ε can varies into a small neighbourhood of 0. For example, Papalexiou and Koutsoyiannis (2013) analyzed a wide dataset of rainfall time series from all over the world and found that ε can lie in the interval $(-0.23,0)$.

As noted in the definition of GEV, this distribution family comprehend distributions of Gumbel, Fréchet and Weibull, that now can be written in their definitive form according relative values of ε:

- $\varepsilon = 0$ - Gumbel distribution. This distribution is characterized only by location and scale parameters. It is an unbounded distribution, whose CDF can take the form:

$$F(x) = exp\left\{-exp\left[-\left(\frac{x-\zeta}{\sigma}\right)\right]\right\} \quad (2.6)$$

- $\varepsilon < 0$ - Fréchet distribution. Taking $y = 1-\varepsilon\left(\frac{x-\zeta}{\sigma}\right)$, this distribution assumes the following expression:

$$\begin{cases} e^{(-y)^{1/\varepsilon}} & y > 0 \\ 0 & y \leq 0 \end{cases} \quad (2.7)$$

and $x \in \left(\frac{x-\zeta}{\sigma}, +\infty\right)$, i. e. is bounded on the left.

– $\varepsilon < 0$ – Weibull distribution: in this last case, taking $y = \left[1 - \varepsilon\left(\frac{x-\zeta}{\sigma}\right)\right]$, this CDF can be written as

$$\begin{cases} e^{-(-y)^{-1/\varepsilon}} & y > 0 \\ 0 & y \leq 0 \end{cases} \quad (2.8)$$

Domain if Weibull distribution is $x \in \left(-\infty, \frac{x-\zeta}{\sigma}\right)$ and is right-bounded.

2.4.1.1 – Pearson and Log-Pearson type 3 distributions

Pearson distributions can be considered as a family of continuous distributions which generalized the normal distribution and was introduced in 1895 by the mathematician Karl Pearson. In particular, Pearson type 3 based distributions play an important role in hydrological applications.
Pearson type 3 (P3) density function has the following expression:

$$f(x) = |\beta|[\beta(x-\zeta)]^{\alpha-1}\frac{exp[-\beta(x-\zeta)]}{\Gamma(\alpha)} \quad (2.9)$$

where α, β, ζ are its parameters under the condition that $\alpha > 0$. Depending on the values assumed by β, two cases have to be discussed: if $\beta > 0$, then $\zeta \leq x < +\infty$ and distribution is positively skewed; if $\beta > 0$, then skewness is negative and $-\infty < x \leq \zeta$.

Special case is when $\zeta = 0$: in this case Gamma distribution is obtained.
Log-Pearson type 3 (LP3) distribution can be obtained from Pearson type 3 distribution taking (Bobée, 1975):

$$y = \log_a x$$

where the base *a*, here reported as general, is usually set equal to *10* or to *e*. The pdf of LP3 will take the form:

$$f(x) = |\beta|[\beta(\log_a x - \zeta)]^{\alpha-1} \frac{exp[-\beta(\log_a x - \zeta)]}{\mathrm{x}\Gamma(\alpha)} k \qquad (2.10)$$

where *k* is a constant dependent on the base *a* (assumed greater than 0) so that

$$k = \log_a e$$

As in (2.9) α, β, ζ are its parameters under the condition that $\alpha > 0$ and for $\alpha > 0$ (positive skew) $exp(\zeta) < x < \infty$, while for $a < 0$ (negative skew) $0 < x < exp(\zeta)$.
This a really important distribution, also because its use is recommended U.S. federal agencies for flood frequency analyses as described by Bulletin 17B (IACWD, 1982). A wide review on LP3 distribution (and parameters estimation techniques) can be found in (Bobée, 1975; Phine and Hira, 1983; Arora and Singh, 1989; Griffis and Stedinger, 2007a, 2007b).

2.4.2 – Peak-Over-Threshold Approach

As described paragraph 2.3, Partial Duration Series (or Peak Over Threshold) approach is an alternative model widely used in hydrology for modelling distributions of extreme values.

As usual, let consider a sequence of independent and identically distributed random variables $(X_1, X_2, \ldots, X_n)$ sampled from a marginal distribution *F*. In this case extreme events are considered those who exceed a high threshold q_0. The stochastic behavior of extreme events is then given by the conditional probability:

$$Pr\{X > q_0 + y | X > q_0\} = \frac{1 - F(q_0 + y)}{1 - F(q_0)} \qquad y > 0 \qquad (2.11)$$

However, usually *F* is unknown, requiring the approximation given by the next theorem (Balkema and de Haan, 1974; Picklands, 1975; Coles, 2001):

Pickands-Balkema-de Haan theorem: be $(X_1, X_2, \ldots, X_n)$ a sequence of random variables with common distribution *F*, and

$$M_n = max\{X_1, X_2, \ldots, X_n\}$$

Under the validity of Fisher-Tippett theorem, for large *n*

$$Pr\{M_n \leq z\} \approx G(z)$$

with $G(z)$ the GEV distribution. If the value of the threshold q_0 is large enough, the distribution function of $(X - q_0 | X > q_0)$ is approximately given by

$$1 - \left(1 - \varepsilon \frac{x - q_0}{\sigma^*}\right)^{1/\varepsilon} \qquad (2.12)$$

with $1 - \varepsilon \frac{x - q_0}{\sigma^*} > 0$.
The limit case that is for $\varepsilon \to 0$ lead to:

$$1 - exp\left(\frac{x - q_0}{\sigma^*}\right) \tag{2.13}$$

i.e. an exponential distribution with parameter $1/\sigma^*$. The family which comprehend both distributions (2.12) and (2.13) is called the *Generalized Pareto family*, and the famous *Generalized Pareto (GP) distribution* can be expressed as:

$$F(x|q_0) = \begin{cases} 1 - \left(1 - \varepsilon\frac{x - q_0}{\sigma^*}\right)^{1/\varepsilon} & \varepsilon \neq 0 \\ 1 - exp\left(\frac{x - q_0}{\sigma^*}\right) & \varepsilon = 0 \end{cases} \tag{2.14}$$

An important property of such made truncated GP distributions is their invariance respect of the choice of the threshold (Madsen et al., 1997).

2.5 – Theoretical probability distributions for flood peaks

Rasmussen and Rosbjerg (1991) reported that Partial Duration Series model was firstly adopted in hydrology by Shane and Lynn (1964) and Todorovic and Zelenhasic (1970). This approach could be illustrated in respect of the variable describing a streamflow hydrograph (fig. 2.1) or a precipitation hyetograph.

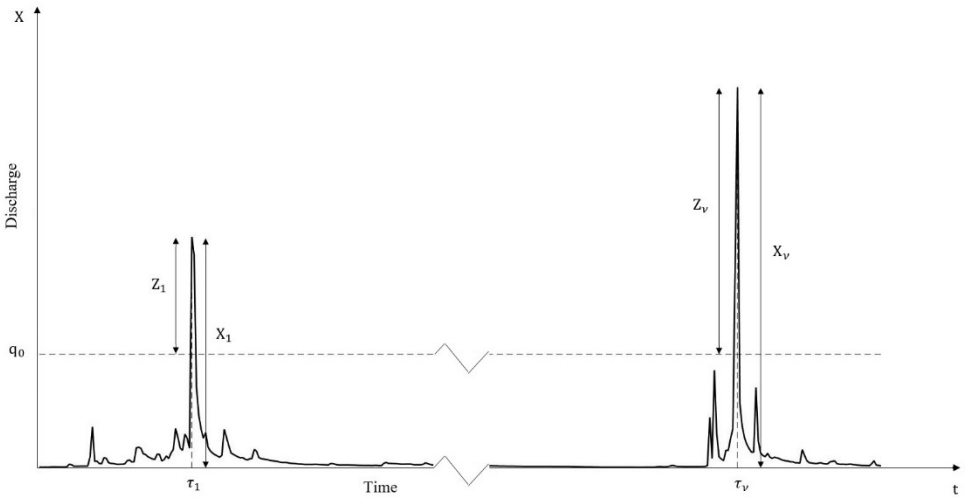

Figure 2.1 – streamflow hydrograph of instantaneous discharges

Peak Over Threshold method consists in analyzing only the exceedances over an assigned threshold level q_0 in a certain interval $[0, t]$. Denoting with ν the number of those exceedances happening at time τ_ν, exceedances over q_0 are:

$$Z_\nu = X_\nu - q_0 \tag{2.15}$$

With these previous statements, the number of exceedances in time interval $[0, t]$, their occurrence time and their magnitudes are random variables. This allows to deals with these events using the theory of stochastic processes. In this framework, a process that associate a time of occurrence with the magnitude of the relative observation is called *marked point process* (Snyder and Miller, 2012). A graphical representation of such defined process

$$\{\tau_i, Z_i;\ i = 1,2,\dots,\nu;\ \nu \in \mathbb{N}\}$$

is in fig. 2.2:

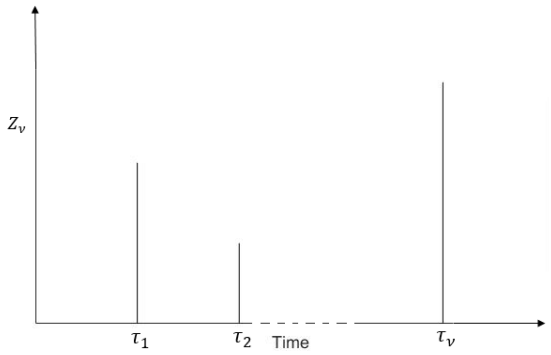

Figure 2.2 – A realization of the stochastic process of flood peak exceedances

Be $\eta(t)$ a discrete value process which gives the number of exceedances in $[0, t]$. It is straightforward that, for every $t \geq 0$ and $\Delta t > 0$, $\eta(t) \leq \eta(t + \Delta t)$.
A rigorous mathematical formulation on these processes requires the definition of the probability space Ω (Todorovic and Yevjevic, 1969) associated to an experiment:

$$(\Omega = \{\omega\}, \mathcal{A}, P)$$

where

- Ω, the *sample space*: is the set of all possible outcomes ω of the experiment. In the present case, the hydrograph of fig. 2.4 can be regarded as an outcome ω of random observations, while Ω is the set of all possible hydrographs, and this allows to give it the quality of a functional space. Realizations of the stochastic process Z_v are, finally, function defined on the space Ω;
- $\mathcal{A}$, a *σ-algebra* over Ω: is a set of subsets (i.e. events) of Ω and whose probability can be evaluated. It has the property to be a class closed under all countable set operations (Loeve, 1977)
- P, the *probability measure*: is a function with the property of associating a probability of an event defined on $\mathcal{A}$, i.e. $P: \mathcal{A} \rightarrow [0,1]$.

For example, a measurable subset belonging to the *σ-algebra* is that who comprehend all the events (and their functions) in which the number of events in $[0, t]$ is less than ν, i.e. $\eta_t < \nu$.

2.5.1 – Distribution of the number of exceedances

The search for the probabilistic interpretation of this kind of process requires the introduction of a particular subset of Ω, E_ν^t, which is characterized by having all the realizations of the stochastic process Z_t with only ν points in $[0, t]$.
This means that (Zelenhasic, 1970):

$$E_\nu^t = \{\eta_t = \nu\} \tag{2.16}$$

and

$$E_i^t \cap E_j^t = \emptyset \qquad \bigcup_{\nu=0}^{+\infty} E_\nu^t = \Omega \qquad \text{for all } i \neq j = 0,1,\ldots$$

Be $\Lambda(t)$ the expected value of η_t:

$$\Lambda(t) = \sum_{\nu=1}^{\infty} \nu P(E_\nu^t) \tag{2.17}$$

and

$$F_k(t) = P\{\tau(k) \leq t\} \tag{2.18}$$

with $\tau(k)$ time of occurrence of the *k-th* exceedance. According to Todorovic and Yevjevic (1969), because of (2.16)

$$E_k^t = \{\tau_k \leq t < \tau_{k+1}\} = \{\tau_k \leq t\} - \{\tau_{k+1} \leq t\} \tag{2.19}$$

Combining (2.18) and (2.19) follows that

$$P(E_k^t) = P\{\tau_k \leq t\} - P\{\tau_{k+1} \leq t\} = F_k(t) - F_{k+1}(t) \tag{2.20}$$

Finally, according to Zelenhasic (1970), it can be written that

$$F_k(t) = 1 - \sum_{j=0}^{k-1} P\left(E_j^t\right) = \sum_{j=k}^{\infty} P\left(E_j^t\right) \tag{2.21}$$

Define, from (2.16), E_k^t as the event that only k exceedances occur in $[0, t]$, and $E_1^{t,t+\Delta t}$ as the event that only one exceedance occurs in $[t, t + \Delta t]$, with $\Delta t > 0$ an assigned interval of time. In Zelenhasic (1970) can be ulteriorly found that probabilities $P(E_k^t)$, $k = 0,1,\ldots$ satisfy this set of differential equations:

$$\begin{cases} \dfrac{dP(E_k^t)}{dt} = \lambda_1(t, k-1)P(E_{k-1}^t) - \lambda_1(t,k)P(E_k^t) & k = 1,2,\ldots \\ \dfrac{dP(E_0^t)}{dt} = -\lambda_1(t,0)P(E_0^t) & k = 0 \end{cases} \tag{2.22}$$

where

$$\lambda_1(t,k) = \lim_{\Delta t \to 0} \frac{P\left(E_1^{t,t+\Delta t} \middle| E_k^t\right)}{\Delta t}$$

and

$$E_1^{t,t+\Delta t} = \{\eta(t + \Delta t) - \eta(t) = 1\}$$

It follows that:

$$\lambda_1(t) = \sum_{k=0}^{\infty} \lambda_1(t,k) P(E_k^t)$$

If $\lambda_1(t,k)$ is an integrable function for $k \in \mathbb{Z}_0^+$, and

$$\lim_{t\to 0} P(E_k^t) = \begin{cases} 0, & if\ k > 0 \\ 1, & if\ k = 0 \end{cases}$$

then solution of system (2.22) is:

$$\begin{aligned} P(E_k^t) = exp\left\{-\int_0^t \lambda_1(s,0)\,ds\right\} \cdot \int_0^t \lambda_1(t_1, k-1)\, exp\left\{\int_0^{t_1} \Delta\lambda_1(s,k) ds\right\} \cdot \\ \cdot \int_0^{t_1} \lambda_1(t_2, k-2)\, exp\left\{\int_0^{t_2} \Delta\lambda_1(s,k-1) ds\right\} \cdot \cdots \\ \cdot \int_0^{t_{k-2}} \lambda_1(t_{k-1}, 1)\, exp\left\{\int_0^{t_{k-1}} \Delta\lambda_1(s,2) ds\right\} \\ \cdot \int_0^{t_{k-1}} \lambda_1(t_k, 0) exp\left\{\int_0^{t_\nu} \Delta\lambda_1(s,1) ds\right\} dt_k dt_{k-1} \ldots dt_1 \end{aligned} \tag{2.23}$$

From (2.17), instead:

$$\Lambda(t) = \int_0^t \lambda_1(s) ds \tag{2.24}$$

Different cases can arise from the values that function $\lambda_1(t,k)$ can get. According to Bačová-Mitková and Onderka (2010) and Vukmirovic and Vukmirovic (2017), we can distinguish:

Case 1: $\lambda_1(t,k) = \lambda_1(t)$

$$P(E_k^t) = \left\{ \int_0^t \lambda(s)ds \right\}^k \frac{e^{-\int_0^t \lambda(s)ds}}{k!} \tag{2.25}$$

Eq. (2.25) expresses a time dependent Poisson process. Because of (2.24), (2.25) becomes:

$$P(E_k^t) = \frac{[\Lambda(t)]^k e^{-\Lambda(t)}}{k!} \tag{2.26}$$

$\Lambda(t)$ is the mean number of exceedances in $[0,t]$.

Case 2: $\lambda_1(t,k) = \lambda_1(t) \cdot \left(1 - \frac{k}{a}\right), a > 0$

$$P(E_k^t) = \frac{\Gamma(a+1)}{\Gamma(k+1)\Gamma(a+1-k)} e^{-\Lambda(t)} \left(e^{-\Lambda(t)/a} - 1\right)^k \tag{2.27}$$

This equation can be traced back to that of a binomial distribution for

$$e^{-\frac{\Lambda(t)}{a}} - 1 \to \frac{p}{1-p}$$

Case 3: $\lambda_1(t,k) = \lambda_1(t) \cdot \left(1 + \frac{k}{\Theta}\right), \Theta > 0$

$$P(E_k^t) = \frac{\Gamma(k+\Theta)}{\Gamma(k+1)\Gamma(\Theta)} e^{-\Lambda(t)} \left(e^{-\Lambda(t)/\Theta} - 1\right)^k \tag{2.28}$$

Analogously to (2.27), (2.28) takes the expression of a negative binomial distribution for

$$e^{-\frac{\Lambda(t)}{\Theta}} \to p$$

2.5.2 – Distribution of the largest peak

The other variable which requires to be analyzed is the largest exceedance in the time interval $[0, t]$, here denoted with $\chi(t)$. This is a random variable defined as:

$$\chi(t) = \sup_{\tau(v) \leq t} Z_v \tag{2.29}$$

The nature of $\chi(t)$ is such that, for every $\Delta t \geq 0$, is:

$$\chi(t) \leq \chi(t + \Delta t)$$

In this way, $\chi(t)$ is a non-decreasing function of t(fig. 2.3):

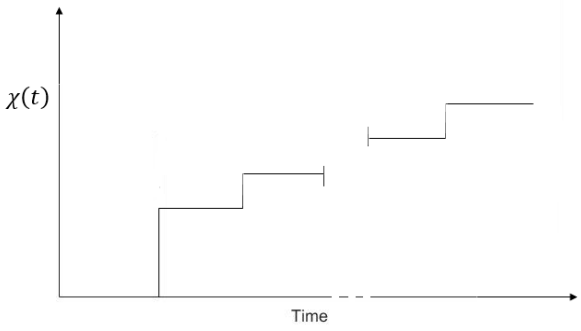

Fig. 2.3 – A sample function of the process $\chi(t)$

With continuity of notations, its CDF is defined as:

$$F_t(x) = P(\chi(t) \leq x)$$

for $t \geq 0$ and $x \geq 0$.

Todorovic (1970, theorems 1 and 3) computed the value of $F_t(x)$ as the mathematical expectation of

$$P\left\{ \sup_{\tau(\nu)\leq t} Z_\nu \leq x \middle| \eta(t) \right\}$$

i.e. (theorem 1):

$$F_t(x) = \sum_{k=0}^{\infty} P\left[\bigcap_{\nu=0}^{k} \{Z_\nu \leq x\} \cap E_k^t \right] \tag{2.30}$$

and (theorem 3):

$$F_t(x) = P(E_0^t) + \sum_{\nu=1}^{\infty} P(E_\nu^t) P\{\chi_\nu \leq x | E_\nu^t\} \tag{2.31}$$

Zelenhasic (1970) reported that the same result can be obtained considering that

$$F_t(x) = P(\chi(t) \leq x) = P\{[\chi(t) \leq x] \cap \Omega\}$$

which is equivalent to write[4]:

$$F_t(x) = P\left\{ [\chi(t) \leq x] \cap \left[\bigcup_{\nu=0}^{\infty} E_\nu^t \right] \right\} = \sum_{\nu=0}^{\infty} P\{[\chi(t) \leq x] \cap E_\nu^t\} \tag{2.32}$$

[4] In this deduction has been considered that events E_0^t, E_1^t… are exhaustive and mutually independent.

An important result for extreme value theory can be derived considering that exceedances Z_ν occurring in an interval $[0, t]$ are independent and identically distributed random variables (Rossi and Versace, 1982).
In particular, if process $\{Z_\nu;\ \nu = 1,2,\dots\}$ is composed by a sequence of random variables independent both mutually and from the process $\{\eta_t, t \geq 0\}$

$$P\{\chi_\nu \leq x | E_\nu^t\} = P\{\chi_\nu \leq x\} = \prod_{i=1}^{\nu} P[Z_i \leq x]$$

If process Z_i has identically distributed random variables with a common CDF $H(x)$, then:

$$\prod_{i=1}^{\nu} P[Z_i \leq z] = [H(x)]^\nu$$

The *intensity,* $\lambda(t)$*,* of process $\eta(t)$ is representative of the mean number of exceedances in the time unity. $\eta(t)$ is said a *homogeneous* process if λ is independent of time, and $\Lambda(t) = \lambda \cdot t$.
Process $\eta(t)$ assumes a Poissonian distribution for high thresholds because of the poissonian distribution of rainfalls (Todorovic and Yevjevic, 1969) and the distance in time between big floods, which guarantee their independence. Under this statement, process results to be seasonally nonhomogeneous (Todorovic, 1978). Concluding, $\eta(t)$ is here considered as a nonhomogeneous Poisson process, with a cyclical intensity function having a yearly period.

Defining with $H(x)$ the distribution of the exceedances in the same interval, this leads to (Todorovic and Zelehnasic, 1970):

$$F_t(x) = \sum_{\nu=0}^{\infty} \{[H(x)]^{\nu} \, P(E_{\nu}^{t})\} \tag{2.33}$$

Consistent with eq. 2.32:

$$F_t(x) = P(E_0^t) + \sum_{\nu=1}^{\infty} \{[H(x)]^{\nu} P(E_{\nu}^{t})\} \tag{2.34}$$

Substituting the expression on (2.26) into (2.34), will be:

$$F_t(x) = e^{-\Lambda(t)} \left\{ 1 + \sum_{\nu=1}^{\infty} \frac{[\Lambda(t) H(x)]^{\nu}}{\nu!} \right\}$$

Because the term into brackets is the Taylor expansion of $e^{[\Lambda(t)H(x)]}$, is (Todorovic and Zelehnasic, 1970):

$$F_t(x) = e^{\{-\Lambda(t)[1-H(x)]\}} \tag{2.35}$$

2.5.3 – Probability distributions of the largest exceedance in one year

Different solutions were proposed in literature for describing the distribution of the largest exceedance under the hypothesis of Poissonian distribution of the number of occurrences. In this chapter, all the distribution will be derived under the assumption of the yearly duration of the interval of analysis:

$$\Lambda_t = \Lambda$$

For example, Zelenhasic (1970) proposed to use the Gamma distribution for describing the distribution of exceedances $H(x)$, or its PDF $h(x)$:

$$h(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x \geq 0; \quad \alpha, \beta > 0$$

If α is a positive integer:

$$H(x) = \int_0^x \frac{\beta^{\alpha}}{\Gamma(\alpha)} u^{\alpha-1} e^{-\beta u}\, du \qquad x \geq 0; \qquad \beta > 0$$

Special case $\alpha = 1$ leads to:

$$H(x) = \int_0^x \beta e^{-\beta u}\, du = \left[-e^{-\beta u}\right]_0^x = 1 - e^{-\beta x} \qquad (2.36)$$

which is an exponential distribution.
Substituting (2.36) into (2.35) leads to

$$F(x) = e^{\left[-\Lambda e^{-\beta x}\right]} \qquad (2.37)$$

Eq. (2.37) is compatible with asymptotic laws described in paragraph 2.3. In fact, denoting with *X* the random variable representative of the annual maximum distribution, is easy to derive that:

$$H(x) = x = X - q_0$$

and, for

$$\beta = \sigma \qquad \Lambda = e^{\sigma(\zeta - q_0)}$$

Eq. (2.37) is much the same as Gumbel distribution (2.6) (Rossi et al., 1984). Same approach can be used for obtaining Fréchet distribution setting

$$H(x) = \ln(x)$$

and

$$\beta = \alpha \qquad \Lambda = e^{\sigma \ln(\zeta - q_0)}$$

Instead of exponential type, distribution of the annual maximum exceedance can be described by the generalized Pareto distribution in Eq. (2.14). Combining (2.35) and (2.14), under the hypothesis that $x > q_0$, is:

$$F(x) = exp\left\{-exp\left[-\frac{x - [q_0 + \sigma^* \ln(\Lambda)]}{\sigma^*}\right]\right\} \qquad \varepsilon = 0$$

$$F(x) = exp\left\{-\left[1 - \varepsilon \frac{x - \left[q_0 + \frac{\sigma^*(1 - \Lambda^{-\varepsilon})}{\varepsilon}\right]}{\sigma^* \Lambda^{-\varepsilon}}\right]^{1/\varepsilon}\right\} \quad \varepsilon \neq 0 \qquad (2.38)$$

Eq. (2.38) is a GEV distribution with the same shape parameter of the Generalized Pareto distribution.
Relationships between GEV and GP parameters are:

$$\zeta = q_0 + \sigma^* \ln(\Lambda) \qquad \varepsilon = 0$$

$$\zeta = q_0 + \frac{\sigma^*}{\varepsilon}(1 - \Lambda^{-\varepsilon}) \qquad \varepsilon \neq 0$$

$$\sigma^* = \sigma\Lambda^{\varepsilon}$$

At the point $x = q_0$ annual maximum cumulative distribution assumes the value $exp(-\Lambda)$, that is the probability of having no exceedances into one year.

Use of the Generalized Pareto distribution for characterizing the process of exceedances can be coupled with the hypothesis that the distribution of the number of events into one year, ν, has a binomial distribution (case 2 in paragraph 2.5.1) with parameters *n* and *p*. It follows that (Hosking, 1994):

$$\begin{aligned} F(x) &= \sum_{j=0}^{n} \binom{n}{j} p^j (1-p)^{n-j} [H(x)]^j = \\ &= \{1 - p[1 - H(x)]\}^n = \\ &= \left\{1 - p\left[1 - \varepsilon \frac{(x - q_0)}{\sigma^*}\right]^{1/\varepsilon}\right\}^n \end{aligned} \tag{2.39}$$

From Eq. (2.39), Hosking derived the following relationship:

$$F(x) = \left\{1 - h\left[1 - \varepsilon \frac{(x - \zeta)}{\sigma}\right]^{1/\varepsilon}\right\}^{1/h} \tag{2.40}$$

Eq. (2.40) is the expression of the four-parameters **Kappa** distribution, with $h = 1/n$.

Hosking (1994) noted that $1/h$ can be interpreted as the maximum number of independent events into one year.

In the end, according to classical theory of probability, for $n \to \infty$ and $n \cdot p$ remains constant, binomial degenerates into a Poisson distribution, and Eq. (2.40) becomes as the GEV in Eq. (2.38).

2.5.4 – Double component Poissonian model

As reported in Cunnane (1985), floods can be phenomenologically linked to different parent mechanisms (frontal or cyclonic rain, hurricanes and snowmelt), and this can explain the coexistence of different populations into a single sample. The presence of a second component into flood time series can be recognized, for example, by outliers, that can trigger to an unsatisfactory fit of traditional two- or three-parameter distributions.
Statistical modelling of this statement can be obtained replacing the hypothesis that variables Z_i are identically distributed with the hypothesis that the same process arises from a mixture of two component. For example (Singh and Sinclair, 1972):

$$H(x) = p \cdot H_1(x) + (1-p) \cdot H_2(x) \tag{2.41}$$

Under the hypothesis that exceedances are exponentially distributed, eq. (2.41) becomes (Rossi and Versace, 1982):

$$H(x) = p \cdot \left[1 - e^{-\theta_1 x}\right] + (1-p) \cdot \left[1 - e^{-\theta_2 x}\right] \tag{2.42}$$

where subscripts 1 and 2 are referred to respective components.
Introducing the additional hypothesis that the annual number of exceedances of each component, ν_1 and ν_2, follows a Poisson process with parameters Λ_1 and Λ_2, then, reproductive property guarantees that total annual number of exceedances $\nu = \nu_1 + \nu_2$ is Poisson distributed with parameter $\Lambda = \Lambda_1 + \Lambda_2$.
Then, proportions p and $(1-p)$ in eq. (2.41) will be:

$$p = \frac{\Lambda_1}{\Lambda_1 + \Lambda_2} \qquad 1 - p = \frac{\Lambda_2}{\Lambda_1 + \Lambda_2}$$

Combining (2.35) and (2.42) (Rossi et al., 1984)

$$F(x) = e^{\left(-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_2 e^{-\frac{x}{\theta_2}}\right)} \tag{2.43}$$

Eq. (2.43) represents the ***Two Component Extreme Value*** (**TCEV**) distribution, and is defined for $\mathrm{x} \geq 0$, $\Lambda_1 > \Lambda_2 \geq 0$ and $\theta_2 > \theta_1 > 0$.
Summing up, its four parameters represent the mean number of independent peaks into one year (Λ_1 and Λ_2) and the mean peak amplitude ($\theta_1$and θ_2) of the basic and the outlying component respectively (Rossi et al., 1984).

2.6 – Annual Maxima (AM) vs. Peak-Over-Threshold (POT) approaches

In this chapter Annual Maximum and Peak Over Threshold approaches for modelling hydrological extremes have been described. Here some important advantages and disadvantages of these models will be illustrated.

As noted in paragraph 2.3, Annual Maximum approach is a particular case of the more general theory of Block Maxima.
The choice of block size is a critical issue. Blocks too small are likely to lead to a poor approximation by the limit model in Fisher-Tippett theorem. This would lead to biases in the estimation of parameters and, consequently, in extrapolation. Large blocks would generate few block maxima, leading to a large estimation variance. It is therefore necessary to find a balance between the bias and the size of variances. Pragmatic considerations often lead to the adoption of blocks of length equal to one year, resulting in a series of annual maxima data. Furthermore, with this choice seasonal inhomogeneity problems are avoided. (Coles, 2001)

One of the most relevant limits of the AM approach is to consider only one value per year, excluding all other values that happens. Is often reported (e.g. Stedinger et al., 1993) that the second biggest value into one year can be greater than the largest in other years. As will be seen in chapter 6, this led to the incorporation into the analysis of values too small, not associated to any significant storm event. Including these data in FFA can influence the estimation of the right-hand tail of a frequency distribution. However, according to Cohn et al. (2013), this risk is very low for small probabilities of occurrence. Detection of this Potentially Influential Low Flows (PILFs) is an important step, and different tools have been proposed in literature for their identification (e.g. the Grubbs-Beck test).

Finally, AM approach conduces to relatively small time series, which can give high uncertainty in the final results. However, dealing with only one data per year can be considered as a guarantee that time series is composed of independent random variables, in agreement with the basic hypothesis of EVT.

The main quality of the POT model is basically its ability in the description of the flood phenomena, compounding modeling both of exceedance and base processes (Lang et al., 1999). Furthermore, this methodology gives the possibility of obtaining more numerous samples with respect to block maxima, leading to a more efficient estimate of parameters and quantiles.

On the other hand, problems can arise in the threshold selection, which is an arbitrary operation: in fact, an improper choice can violate the independence assumption of random variables. Moreover, from different thresholds contradictory results can be.

From a practical point of view, the choice between AM and POT models is strongly influenced by the availability of recorded data. For example, while countries as United States of America (USA), Canada and Australia have their own department which collect, validate and share water data, there are others in which this service is not efficient. It is the case of Italy, where water data recording is delegated to different regional or inter-regional departments, and time series of rainfall and discharges are only available in .pdf of in .xls files. An

exception are Calabria and Toscana region, in which data are provided in a user-friendly electronic format. In these cases, it is evident how collect one data per year is easier and faster for the practitioner, which is so tempted to use the AM model.

A last important remark has to be made about approaches used for describing probabilities of flood events. This work analyzes mainly Australian datasets and, in order to frame in previous studies about this topic, was considered reasonable to use notations proposed by Ball et al. (2019). This means that the measure of probability for AM approach will be the Annual Exceedance Probability (AEP), defined as the probability that a certain event is equaled or exceeded into one year.

CHAPTER 3: CLASSICAL APPROACH TO STATISTICAL INFERENCE

3.1 - Overview

In chapter 2 main probability distributions, according to their theoretical framework, were illustrated. One of the goals of this thesis, however, is to highlight properties and problems in using four-parameter distributions for flood frequency analysis. In this chapter, traditional methods of parameters estimation in the frequentist inference will be described, in order to give a rigorous basis for next chapter, where a comprehensive description of TCEV and Kappa will be provided in the same framework. In particular, paragraph 3.3 will deals with L-Moments Ratio Diagram, with the aim to describe this useful tool for discriminating between candidate distributions. Its usefulness in detecting the best fitting distribution will be illustrated in chapter 6.

3.2 – Methods of estimation

3.2.1 – Method of Moments

Velickov (2014) reported that this method was introduced before the era of Johann Bernoulli (1667-1748). Van Gelder (2004), in fact, noted that he made use of this method in his scientific production.

Method of Moments (MoM) is a very simple approach for parameters estimation and consists in their calculation by equating theoretical moments of an assigned probability distribution with sample moments. This means that a number of equations equal to the number to unknown parameters is required to be solved. As will be explained later, bias in sample estimation suggests to use the lowest order of moments as possible.

Define $F(x)$ and $f(x)$ respectively as the cumulative distribution function and the probability distribution function of a random variable x, while $\boldsymbol{\theta} =$

$[\theta_1, \ldots, \theta_k]$ is a *k*-dimensional vector containing the parameters of the distribution.

The *non-central moment about the origin* of order *r* of a probability distribution is:

$$\mu'_r = \int_{-\infty}^{+\infty} x^r f(x) dx \tag{3.1}$$

For $r = 1$ the non-central moment μ'_1 is equal to the mean.

Introducing a bias represented by the mean of this distribution, the *central moment of order r* about the mean is:

$$\mu_r = \int_{-\infty}^{+\infty} (x - \mu'_1)^r f(x) dx \tag{3.2}$$

Obviously, from (3.2), $\mu_1 = 0$.

Relationships between central and non-central moments of order *r* can be found in Kendall and Stewart (1977, pag. 58):

$$\mu_r = \sum_{i=0}^{r} \binom{r}{i} \mu'_{r-i} (-\mu'_1)^i$$

$$\mu'_r = \sum_{i=0}^{r} \binom{r}{i} \mu_{r-i} (\mu'_1)^i$$

Respective sample moments (with *n* length of sample), instead, are:

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{3.3}$$

$$m_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r \tag{3.4}$$

where $\bar{x} = m'_1$ and, as before, $m_1 = 0$.

Cunnane (1989) reported how (3.3) and (3.4) are biased estimates (larger the smaller the sample is), and suggested corrections. For central moments, for example, the following expressions can be applied:

$$\hat{m}_2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3.3a}$$

$$\hat{m}_3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n} (x_i - \bar{x})^3 \tag{3.4a}$$

However, these corrections are not sufficient for eliminate bias. Wallis et al. (1974) investigated the bias in the sample estimates of the standard deviation σ, $\hat{s} = (m_2)^{0.5}$, and in the skewness C_s, $m_3/m_2^{1.5}$ and noted that is function of sample size and skewness and form of the parent population.

Despite its formal simplicity, MoM estimates are less efficient is compared with other methods (e.g. Maximum Likelihood). High bias in relatively small sample makes MoM are *de facto* not convenient for multi-parameter distributions (three or more). Matalas and Wallis (1973) obtained the same conclusions comparing MoM and ML for the three parameters distribution Pearson type 3.

3.2.2 – Maximum Likelihood Method

It is a complicate study discovering who introduced Maximum Likelihood (ML) in the scientific literature. Kendall (1961) reported how traces can be found in Daniel Bernoulli's opera. Given the relevant statistical importance of this method, the history about it discover has been the subject of a special paper written by Aldich (1997).

Omitting the interesting historical notes, likelihood function can be defined as the joint pdf of the observations conditional on given values of the set of k parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_k]$ (Ramachandra Rao and Hamed, 2000), i.e.:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} f\big((x_i)\big|\boldsymbol{\theta}\big) \tag{3.5}$$

Unknown parameters can be estimated maximizing this function with respect the same parameters. Usually, instead of the Likelihood function, its logarithm is used:

$$\ln[\mathcal{L}(\boldsymbol{\theta})] = \sum_{i=1}^{n} \ln\big[f\big((x_i)\big|\boldsymbol{\theta}\big)\big] \tag{3.6}$$

The monotonic property of the logarithmic function guarantees the same maximum of the likelihood function $\mathcal{L}(\boldsymbol{\theta})$.

From a strictly mathematical perspective, it means to solve the k-dimensional homogeneous system of equations obtained by differentiating (2.6) with respect to parameters $\boldsymbol{\theta}$, i.e.:

$$\frac{\partial \ln[\mathcal{L}(\boldsymbol{\theta})]}{\partial \theta_i} = 0 \qquad i = 1,2,\ldots,k \tag{3.7}$$

ML estimator has different properties which make it a reliable tool for hydrological applications. It can be retained an efficient method since provides the smallest sampling variance of the estimated parameters (and hence of the quantiles) with respect of other methods. It is important to note that ML method frequently gives biased estimates, but corrections can be found in literature (Fiorentino and Gabriele, 1984). Problems in getting ML estimates can be found when dealing with small samples and multi-parametric models. The main problem of ML estimation has been assumed for long time as the computational effort for solving system (3.7). Evolution and diffusion of computer machines allowed to overcome this issue. Actually, for example, parameter estimation with ML can be easily obtained using the statistical software *R*: *extRemes* (Gilleland and Katz, 2016) and *ismev* are an example of affordable packages than can give parameter estimation in both stationary and non-stationary framework.

3.2.2.1 – Likelihood based model selection criteria

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are criteria for model selection widely employed in literature, due to their simplicity and clarity. They are likelihood-based criteria, and their use relies on the maximum-likelihood method.

Information criteria are useful tools for model selection. It is reasonable to retain that Akaike Information Criterion (*AIC*; Akaike, 1974) is the most famous among them. Based on the Kullbach-Leibler discrepance measure, if $\boldsymbol{\theta}$ is the parameter set of a *k*-dimensional model ($k = dim(\boldsymbol{\theta})$), AIC is defined as:

$$AIC = -2\ell\left(\widehat{\boldsymbol{\theta}}\right) + 2k.$$

The model that best fits data has the lowest value of *AIC* between candidates. It is useful to observe that the term proportional to the number of model

parameters allows to account for the increased estimator variance due to a larger parametrization and embodies the principle of parsimony.
Sugiura (1978) observed that *AIC* can lead to misleading results for small samples; he proposed a new measure for AIC:

$$AIC_c = -2\ell\left(\hat{\theta}\right) + \frac{2k(k+1)}{L-k-1}$$

where a second-order bias correction is introduced and *L* is the sample size. Burnham and Anderson (2004) suggested to use this version only when $L/k_{max} < 40$, being k_{max} the maximum number of parameters between the compared models. However, for larger *L*, AIC_c converges to AIC. For a quantitative comparison between *AIC* and AIC_c in the extreme value stationary model selection framework see also Laio et al. (2009).
AIC is widely also used in model selection between non-stationary and stationary distributions. Dealing with GEV and Gumbel distributions with a linear trend in the location parameter, Totaro et al. (2019) proposed a new way for using AIC, introducing the AIC_{ratio}, and noted how the power of this statistic in model selection is strongly influenced by sample size and parent distribution.

Bayesian Information Criterion (BIC) was originally introduced by Schwarz (1978) in a Bayesian framework. Its expression is:

$$BIC = -2\ell\left(\hat{\theta}\right) + \ln(L) \cdot k$$

In practical application, after the computation of the BIC, for all of the operating models one selects the model with the minimum BIC value (Laio et al., 2009).

3.2.3 – Probability Weighted Moments

Probability Weighted Moments (PWM) of a probability distribution having CDF $F = F(x) = P(X \leq x)$ can be defined as (Greenwood et al., 1979):

$$M_{p,r,s} = E[X^p F^r (1-F)^s] = \int_0^1 [x(F)]^p F^r (1-F)^s dF \qquad (3.8)$$

with $[p, r, s] \in \mathbb{R}$. Conventional noncentral moments can be derived from (3.8) for *r* and *s* equal to 0 and $p \in \mathbb{Z}$:

$$E(X^p) = \int_{-\infty}^{+\infty} x^p f(x) dx = \int_0^1 [x(F)]^p dF = M_{p,0,0}$$

Parameters estimation with PWMs is conceptually analogous to MoM, and consists in equaling sample and theoretical PWM.
Greenwood et al. (1979) reported also that if $p, r, s \in \mathbb{Z}_0^+$ then $M_{p,r,s}$ is proportional to the *p-th* noncentral moment of the *(r + 1)-th* order statistic for a sample of size $s + r + 1$:

$$E\left[X^p_{r+1,r+s+1}\right] = \frac{M_{p,r,s}}{B[r+1, s+1]}$$

with $B[\cdot,\cdot]$ the Beta function.
Setting $r = 0$, we will get:

$$E\left[X^p_{1,s+1}\right] = \frac{M_{p,0,s}}{B[1, s+1]}$$

It can be proved that:

$$E\left[X_{1,s+1}^{p}\right] = \frac{1}{k+1} M_{p,0,s} \tag{3.9}$$

which is the *p-th* moment about the origin of the first order statistic for a sample of size *s+1*.

Using *p* equal to 1 allows to use moments $M_{1,r,s}$ for parameter estimation: in this way, only the first power of *x* is used, avoiding problems arising in the method of moments with higher orders. Furthermore, use of high powers gives more weight to large observation (Rasmussen, 2001). Particularly useful PWM are:

$$\alpha_s = M_{1,0,s} = \int_0^1 x(F)\,(1-F)^s dF \tag{3.10a}$$

$$\beta_r = M_{1,r,0} = \int_0^1 x(F) F^r\, dF \tag{3.10b}$$

with *r* and *s* $\in \mathbb{Z}_0^+$.

Greenwood et al. (1979) reported that if *r* and *s* $\in \mathbb{Z}_0^+$ the following relationships hold:

$$M_{1,0,s} = \sum_{i=0}^{s} \binom{s}{i} (-1)^i M_{1,i,0} = \sum_{i=0}^{s} \binom{s}{i} (-1)^i \beta_i \tag{3.11a}$$

$$M_{1,r,0} = \sum_{i=0}^{r} \binom{r}{i} (-1)^i M_{1,0,s} = \sum_{i=0}^{r} \binom{r}{i} (-1)^i \alpha_i \tag{3.11b}$$

Application of the method of Probability Weighted Moments for parameters estimation allows to obtain in much cases simply relationships between parameters and moments $M_{p,r,s}$, especially when (3.10) are used.

Given a sample of *n* observations $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, be $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ its elements ranked in ascending order, i. e. the *order statistics*. An unbiased estimate of α_s and β_r can be obtained using the following relationships (Hosking, 1986), valid for $r, s = 0,1,\ldots, n-1$[5]:

$$a_s = \frac{1}{n}\sum_{i=1}^{n}\frac{\binom{n-i}{s}}{\binom{n-1}{s}}x_{(i)} = \sum_{i=1}^{n}\frac{(n-i)(n-i-1)\ldots(n-i-s+1)}{(n-1)(n-2)\ldots(n-s)}\,x_{(i)} \qquad (3.12a)$$

$$b_r = \frac{1}{n}\sum_{i=1}^{n}\frac{\binom{i-1}{r}}{\binom{n-1}{r}}x_{(i)} = \sum_{i=1}^{n}\frac{(i-1)(i-2)\ldots(i-r)}{(n-1)(n-2)\ldots(n-r)}x_{(i)} \qquad (3.12b)$$

Other estimators for α_s and β_r are (Hosking et al., 1985; Hosking and Wallis, 1987):

$$\hat{\alpha}_s = \frac{1}{n}\sum_{i=1}^{n}\left(1-p_{i,n}\right)^s x_{(i)} \qquad (3.13a)$$

$$\hat{\beta}_r = \frac{1}{n}\sum_{i=1}^{n}\left(p_{i,n}\right)^r x_{(i)} \qquad (3.13b)$$

where $p_{i,n}$ is

$$p_{i,n} = \frac{i+\gamma}{n+\delta}$$

i.e. a plotting position estimate of $F\left(x_{(i)}\right)$, with *γ* and *δ* constants.

Parameters of a probability distribution can be derived equaling theoretical expressions of PWMs (3.10) with sample ones. Generally, for two parameters

[5] It is useful remark that binomial coefficient $\binom{n}{k}$ is defined for *n, k* ∈ ℕ and $0 \le k \le n$.

distributions simple analytical expressions between parameters and PWMs can be obtained for *r* or *s* equal to 0 and 1. In particular, (3.11) allow the practitioner to choose the more convenient quantity α_s and β_r because of their mutual links. For a three parameters distribution, moment of order 2 can be used in addition (e.g. Hosking et al., 1985, for the GEV). For the GEV distribution, Wang (1997b) noted how using for β orders of 1,2,3 leads to a better fit of the right tail of the distribution.

Stedinger (1993) reported that (3.13) should be used in at-site analyses, because mean square error in quantile estimates is lower than that obtained using the unbiased (3.12).

Different applications have been made in parameters and quantile estimation using PWMs. Greenwood et al. (1979) derived expressions between parameters and PWMs for different distributions, including Weibull and Gumbel. Comparison between MoM, ML and PWM in estimation of parameters and quantiles has been made by Landwer et al. (1979) for Gumbel distribution and by Hosking and Wallis (1987) for Generalized Pareto distribution. Hosking et al. (1985) deriver parameters expressions in terms of PWMs for the GEV distribution. A review on PWMs theory and estimation can be found in Hosking (1986). Probability Weighted Moments for Kappa distribution are in Hosking (1994), while for TCEV are presented in Beran et al. (1986).

Hosking and Wallis (1997) noted that is not easy to get a direct interpretation of PWM as a measure of the scale and the shape of a distribution. However, this link can be obtained using *L-moments*, that will be introduced and described in the following paragraph.

3.2.4 – L-Moments

L-moments of a probability distribution have been introduced in Hosking (1986) as linear (from which the "*L*" of "L-moments") functions of the expected order statistics and can be defined as:

$$\lambda_r = \frac{1}{r}\sum_{i=0}^{r-1}(-1)^i \binom{r-1}{i} E\{X_{r-i:r}\} \qquad r \in \mathbb{Z}^+ \qquad (3.14)$$

being $E(\cdot)$ the expectation of an order statistic defined by:

$$E\{X_{j:r}\} = \frac{r!}{(r-j)!\,j!}\int x\{F(X)\}^{j-1}\{1-F(X)\}^{r-j}\,dF(X) \qquad (3.15)$$

Combining (3.14) and (3.15):

$$\lambda_r = \int_0^1 x(F)P_{r-1}^*(F)\,dF \qquad r \in \mathbb{Z}^+ \qquad (3.16)$$

where is the r-th shifted Legendre polynomial:

$$P_r^*(F) = \sum_{i=0}^{r} p_{r,i}^*\,F^i$$

and

$$p_{r,i}^* = (-1)^{r-i}\binom{r}{i}\binom{r+i}{i}$$

Shifted Legendre polynomials are related to the ordinary Legendre polynomials $P_r(u)$ as $P_r^*(u) = P_r(2u-1)$ and are orthogonal on the interval $(0,1)$ with constant weight function (Singh, 1998).
First four L-moments can be expressed as:

$$\lambda_1 = E(X) = \int_0^1 x(F)\, dF \tag{3.17}$$

$$\lambda_2 = \frac{1}{2} E\left(X_{(2:2)} - X_{(1:2)}\right) = \int_0^1 x(F)\, (2F - 1)\; dF \tag{3.18}$$

$$\lambda_3 = \frac{1}{3} E\left(X_{(3:3)} - X_{(2:3)} + X_{(1:3)}\right) = \int_0^1 x(F)\, (6F^2 - 6F + 1)\; dF \tag{3.19}$$

$$\begin{aligned} \lambda_4 = \frac{1}{4} E\left(X_{(4:4)} - 3X_{(3:4)} + 3X_{(2:4)} - X_{(1:4)}\right) \\ = \int_0^1 x(F)\, (20F^3 - 30F^2 + 12F - 1)\; dF \end{aligned} \tag{3.20}$$

L-moments can be expressed in terms of PWM using this general relationship (Hosking and Wallis, 1997):

$$\lambda_{r+1} = (-1)^r \sum_{i=0}^{r} p^*_{r,i}\, \alpha_i = \sum_{i=0}^{r} p^*_{r,i}\, \beta_i \tag{3.21}$$

Using (3.10), L-moments (3.17-3.20) become:

$$\lambda_1 = \alpha_0 \qquad = \beta_0 \tag{3.22}$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 \qquad = 2\beta_1 - \beta_0 \tag{3.23}$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 \qquad = \beta_0 - 6\beta_1 + 6\beta_2 \tag{3.24}$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 \quad = \beta_0 - 12\beta_1 - 30\beta_2 + 20\beta_3 \quad (3.25)$$

As for PWMs, parameters can be estimated equating (3.22-3.25) to their sample estimates ℓ obtained using (3.12) or (3.13).

In the previous paragraph has been noted that PWM cannot be easily explained as property of an assigned distribution. This is possible, instead, for L-moments, where λ_1 is the mean (a measure of location) and λ_2 is a measure of the scale of the distribution. Higher moments $\lambda_r, r \geq 3$ can be standardizes giving L-moments ratios:

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad (3.26)$$

Usually, the following quantities are mostly used:

L-Coefficient of Variation $$\tau = \frac{\lambda_2}{\lambda_1} \quad (3.27)$$

L-Skewness $$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (3.28)$$

L-Kurtosis $$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad (3.29)$$

Some properties of L-moments and L-moments ratios have to be remarked, also because will be useful for the analysis (Hosking and Wallis, 1997):

- Existence: if the distribution mean exists, then all of the L-moments exist.
- Uniqueness: if the distribution exists, then it is uniquely defined by the L-moments; two different distributions cannot have the same L-moments.

- sample L-moments are an unbiased estimate of the theoretical L-moments. The estimate of the L-moments ratios is not unbiased, but them bias is very small for large samples.

Some numerical limits have to be remarked too. These limits will be reported as a pointed list, because will be recalled in the next chapter about L-moments parameter estimation of TCEV and Kappa.

1. $\lambda_2 \geq 0$
2. $\|\tau\| < 1$ for $r \geq 3$. Given τ_3, there are limits for τ_4

$$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 < 1 \tag{3.30}$$

If distribution can take only positive values (as TCEV), following constraints can be applied:

$$2\tau - 1 \leq \tau_3 < 1 \tag{3.31}$$

3.3 – L-moments ratio diagram

L-moments ratio diagram is a useful graphical tool for representing L-moments of different distributions (Stedinger et al., 1993). We are dealing with a diagram that has on the abscissa the L-skewness τ_3 and on the ordinate the L-kurtosis τ_4. Basically, two parameters distributions are identified by a point on this diagram, while when parameters are three, they are represented by a line obtained for different values of the shape parameter (Hosking and Wallis, 1997).
LMRD for GEV, Generalized Logistic (GLO), Generalized Pareto (GPA), Pearson Type III (PE3), Exponential (EXP) and Gumbel (GUM) are reported in fig. 3.1 This diagram has been realized using the R package *lmomco* (Asquit, 2018).

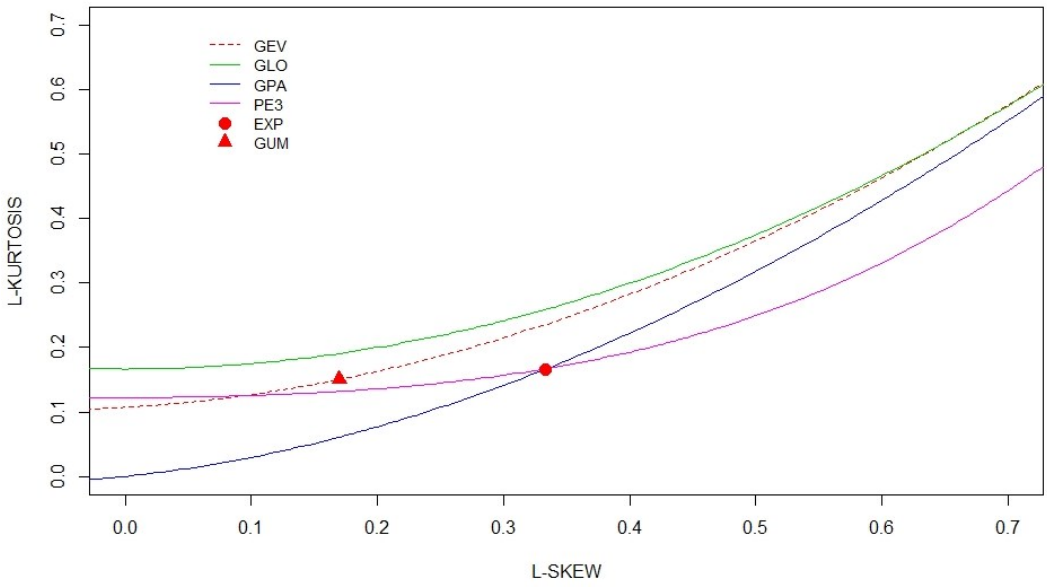

figure 3.1 – L-Moments Ratio Diagram

Build a such-made diagram can have some degree of complexity due to analytical expression of distributions or to long time required write functions required. Hosking and Wallis (1997) noted that plotting this diagram can be made easier using a polynomial approximation

$$\tau_4 = \sum_{k=0}^{8} A_k \tau_3^k$$

for each distribution. Coefficients for each distribution can be found in Hosking and Wallis (1997).

It is important to remark that in fig. 3.1, is reported the Pearson type 3 distribution and not the Log-Pearson type 3. Vogel et al. (1993a) noted that is not possible derive the theoretical relationships $\tau_3 - \tau_4$ for LP3. However, this issue can be overcome by compute sampling L-Moments for the logarithm of data.

Four parameters distribution, instead, are represented by an area on LMRD. Plots and considerations for Kappa and TCEV will be described in the next chapter.

CHAPTER 4: A REVIEW OF TCEV AND KAPPA DISTRIBUTIONS

4.1 - Introduction

Goal of this chapter is to provide a review of concepts described in the previous chapter for TCEV and Kappa distributions with regard to their point estimation. In particular, focus will be concentrated on the most popular methods in parameter estimation, such as Maximum Likelihood (MLE), Probability Weighted Moments (PWMs) and L-Moments (LME). The need of this review arises from the account that there are only some scientific papers which deal with the subject concerning the at-site Kappa distribution (Hosking, 1994; Parida, 1999; Winchester, 2000; Dupuis and Winchester, 2001), while only some notes about were given for TCEV distribution (Rossi et al., 1984; Beran et al., 1986; Arnell and Beran, 1988; Gabriele and Iiritano, 1994).
Particularly, TCEV L-Moments theoretical relationships will be applied in order to introduce this distribution in L-Moments Ratio Diagram (LMRD), a useful tool for identify appropriate distributions to fitted data (Ramachandra Rao ad Hamed, 2000). Indeed, only in some papers this diagram for TCEV was plotted; a visual comparison with other distributions can lead to interesting insight about goodness-of-fit.

4.2 – Kappa distribution

Kappa distribution was introduced by Hosking (1994) as a generalization of the three-parameter kappa distribution proposed by Mielke (1973). It is reasonable to assume that the most known use of this distribution is in the definition of the measures of heterogeneity and goodness of fit proposed by Hosking and Wallis (1993). However, multiple use can be found in literature.
Castellarin et al. (2007) applied Kappa to empirical series of dimensionless daily streamflows with a modified procedure for parameters estimation that allows to

ensure feasibility of solution. Blum et al. (2017) found that Kappa has a good fit to daily streamflows in the conterminous United States, and reported computational problems in parameters estimation. Kjeldsen et al. (2017) modelled Kappa for FFA in United Kingdom and found it having the best fit in small and wet catchments.

Point estimation of Kappa distribution was discussed in different studies. Parida (1999) applied Kappa to summer monsoon rainfall data in India with LME and found problems in computations due to feasibility region of searching. Winchester (2000) and Dupuis and Winchester (2001) discussed about problems in parameters estimation with MLE and LME and compared their performances both in parameters and quantiles estimation. Park and Jung (2002) proposed a penalized MLE for KAP. Singh and Deng (2003) proposed the application of Maximum Entropy for estimating parameters of Kappa and concluded that its performances are similar to LME. Park and Kim (2007) derived the exact expression and conditions of existence for Kappa Fisher information matrix. Murshed et al. (2014) derived the analytical distribution of LH moments of Kappa distribution and concluded that in the estimation of upper quantiles of distribution its performances are comparable to traditional LME. A good fit to wind speed data of Kappa distribution was found by Morgan et al. (2011) and Ouarda et al. (2015).

4.2.1 – Definition and moments

Kappa cumulative distribution can be expressed in this general form

$$F(x) = \left\{1 - h\left[1 - \varepsilon\frac{(x-\zeta)}{\sigma}\right]^{1/\varepsilon}\right\}^{1/h} \tag{4.1}$$

while its pdf is

$$f(x) = [F(x)]^{1-h}\frac{1}{\sigma}\left[1-\varepsilon\frac{(x-\zeta)}{\sigma}\right]^{\frac{1}{\varepsilon}-1} =$$

$$= \frac{1}{\sigma}\left\{1-h\left[1-\varepsilon\frac{(x-\zeta)}{\sigma}\right]^{1/\varepsilon}\right\}^{\frac{1}{h}-1}\left[1-\varepsilon\frac{(x-\zeta)}{\sigma}\right]^{\frac{1}{\varepsilon}-1} \qquad (4.2)$$

From eq. (4.1) it is possible to see that an explicit expression for the quantile of this distribution can be derived:

$$x(F) = \zeta + \frac{\sigma}{\varepsilon}\left[1-\left(\frac{1-F^h}{h}\right)^{\varepsilon}\right] \qquad (4.3)$$

According to the limiting values that parameters ε and h can assume, the CFD pf Kappa can assume different expressions:

$$F(x) = \begin{cases} \left\{1-h\left[1-\varepsilon\frac{(x-\zeta)}{\sigma}\right]^{\frac{1}{\varepsilon}}\right\}^{\frac{1}{h}} & if\ k\neq 0, h\neq 0 \\ e^{-\left[1-\varepsilon\frac{(x-\zeta)}{\sigma}\right]^{\frac{1}{\varepsilon}}} & if\ k\neq 0, h=0 \\ \left[1-he^{-\frac{(x-\zeta)}{\sigma}}\right]^{\frac{1}{h}} & if\ k=0, h\neq 0 \\ e^{-e^{-\frac{(x-\zeta)}{\sigma}}} & if\ k=0, h=0 \end{cases} \qquad (4.4)$$

As reported in Parida (1999), Kappa distribution can include multiple distributions for different combinations of h and ε:

h	ε	Distributions
1	≠ 0	Generalized Pareto
0	≠ 0	Generalized Extreme Value
-1	≠ 0	Generalized Logistic
1	0	Exponential
0	0	Gumbel
-1	0	Logistic
1	1	Uniform
0	1	Reverse Exponential

It is evident to note that, for different combinations of h andε, boundaries of distribution change:

$$\zeta + \frac{\sigma}{\varepsilon}(1 - h^{-\varepsilon}) \leq x \leq \zeta + \frac{\sigma}{\varepsilon} \qquad h > 0 \qquad \varepsilon > 0$$

$$\zeta + \sigma\, ln(h) \leq x < +\infty \qquad h > 0 \qquad \varepsilon = 0$$

$$\zeta + \frac{\sigma}{\varepsilon}(1 - h^{-\varepsilon}) \leq x < +\infty \qquad h > 0 \qquad \varepsilon < 0$$

$$-\infty < x < \zeta + \frac{\sigma}{\varepsilon} \qquad h \leq 0 \qquad \varepsilon > 0$$

$$-\infty < x < +\infty \qquad h \leq 0 \qquad \varepsilon = 0$$

$$\zeta + \frac{\sigma}{\varepsilon} \leq x < +\infty \qquad h \leq 0 \qquad \varepsilon < 0$$

These bounds are very important, because observations from a sample must satisfy them.

The r-th noncentral moment[6] of Kappa distribution can be obtained (in the spite of the discussion about PWM) as:

$$\mu'_r = \int_0^1 [x(F)]^r = \int_0^1 \left\{ \zeta + \frac{\sigma}{\varepsilon}\left[1 - \left(\frac{1-F^h}{h}\right)^{\varepsilon}\right]\right\}^r dF \qquad (4.5)$$

According to Winchester (2000), introducing the variable *y* and letting

$$y = 1 - \varepsilon\frac{(x-\zeta)}{\sigma}$$

the related CDF will be

$$G(y) = 1 - \left[1 - \varepsilon y^{\frac{1}{\varepsilon}}\right]^{\frac{1}{h}}$$

and the quantile function is

$$y(G) = \left[\frac{1-(1-G)^h}{h}\right]^{\varepsilon}$$

Setting *F = 1 – G*, will be

$$E[y^r] = \int_0^1 \left[\frac{1}{h}(1-F^h)\right]^{\varepsilon r} dF \qquad (4.6)$$

The evaluation of this integral should be done separately for the cases $h > 0$ and $h < 0$ (Hosking, 1994).

[6] r is a nonnegative integer

Case 1: h > 0

Setting $u = F^h$, integral (4.6) become:

$$E[y^r] = h^{-(1+r\varepsilon)} \int_0^1 (1-u)^{r\varepsilon} u^{\frac{1}{h}-1} du$$

Gradhteyn and Ryzhik (2007, eq. 3.191.3) state that, if $\mu, \nu > 0$, then

$$\int_0^1 x^{\nu-1}(1-x)^{\mu-1} dx = \mathfrak{B}(\mu, \nu)$$

with $\mathfrak{B}(\mu, \nu)$ the *beta function*.
Because (Gradhteyn and Ryzhik, 2007; eq. 8.384.1)

$$\mathfrak{B}(\mu, \nu) = \frac{\Gamma(\mu)\,\Gamma(\nu)}{\Gamma(\mu+\nu)}$$

Being $\Gamma(\mu, \nu)$ the *gamma function*. It follows that

$$\int_0^1 (1-u)^{r\varepsilon} u^{\frac{1}{h}-1} du = \frac{\Gamma(1+r\varepsilon)\,\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+r\varepsilon+\frac{1}{h}\right)}$$

Conditions of existence of this moment are:

$$\text{if } \varepsilon > 0 \qquad \Rightarrow \qquad r \in \mathbb{Z}_0^+$$

$$\text{if } \varepsilon \leq 0 \qquad \Rightarrow \qquad r\varepsilon < -1$$

Case 2: h < 0

In this case, solving the integral require the substitution $u = F^h - 1$. In this way, will be:

$$E[y^r] = (-h)^{-(1+r\varepsilon)} \int_0^{\infty} u^{r\varepsilon}(1+u)^{\frac{1}{h}-1} du$$

According to (Gradhteyn and Ryzhik, 2007; eq. 3.196.2), will become

$$E[y^r] = (-h)^{-(1+r\varepsilon)} \frac{\Gamma(1+r\varepsilon)\,\Gamma\left(-r\varepsilon - \frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}$$

Conditions of existence allow to write that:

if $\varepsilon > 0$ $\qquad r < -\frac{1}{rh}$

if $\varepsilon \leq 0$ $\qquad \begin{cases} rk > -1 \\ rk < -\frac{1}{h} \end{cases}$

Hosking (1994) reported that when using MoM for estimation, the first four moments of this distribution can be obtained for different values of (h, k). This has as direct consequence that MoM cannot be considered as a reliable tool for parameter estimation in the Kappa case.

4.2.2 – Maximum Likelihood

According to paragraph 3.2.2, given a sample of length *n*, likelihood function for Kappa distribution is:

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{\sigma^n}\prod_{i=1}^{n}\left\{1 - h\left[1 - \varepsilon\frac{(x-\zeta)}{\sigma}\right]^{1/\varepsilon}\right\}^{\frac{1}{h}-1}\left[1 - \varepsilon\frac{(x-\zeta)}{\sigma}\right]^{\frac{1}{\varepsilon}-1} \tag{4.7}$$

Log-likelihood function, therefore, will have the following structure:

$$\begin{aligned}\ln[\mathcal{L}(\boldsymbol{\theta})] = -n\ln(\sigma) + \left(\frac{1}{\varepsilon} - 1\right)\sum_{i=1}^{n}\ln\left\{1 - \left[\frac{\varepsilon(x_i-\zeta)}{\sigma}\right]\right\} \\ + \left(\frac{1}{h} - 1\right)\sum_{i=1}^{n}\ln\left\{\left[1 - h\left(1 - \frac{\varepsilon(x_i-\zeta)^{\frac{1}{\varepsilon}}}{\sigma}\right)\right]\right\}\end{aligned} \tag{4.8}$$

Reminding that the set of parameters $\boldsymbol{\theta}$ is equal to: $\boldsymbol{\theta} = [\zeta, \sigma, \varepsilon, h]$, maximum of (4.8) can be found setting equal to zero the first derivatives of (4.9) with respect to the four Kappa parameters:

$$S_1 = \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\zeta} = 0$$

$$S_2 = \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\sigma} = 0 \tag{4.9}$$

$$S_3 = \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\varepsilon} = 0$$

$$S_4 = \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial h} = 0$$

Introducing the variable $z_i = \varepsilon\frac{(x_i-\zeta)}{\sigma}$, $i = 1,\dots,n$, equations (4.9) become:

$$S_1 = \frac{\left(\frac{1}{\varepsilon}-1\right)}{\sigma}\varepsilon\sum_{i=1}^{n}\frac{1}{(1-z_i)} - \frac{\left(\frac{1}{h}-1\right)}{\sigma}\sum_{i=1}^{n}\frac{h(1-z_i)^{\frac{1}{\varepsilon}}}{\left[1-h(1-z_i)^{\frac{1}{\varepsilon}}\right](1-z_i)} = 0$$

$$S_2 = -\frac{n}{\sigma} + \frac{\left(\frac{1}{\varepsilon}-1\right)}{\sigma}\sum_{i=1}^{n}\frac{z_i}{(1-z_i)} - \frac{\left(\frac{1}{h}-1\right)}{\varepsilon\sigma}\sum_{i=1}^{n}\frac{h(1-z_i)^{\frac{1}{\varepsilon}}z_i}{\left[1-h(1-z_i)^{\frac{1}{\varepsilon}}\right](1-z_i)} = 0 \tag{4.10}$$

$$S_3 = -\frac{1}{\varepsilon^2}\sum_{i=1}^{n}\ln(1-z_i) - \frac{\left(\frac{1}{\varepsilon}-1\right)}{\varepsilon}\sum_{i=1}^{n}\frac{z_i}{(1-z_i)} + \\ -\left(\frac{1}{h}-1\right)\sum_{i=1}^{n}\frac{h(1-z_i)^{\frac{1}{\varepsilon}}\left[-\frac{1}{\varepsilon^2}\ln(1-z_i) - \frac{1}{\varepsilon^2}\frac{z_i}{(1-z_i)}\right]}{\left[1-h(1-z_i)^{\frac{1}{\varepsilon}}\right]} = 0$$

$$S_4 = -\frac{1}{h^2}\sum_{i=1}^{n}\ln\left[1-h(1-z_i)^{\frac{1}{\varepsilon}}\right] - \left(\frac{1}{h}-1\right)\sum_{i=1}^{n}\frac{(1-z_i)^{\frac{1}{\varepsilon}}}{\left[1-h(1-z_i)^{\frac{1}{\varepsilon}}\right]} = 0$$

that is a nonlinear system with four unknown parameters.

4.2.3 – PWM and L-Moments

According to the definition of Probability Weighted Moments given in the previous Chapter, has to be noticed that PWMs of a distribution exists only if the mean of the distribution exists. Following the development of (4.6) for $r = 1$, this

condition of existence is verified for $\varepsilon > -1$ if $h \geq 0$ and for $-1 < \varepsilon < -1/h$ for $h < 0$. (3.10b) can be written as:

$$\beta_r = \frac{1}{r+1}\left(\zeta + \frac{\sigma}{\varepsilon}\right) - \frac{\sigma}{\varepsilon}\int_0^1 \left[\frac{1}{h}(1-F^h)\right]^{\varepsilon} F^r \, dF \qquad (4.11)$$

Solution of this integral can be obtained in the same way of conventional moments. For $\varepsilon \neq 0$, Hosking (1994):

$$r\beta_{r-1} = \begin{cases} \zeta + \frac{\sigma}{\varepsilon}\left[1 - \frac{r\Gamma(1+\varepsilon)\Gamma(r/h)}{h^{1+\varepsilon}\Gamma(1+\varepsilon+r/h)}\right] & h > 0, \varepsilon > -1 \\ \zeta + \frac{\sigma}{\varepsilon}\left[1 - \frac{1}{r}\Gamma(1+\varepsilon)\right] & h = 0, \varepsilon > -1 \\ \zeta + \frac{\sigma}{\varepsilon}\left[1 - \frac{r\Gamma(1-\varepsilon)\Gamma(-\varepsilon-r/h)}{(-h)^{1+\varepsilon}\Gamma(1-r/h)}\right] & h < 0, -1 < \varepsilon < -1/h \end{cases} \qquad (4.12)$$

If $\varepsilon = 0$, instead:

$$r\beta_{r-1} = \begin{cases} \zeta + \alpha[\gamma + \ln(h) + \psi(1+r/h)] & h > 0 \\ \zeta + \sigma[\gamma + \ln(r)] & h = 0 \\ \zeta + \sigma[\gamma + \ln(-h) + \psi(-r/h)] & h < 0 \end{cases} \qquad (4.13)$$

where ψ is the *digamma function* and γ is the *Euler's constant*.
From equations (4.12) and (4.13) L-moments can be easily deducted. According to the theoretical framework, the following set of equations has to be numerically solved:

$$\begin{cases} \lambda_1 = \ell_1 \\ \lambda_2 = \ell_2 \\ \tau_3 = t_3 \\ \tau_4 = t_4 \end{cases} \qquad (4.14)$$

Distinguishing the three cases of eqs. (4.12) and (4.13) allows to evaluate the values of β_0, β_1, β_2 and β_3 for $\varepsilon \neq 0$ and $\varepsilon = 0$. Results are reported in Appendix A.

As reported by Hosking and Wallis (1997), four parameter distributions on L-Moments Ratio Diagram are not represented by a line, but by an area. Kjeldsen et al. (2017) reported its position on LMRD for $-1 \leq h \leq 1$, as in fig. 4.1:

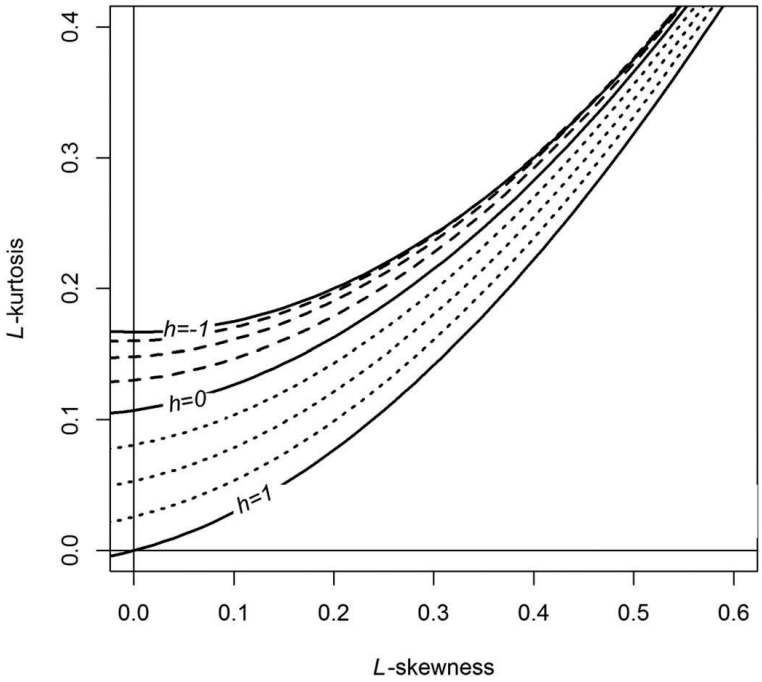

fig. 4.1 – LMRD for some values of h (from Kjeldsen et al., 2017)

A Fortran subroutine for estimating Kappa parameters with the methods of L-moments was provided by Hosking (2000). The availability of this subroutine gave a strong contribution for at-site applications with Kappa distribution. Is should be remarked how Hosking found that solutions of hydrological interest lie in the domain of $h > -1$: in his code, this constraint is applied. Furthermore, the same

code privileges solutions with the highest value of h (is more than one solution is available).

Again, Kjeldsen et al. (2017) reported polynomials expressions for Kappa curves on LMRD (table 4.1). Coefficient meaning was described in paragraph 3.3 about L-Moments Ratio Diagram.

	h							
	−1.00	−0.75	−0.50	−0.25	0.00	0.25	0.5	0.75
A_0	0.16667	0.15993	0.14804	0.13031	0.10701	0.08080	0.05313	0.02588
A_1	-	0.02101	0.04803	0.08044	0.11090	0.14431	0.16889	0.18734
A_2	0.83333	0.83146	0.82980	0.83009	0.84838	0.86000	0.88910	0.92319
A_3	-	−0.01700	−0.03850	−0.06646	−0.06669	−0.12105	−0.14619	−0.17023
A_4	-	0.00635	0.01946	0.04241	0.00567	0.05481	0.04945	0.04428
A_5	-	−0.00151	0.00324	0.01688	−0.04208	0.00739	−0.00501	−0.01053
A_6	-	0.00071	−0.01072	−0.04121	0.03763	−0.02960	−0.00823	0.00197
A_7	-	−0.00248	−0.01255	−0.03002	-	−0.03004	−0.01744	−0.00649
A_8	-	0.00152	0.01334	0.03802	-	0.03380	0.01647	0.00465

Table 4.1 - Polynomial expression for Kappa L-Moments (from Kjeldsen et al., 2017)

4.3 – Two Component Extreme Value (TCEV) distribution

Two Component Extreme Value (TCEV) distribution was introduced by Rossi et al. (1984) as the natural evolution of several works on outliers in hydrological records in southern Italy (Penta et al., 1980; Rossi and Versace, 1982). These studies were motivated by the need of finding a probability model able to justify and manage the systematic presence of outliers in different Italian floods and short-duration rainfalls time series. Beran et al. (1986) and Cunnane (1989) noticed that both TCEV and Wakeby (Houghton, 1978) distributions are able to justify the condition of separation introduced by Matalas et al. (1975) for Italian

and British sites. From the work of Cunnane (1989) emerges how since its publication no studies were produced for assessing the robustness of quantile estimates and advised against the at-site use of multiparameter distributions, such as TCEV and Wakeby because of large standard error in parameter estimation. Point estimation of TCEV parameter with MLE was proposed by Rossi et al. (1984), but authors noted that, when estimating from a single time series of annual maxima, uncertainty is high. However, they did not give any quantitative assessment of this uncertainty. Cunnane (1987) noticed the instability of the solution of at-site TCEV. Analytical derivation of PWM can be found in Beran et al. (1986), while LME of TCEV parameters was described in Arnell and Beran (1988), Gabriele and Arnell (1991) and Gabriele and Iiritano (1994). Fiorentino et al. (1987) and Singh (1998) derived a maximum entropy-based method for estimating both at-site and regional TCEV parameters and quantiles. They found this method be simpler than MLE and in, a regional approach, they concluded that performances are of the same order of MLE approach. Connell and Pearson (2001) applied at-site TCEV to annual maximum floods for the Canterbury region in New Zealand using least squared method. They found a good fit of this distribution for rivers located in the South Canterbury East Coast and justified differences between northern and southern part of the above region with orographic effects and storms directions. It is interesting to note the first use of L-moments ratio diagram for TCEV.

Conclusions about non optimal performances of point estimation of TCEV parameters probably led scientists to focus their attention only on the regional applications. This motivation was also justified by the above cited verification of the condition of separation. Fiorentino et al. (1987) proposed a hierarchical approach for regional frequency analysis with TCEV, able to combine regional and at-site estimation. Arnell and Gabriele (1988) analyzed performances of regional TCEV and noted that: i) failure in estimation can arise when outliers in time series are few; ii) TCEV regional procedure is robust; iii) performances of TCEV are directly proportional to region size, and inversely to sample size. Gabriele and

Arnell (1991) highlighted the importance of using a hierarchical approach and the reduction in quantiles estimation when into a region with a constant skewness is subdivided in sub-regions with constant coefficient of variation. Furthermore, they found the TCEV-MLE be more efficient than TCEV-PWM method.
Application of TCEV in regional frequency analysis was proposed for Italy and for some regions of Spain (Castellarin et al., 2012). In particular, for Italy the hierarchical approach proposed by Gabriele and Arnell (1991) was applied; in fig. 4.2 homogeneous regions for coefficient of skewness (first level of regionalization) are reported.

fig. 4.2 – Homogeneous areas for TCEV at first level of regionalization in Italy (from Castellarin et al., 2012)

This approach was formalized with VA.PI. and CUBIST projects. This latter was proposed in the Prediction in Ungauged Basins (PUB) research area.

In Spain, instead, TCEV model for preferred in those Mediterranean regions which showed a different behavior in rainfall mechanism: frontal storms, that are more frequent and representative of the first component of TCEV, and heavy convective storms, which described better the second component (Castellarin et al., 2012).

4.3.1 – Definition and moments

TCEV cumulative distribution function is:

$$F(x) = e^{\left(-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_2 e^{-\frac{x}{\theta_2}}\right)}$$

and its density function:

$$f(x) = e^{\left(-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_2 e^{-\frac{x}{\theta_2}}\right)} \left(\frac{\Lambda_1}{\theta_1} e^{-\frac{x}{\theta_1}} + \frac{\Lambda_2}{\theta_2} e^{-\frac{x}{\theta_2}}\right) \qquad (4.15)$$

Both equations are valid for $x \geq 0$ and $\Lambda_1 > \Lambda_2 \geq 0$ and $\theta_2 > \theta_1 > 0$.
No explicit solutions can be found for evaluating quantiles of TCEV. This requires the use of root solver, such as the *pegasus* or its modifications (e.g. King, 1973).
TCEV is an exponential type distribution, and its left-hand limit can be found for $x = 0$ and is

$$F(0) = e^{-\Lambda_1 - \Lambda_2}$$

On the right-hand, for $F(x) \leq 1$, limit is given by:

$$-\Lambda_1 e^{-\frac{x}{\theta_1}} - \Lambda_2 e^{-\frac{x}{\theta_2}} < 0$$

TCEV distribution can degenerate into Gumbel distribution with different combinations of its four parameters. The basic condition is, obviously, that Λ_2 is equal to zero. Another condition is that parameters are such that the absolute value of the second component will asymptotically be close to zero for all values of *x*, i. e.

$$\left|-\Lambda_1 e^{-\frac{x}{\theta_1}}\right| >> \left|-\Lambda_2 e^{-\frac{x}{\theta_2}}\right|$$

TCEV noncentral moments have been derived in Beran et al. (1986), and are:

$$\mu'_r = m'_r + \theta_1 \sum_{j=0}^{\infty} \frac{(-1)^j \Lambda_*^j}{j!} \sum_{k=1}^{r} (-1)^k k \binom{r}{k} [\ln(\Lambda_1)]^{r-k} \Gamma^{(k-1)}(j/\theta_*) \qquad (4.16)$$

with

$$m'_r = \theta_1^r \sum_{k=0}^{r} (-1)^k \binom{r}{k} [\ln(\Lambda_1)]^{r-k} \Gamma^{(k-1)}(1)$$

m'_r represents the *r*th noncentral moment of the basic series. Instead:

$$\theta_* = \frac{\theta_2}{\theta_1} \qquad \Lambda_* = \frac{\Lambda_2}{\Lambda_1^{1/\theta_*}} \qquad (4.17)$$

4.3.2 – Maximum likelihood estimation

Likelihood function for TCEV is:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^{n} e^{\left(-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right)} \left(\frac{\Lambda_1}{\theta_1} e^{-\frac{x_i}{\theta_1}} + \frac{\Lambda_2}{\theta_2} e^{-\frac{x_i}{\theta_2}}\right) \tag{4.18}$$

Its natural logarithm is:

$$\ln[\mathcal{L}(\boldsymbol{\theta})] = \sum_{i=1}^{n} \left[-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right] + \sum_{i=1}^{n} \ln\left[\frac{\Lambda_1}{\theta_1} e^{-\frac{x_i}{\theta_1}} + \frac{\Lambda_2}{\theta_2} e^{-\frac{x_i}{\theta_2}}\right] \tag{4.19}$$

In this case $\boldsymbol{\theta}$ is equal to: $\boldsymbol{\theta} = [\Lambda_1, \theta_1, \Lambda_2, \theta_2]$. As for Kappa distribution, maximum of (4.19) can be found setting equal to zero the first derivatives with respect to the four TCEV parameters (Rossi et al., 1984):

$$\begin{aligned}
T_1 &= \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\Lambda_1} = -\sum_{i=1}^{n}\left(e^{-\frac{x_i}{\theta_1}}\right) + \frac{1}{\theta_1}\sum_{i=1}^{n}\left[\frac{e^{-\frac{x_i}{\theta_1}}}{\left(-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right)}\right] = 0 \\
T_2 &= \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\Lambda_2} = -\sum_{i=1}^{n}\left(e^{-\frac{x_i}{\theta_2}}\right) + \frac{1}{\theta_2}\sum_{i=1}^{n}\left[\frac{e^{-\frac{x_i}{\theta_2}}}{\left(-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right)}\right] = 0 \\
T_3 &= \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\theta_1} = -\left(\frac{\Lambda_1}{\theta_1^2}\right)\left\{\sum_{i=1}^{n} x_i e^{-\frac{x_i}{\theta_1}} + \sum_{i=1}^{n}\left[\frac{e^{-\frac{x_i}{\theta_1}}\left(1-\frac{x_i}{\theta_1}\right)}{\left(-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right)}\right]\right\} = 0 \\
T_4 &= \frac{\partial\{\ln[\mathcal{L}(\boldsymbol{\theta})]\}}{\partial\theta_2} = -\left(\frac{\Lambda_2}{\theta_2^2}\right)\left\{\sum_{i=1}^{n} x_i e^{-\frac{x_i}{\theta_2}} + \sum_{i=1}^{n}\left[\frac{e^{-\frac{x_i}{\theta_2}}\left(1-\frac{x_i}{\theta_2}\right)}{\left(-\Lambda_1 e^{-\frac{x_i}{\theta_1}} - \Lambda_2 e^{-\frac{x_i}{\theta_2}}\right)}\right]\right\} = 0
\end{aligned} \tag{4.20}$$

Solving this set of equations will give the four unknown parameters.

4.2.3 – PWMs and L-Moments

According to notations of eq. (3.8), Probability Weighted Moments for TCEV and Kappa distribution have been derived by Beran et al. (1986) for $p = 1$ and $s = 0$:

$$\beta_r = M_{1,r,0} = PWM_r^{(1)} + \frac{\theta_1}{r+1} T_r \tag{4.21}$$

where

$$PWM_r^{(1)} = \frac{\theta_1}{r+1} [\gamma + \ln(\Lambda_1) + \ln(r+1)] \tag{4.22}$$

is the *r*-th probability weighted moment of the basic series and

$$T_r = \sum_{j=1}^{\infty} (-1)^{j-1} \Lambda_*^j (r+1)^{j\left(1-\frac{1}{\theta_*}\right)} \frac{\Gamma\left(\frac{j}{\theta_*}\right)}{j!} \tag{4.23}$$

According to Gabriele and Arnell (1991), theoretical L-moments for TCEV distribution are:

$$\begin{aligned}
\lambda_1 &= \beta_0 && = \theta_1 [\gamma + \ln(\Lambda_1) + T_0] \\
\lambda_2 &= 2\beta_1 - \beta_0 && = \theta_1 [\ln(2) + D_1] \\
\lambda_3 &= 6(\beta_2 - \beta_1) + \beta_0 && = \theta_1 \left[\ln\left(\frac{9}{8}\right) + 2D_2 - D_1\right] \\
\lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 && = \theta_1 \left[\ln\left(\frac{2^{16}}{3^{10}}\right) + 5D_3 - 5D_2 + D_1\right]
\end{aligned} \tag{4.24}$$

being $D_r = T_r - T_{r-1}$.
L-skewness and L-kurtosis can be easily evaluated:

$$\tau_3 = \frac{\lambda_3}{\lambda_2} = \frac{\ln\left(\frac{9}{8}\right) + 2D_2 - D_1}{\ln(2) + D_1} \qquad (4.25)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} = \frac{\ln\left(\frac{2^{16}}{3^{10}}\right) + 5D_3 - 5D_2 + D_1}{\ln(2) + D_1}$$

According to the general method for estimating parameters with the method of L-Moments, a system of four equations in four unknown parameters has to be solved for estimating the four TCEV parameters (Gabriele and Iiritano, 1994).
Unlike for Kappa, no subroutines or codes are available for estimating TCEV parameters with the method of L-Moments and Maximum Likelihood.
As expressed in previous paragraphs, the estimation of the four at-site TCEV parameters require the solution of a nonlinear system of four equations in four unknown parameters. Different methods are available in literature for finding a numerical solution of such-made systems. Newton-Raphson method is the simplest multidimensional root finding method, but its convergence is conditionate to the starting point of the algorithm. This can lead to a more frequent rejection of the hypothesis that system has no solutions.

Plot of TCEV distribution on LMRD is not so easy to obtain. In literature can be found only in Arnell and Beran (1988) and Gabriele and Arnell (1991). Some of its characteristics are illustrated in Connell and Pearson (2001). However, in all cases this area was not compared with other distributions.
This complete LMRD is illustrated in the following figure 4.3. Note that for building LMRD (except for TCEV), R package *lmomco* has been used (Asquith, 2018).

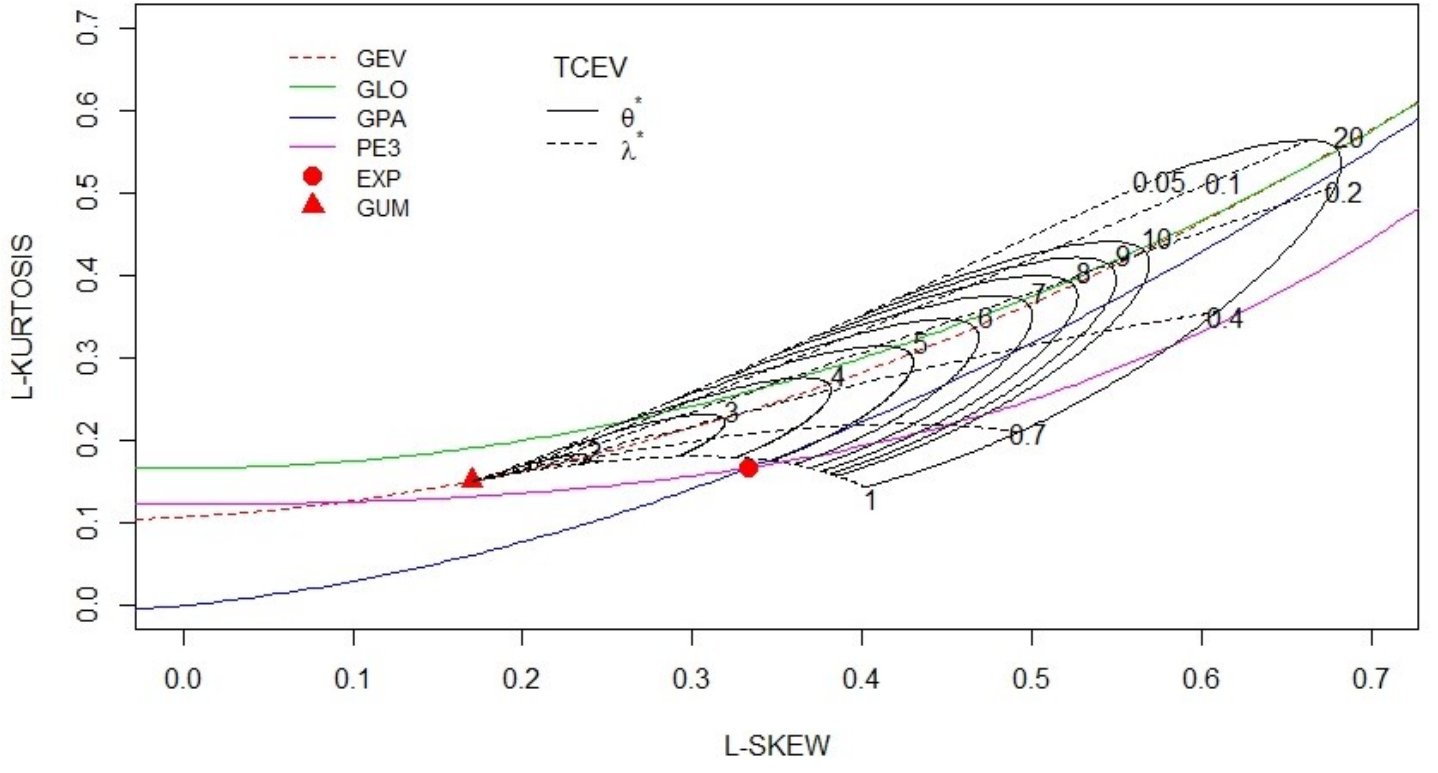

figure 4.3 – L-Moments Ratio Diagram with TCEV distribution

L-Moments for TCEV has been reported for a maximum value of $\theta^*$of 20. However, it is possible to increase it, but for practical purposes can be deemed a reliable upper limit.
When looking at this diagram several considerations arise.

1. TCEV's area moves from the Gumbel point: considering that TCEV comes from two different populations Gumbel distributed, this is a graphical confirmation of its nature;
2. point 1 can be important for discriminating the presence of the second component of TCEV: if presence is weak, it can be hard to be distinguished.
3. TCEV is able to cover areas over the Generalized Logistic (the limit imposed by Hosking for Kappa) and under the PE3 distribution;
4. TCEV cannot explain distributions which combinations of (τ_3, τ_4) lower than those corresponding to Gumbel distribution;
5. GEV line seems cut in half TCEV area, moving from the Gumbel point, which correspond to its limits case.

Reproduce TCEV on LMRD can be tricky. Coefficients for polynomial approximations for curves of Λ^* (as described in Par. 3.3) are reported in table 4.2.

	Λ*					
	1	0.7	0.4	0.2	0.1	0.05
A_0	0.60502	0.16019	0.20163	0.24315	0.31885	0.52281
A_1	-14.39219	-1.37088	-1.97840	-2.74212	-4.35089	-8.94998
A_2	175.47031	16.20462	18.39390	23.11974	36.10041	77.70206
A_3	-1134.23455	-78.48093	-73.88646	-86.24619	-139.81071	-338.82988
A_4	4339.80707	224.11676	177.07711	192.48525	322.84410	878.34550
A_5	-9884.49020	-387.37502	-253.51274	-255.11716	-441.44235	-1348.24547
A_6	12412.34914	370.96819	199.20222	185.17320	330.13728	1133.15577
A_7	-6659.30979	-154.43081	-66.24105	-56.71435	-104.11043	-402.18168
$\tau_{3,\,min}$	0.16988	0.16993	0.16993	0.16993	0.16993	0.16993
$\tau_{3,\,max}$	0.39988	0.48993	0.60493	0.67493	0.65993	0.57993

table 4.2 – Polynomial coefficients for Λ^*

For each of these curves, in last two rows boundaries with limits on LMRD are reported.

Compared to those of other distributions, it is clear how the module of A_k coefficients are much greater. This aspect can be explained with the marked curvature of TCEV lines, which reflect to the analytical form of polynomial coefficients.

Analytical derivations of polynomial relationships for $\theta^*$are more complicated by the not bijection between τ_3 and τ_4 not only in the function $\tau_4 = f(\tau_3)$ but also in its inverse. Basically, an attempt for finding these coefficients can be made by splitting these curves and finding the optimal combination in terms of goodness-of-fit. However, this can be investigated in future.

CHAPTER 5 – BAYESIAN INFERENCE

5.1 – Overview

Bayesian inference is the main topic of this chapter. The ability of Bayesian theory in evaluating the uncertainty in parameters and quantiles is an attractive quality that make it the preferable method for analyzing and interpreting results about four parameters distributions. In paragraph 5.2 an overview on Bayesian inference is illustrated, with a focus on the role of uncertainty and of Markov chain Monte Carlo (MCMC) methods. In 5.3 a description of the proposed Bayesian approach is provided and main characteristics of Metropolis-Hastings algorithm are illustrated.
In paragraph 5.4, finally, a new measure for distinguishing the presence of two components into a population, based on the sampling of posterior distribution, is described. Its skills will be evident in chapter 6.

5.2 – Bayesian framework

5.2.1 - Description

The use of Bayesian methodologies for floods frequency analysis can be dated back to the works of Wood and Rodriguez-Iturbe (1975a, b) and Vicens et al. (1975). However, computational issues for long times did not encourage its use in hydrology. Currently, Bayesian methodologies are widely employed and recognized as an affordable tool, also because of their ability in incorporating several sources of information for analysis.
Bayesian inference relies on the application of Bayes' Theorem, formulated by Thomas Bayes and published by Richard Price in 1763, two years after Bayes' death. This theorem has been applicated in many fields of sciences.

In the Bayesian framework, the set of *n* observations[7] $\boldsymbol{x} = (x_1, x_2, \ldots x_n)$ are considered random realizations depending from a probability distribution $f(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the set of unknown parameters. It is useful to remark that $f(\boldsymbol{x}|\boldsymbol{\theta})$ is known both as *sampling distribution* (in case is used for describing the model that generate $\boldsymbol{x}$ for a given $\boldsymbol{\theta}$) or *likelihood function,* $\ell(\boldsymbol{x}|\boldsymbol{\theta})$ (when data $\boldsymbol{x}$ are known and inference is needed only on $\boldsymbol{\theta}$). In the following all the description will deals with this latter case.

One of the main characteristics of Bayesian approach is that the parameter set $\boldsymbol{\theta}$ is considered a random vector: the advantage is that its probability distribution will contain all the knowledge about the true value of $\boldsymbol{\theta}$.

Another important property of Bayesian inference relies in its ability of modelling the knowledge about $\boldsymbol{\theta}$ trough the introduction of the prior distribution $p(\boldsymbol{\theta})$.

With these statements, Bayes' rule can be formalized as:

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\ell(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int_{\boldsymbol{\Theta}} \ell(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}} \propto \ell(\boldsymbol{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \qquad (5.1)$$

$p(\boldsymbol{\theta}|\boldsymbol{x})$ is known as the *posterior distribution* of the parameters $\boldsymbol{\theta}$ and the integral in the denominator of eq. (5.1), computed in the whole parameter space $\boldsymbol{\Theta}$, is a normalization constant, whose goal is to guarantee that the area under the posterior density is one. This explain the proportional implication in eq. (5.1).

It should be remarked that uncertainty evaluation is possible also in a not-Bayesian framework, using an approach based on parametric bootstrap. However, it has not been implemented into this thesis.

5.2.2 – Computation of the posterior distribution

[7] In our case, these observations are annual maximum floods.

One of the main differences between frequentist and Bayesian approaches relies on presence of the posterior distribution. In fact, likelihood-based inference allows only a point estimate of the parameters, and uncertainty can be described only through an asymptotic approximation of the likelihood function (Reis and Stedinger, 2005). On the other side, the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{x})$ introduced in the Bayesian framework contains all the information about the parameters (considered as random variables), giving a precise quantification of uncertainty.
Nevertheless, achieving of an exact analytical solution for the posterior distribution can be complicated. A tractable relationship for the posterior can be obtained when both the posterior and the prior distributions are in the same probability distribution family: in this case, prior and posterior are called *conjugate distributions*.
Even if this latter case is very useful, it happens very rarely. This issue has been one of the key factors that limited the diffusion of the Bayesian framework in the scientific community (Gelman et al. 2004).
The development of Markov Chain Monte Carlo (MCMC) methods gave a boost to Bayesian theory development, because their ability of sampling values from the posterior distribution without the need of computing the normalization constant (Reis and Stedinger, 2005).

5.2.3 – Markov Chain Monte Carlo (MCMC) methods

Markov chain Monte Carlo (MCMC) techniques can realize an efficient sampling of the posterior distribution, because of their ability in detecting the zone where there is a peak, rejecting parts of parameter space where probabilities are low. Direct consequence is that the number of sampled points grows linearly and not exponentially with the dimension of the sample space.
MCMC is a method that allows to draw samples $\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(n)}$ from an approximate distribution and then making some corrections to those draws, with

the goal of finding a good approximation to a target distribution $\pi(\boldsymbol{\psi})$ (Gelman et al., 2004). Write:

$$\pi(\boldsymbol{\psi}) = C \cdot f(\boldsymbol{\psi})$$

where $f(\boldsymbol{\psi})$ in an unnormalized density and C is a constant of normalization (generally unknown). MCMC methods draw a sequence of samples from a Markov chain whose stationary distribution converges to the target distribution. In this Bayesian framework, target and posterior distributions coincides.
Convergence of the Markov chain to target distribution can be proved with a two-step procedure:

1. prove the convergence of the Markov chain converges to a stationary distribution: this requires that Markov chain be irreducible, aperiodic and not transient. Irreducibility holds if the Markov chain has a positive probability of visiting every part of the target distribution from any other part of the target distribution, while remaining conditions hold for a random walk of any proper distribution;
2. show that the stationary distribution is equal to the target distribution.

According to continuous Markov chain theory, a transition kernel $K(\boldsymbol{\psi}|\boldsymbol{\omega})$ can be defined as the conditional pdf for moving from a point $\boldsymbol{\psi}$ to a point $\boldsymbol{\omega}$. It should be remarked that chain cannot move from the point $\boldsymbol{\psi}$, i.e. there is a transition from $\boldsymbol{\psi}$ to $\boldsymbol{\psi}$.
The stationary pdf must satisfy

$$\pi(\boldsymbol{\omega}) = \int K(\boldsymbol{\omega}|\boldsymbol{\psi})\, \pi(\boldsymbol{\psi})\, d\boldsymbol{\psi} \qquad (5.2)$$

Analysis of (5.2) can lead to the conclusion that if the chain is stationary, then the total probability of moving to a region from any point $\boldsymbol{\psi}$ is equal to the unconditional (or marginal) probability of sampling that region. In conclusion, chain can be retained no longer affect by the starting point $\boldsymbol{\psi}^{(0)}$.

The approach applied in MCMC methods denotes a different nature: the stationary pdf - $\pi(\cdot)$, i.e. the target density - is known up to a constant of probability, unlike the transitional kernel. Within this framework, for generating samples from $\pi(\cdot)$, in MCMC methods a transition kernel $K(\boldsymbol{\omega}|\boldsymbol{\psi})$ able to produce samples converging to $\pi(\cdot)$ is sought.

An affordable transition kernel is found in MCMC theory by construction one able to describes the probability density of either moving from $\boldsymbol{\psi}$ to a new point $\boldsymbol{\omega}$ or remaining at $\boldsymbol{\psi}$:

$$K(\boldsymbol{\omega}|\boldsymbol{\psi}) = p(\boldsymbol{\omega}|\boldsymbol{\psi}) + r(\boldsymbol{\psi}) \cdot \delta(\boldsymbol{\omega} - \boldsymbol{\psi}) = \\ = \begin{cases} p(\boldsymbol{\omega}|\boldsymbol{\psi}) & if\ \boldsymbol{\omega} \neq \boldsymbol{\psi} \\ r(\boldsymbol{\psi}) \cdot \delta(0) & if\ \boldsymbol{\omega} = \boldsymbol{\psi} \end{cases} \tag{5.3}$$

being $p(\boldsymbol{\omega}|\boldsymbol{\psi})$ a nonnegative function with $p(\boldsymbol{\psi}|\boldsymbol{\psi}) = 0$, $\delta(\boldsymbol{\omega} - \boldsymbol{\psi})$ is the well-known Dirac delta function centered about $\boldsymbol{\psi}$, and

$$r(\boldsymbol{\psi}) = 1 - \int p(\boldsymbol{\omega}|\boldsymbol{\psi})\, d\,\boldsymbol{\omega} \tag{5.4}$$

is the probability of the chain remaining at $\boldsymbol{\psi}$.

In order to guarantee the stationarity of a Markov chain, is required to satisfy the reversibility condition:

$$\pi(\boldsymbol{\psi})p(\boldsymbol{\omega}|\boldsymbol{\psi}) = \pi(\boldsymbol{\omega})p(\boldsymbol{\psi}|\boldsymbol{\omega}) \tag{5.5}$$

Testing that $\pi(\cdot)$ is the stationary density of the transition kernel $K(\cdot \mid \cdot)$ if the reversibility condition is satisfied can be performed combining eq. (5.2), (5.3) and (5.5) and using properties of delta function:

$$\int K(\boldsymbol{\omega}|\boldsymbol{\psi})\pi(\boldsymbol{\psi})\, d\boldsymbol{\psi} = \int [p(\boldsymbol{\omega}|\boldsymbol{\psi}) + r(\boldsymbol{\psi})\delta(\boldsymbol{\omega} - \boldsymbol{\psi})]\, \pi(\boldsymbol{\psi}) d\boldsymbol{\psi} =$$

$$= \int p(\boldsymbol{\psi}|\boldsymbol{\omega})\, \pi(\boldsymbol{\omega}) d\boldsymbol{\psi} \; + \int r(\boldsymbol{\psi})\delta(\boldsymbol{\omega} - \boldsymbol{\psi})\pi(\boldsymbol{\psi}) d\boldsymbol{\psi} =$$

$$= \pi(\boldsymbol{\omega}) \int p(\boldsymbol{\psi}|\boldsymbol{\omega})\, d\boldsymbol{\psi} + \boldsymbol{r}(\boldsymbol{\omega})\pi(\boldsymbol{\omega}) \cdot \mathbf{1} = \pi(\boldsymbol{\omega})$$

q.e.d.

5.2.4 – The nature of uncertainty

Evaluation of uncertainty is a very important step in hydrological processes. Since hydrology is intrinsically affected by a high degree of uncertainty (Montanari et al., 2009), practical applications should not ignore its quantification. A review on the supporting reasons for undertaking the analysis of uncertainty was provided by Pappenberger and Beven (2006) For the purpose of allowing an aware management, two types of uncertainty are distinguished (National Research Council, 2000):

- *natural variability*: is the uncertainty that arises from variability that can be observed in nature; in the scientific literature, is also defined with the following terms: *aleatory, external, objective, random, stochastic, inherent*;
- *knowledge uncertainty*: is due to a lack of understanding events and processes. It is also known as *epistemic, functional, internal, subjective* uncertainty.

This latter, due to its intrinsic nature, it the only one that can be reduced, while natural variability can be only characterized.
Defining as *predictive* uncertainty the total uncertainty in the estimates of the design flood (Ball et al., 2019), it can be split in difference sources. In flood frequency analysis with at-site data the most relevant are:

- *data uncertainty*: is the uncertainty connected with streamflow data. Includes, inter alia, uncertainty about the quality of the rating curve employed for estimating streamflows and its stability;
- *parametric uncertainty*: this source of uncertainty is connected with parameters estimation, and all the related problematics (e.g.: estimation method, length of time series). Is a function of data uncertainty in such a way that parametric uncertainty increases as data uncertainty increases);
- *structural uncertainty*: it the uncertainty related to the probabilistic model used for fitting data.

A detailed description of these sources of uncertainty can be found in Merz and Thieken (2005).

5.3 – On the use of the Bayesian approach

Bayesian inference for TCEV and Kappa distribution was performed working on source codes of FLIKE software under the supervision of prof. George Kuczera during a visiting period at The University of Newcastle (Australia) from 26/02/2018 to 26/07/2018. FLIKE was recommended for flood frequency analysis in Australia in ARR[8] 2016 (Paul et al., 2016). Working on original scripts (in Fortran language) allowed to obtain the same approach for FFA, giving results totally comparable with previous studies.

[8] Australian Rainfall and Runoff

FLIKE software structure was thoroughly illustrated in Kuczera (1999), while a guide for users can be found in Ball et al. (2019). A modified version to the original inferential procedure was realized, introducing Metropolis-Hastings (MH) as an alternative algorithm to Importance Sampling for the implementation of the two four-parameters distribution: TCEV and Kappa. Metropolis-Hastings algorithm was added because of a less fine tune than Importance Sampling was needed for TCEV and Kappa. In the analysis of chapter 6, all the reported results were obtained using MH algorithm.

5.3.1 – Metropolis-Hastings algorithm

Metropolis-Hastings algorithm (Hastings, 1970) can be considered a generalization of the classical Metropolis algorithm (Metropolis et al., 1953).
This algorithm requires the choice of a density $q(\boldsymbol{\omega}|\boldsymbol{\psi})$ from which a sample $\boldsymbol{\omega}$ (known as *proposal*) can be generated. Because $\boldsymbol{\omega}$ is to form a Markov chain, the proposal density must be dependent on the current state of the chain $\boldsymbol{\psi}$. Furthermore, it also required that the pdf $q(\boldsymbol{\omega}|\boldsymbol{\psi})$ satisfy the reversibility condition for all $(\boldsymbol{\psi}, \boldsymbol{\omega})$. This cannot occur for some couples $(\boldsymbol{\psi}, \boldsymbol{\omega})$, when

$$\pi(\boldsymbol{\psi})q(\boldsymbol{\omega}|\boldsymbol{\psi}) > \pi(\boldsymbol{\omega})q(\boldsymbol{\psi}|\boldsymbol{\omega}) \tag{5.6}$$

This means that chain is more likely to move from $\boldsymbol{\psi}$ to $\boldsymbol{\omega}$ than from $\boldsymbol{\omega}$ to $\boldsymbol{\psi}$, without satisfying the reversibility condition. In order to achieve a convenient correction, in Metropolis-Hastings algorithm is introduced a "probability of move" $\alpha(\boldsymbol{\omega}|\boldsymbol{\psi}) < 1$: its goal is to reduce the probability of the chain to move from $\boldsymbol{\psi}$ to $\boldsymbol{\omega}$. With this correction, the transition from $\boldsymbol{\psi}$ to $\boldsymbol{\omega}$ (with $\boldsymbol{\psi} \neq \boldsymbol{\omega}$) are made according to

$$p_{MH}(\boldsymbol{\omega}|\boldsymbol{\psi}) \equiv q(\boldsymbol{\omega}|\boldsymbol{\psi}) \cdot \alpha(\boldsymbol{\omega}|\boldsymbol{\psi}) \tag{5.7}$$

for $\boldsymbol{\psi} \neq \boldsymbol{\omega}$ and $\alpha(\boldsymbol{\omega}|\boldsymbol{\psi})$ has to be determined. Increase in the likelihood of the movement from $\boldsymbol{\omega}$ to $\boldsymbol{\psi}$ can be achieved by setting $\alpha(\boldsymbol{\psi}|\boldsymbol{\omega})$ equal to one, i.e. the maximum value that a probability can assume. Reversibility condition is then applied to determine the probability of move $\alpha(\boldsymbol{\omega}|\boldsymbol{\psi})$:

$$\pi(\boldsymbol{\psi})q(\boldsymbol{\omega}|\boldsymbol{\psi})\alpha(\boldsymbol{\omega}|\boldsymbol{\psi}) = \pi(\boldsymbol{\omega})q(\boldsymbol{\psi}|\boldsymbol{\omega})\alpha(\boldsymbol{\psi}|\boldsymbol{\omega}) = \pi(\boldsymbol{\omega})q(\boldsymbol{\psi}|\boldsymbol{\omega}) \tag{5.8}$$

If (5.6) is reversed, then $\alpha(\boldsymbol{\omega}|\boldsymbol{\psi})$ is set to 1 and $\alpha(\boldsymbol{\psi}|\boldsymbol{\omega})$ is derived as on eq. (5.8). The probability of move is then defined as:

$$\alpha(\boldsymbol{\omega}|\boldsymbol{\psi}) = \begin{cases} min\left[\dfrac{\pi(\boldsymbol{\omega})q(\boldsymbol{\psi}|\boldsymbol{\omega})}{\pi(\boldsymbol{\psi})q(\boldsymbol{\omega}|\boldsymbol{\psi})}, 1\right] & if\ \pi(\boldsymbol{\psi})q(\boldsymbol{\omega}|\boldsymbol{\psi}) > 0 \\ 1 & 1 \end{cases} \tag{5.9}$$

It must be remarked that proposal density $q(\boldsymbol{\omega}|\boldsymbol{\psi})$ has to be specified, and this is a crucial step, because affect the overall efficiency of the algorithm. For the purposes of this thesis, a multinormal distribution is a good choice.

Metropolis-Hastings algorithm can be summarized by the following steps:

1. Initialize $\boldsymbol{\psi}$ with a starting value $\boldsymbol{\psi}^{(0)}$;
2. for $i = 1,2,\ldots,n$
    a) sample $\boldsymbol{\omega}$ from $q\left(\cdot\,\middle|\boldsymbol{\psi}^{(i-1)}\right)$ and u from the uniform distribution
    b) if $u < \alpha\left(\boldsymbol{\omega}\middle|\boldsymbol{\psi}^{(i-1)}\right)$
        set $\boldsymbol{\psi}^{(i)} = \boldsymbol{\omega}$
    else
        set $\boldsymbol{\psi}^{(i)} = \boldsymbol{\psi}^{(i-1)}$
3. return the values $\left\{\boldsymbol{\psi}^{(1)}, \boldsymbol{\psi}^{(2)}, \ldots, \boldsymbol{\psi}^{(n)}\right\}$

5.3.2 – Applicative example: TCEV distribution

Introducing TCEV in FLIKE source structure required first to investigate its analytical properties, defining constraints and limits. It is not straightforward to manage this distribution, because of mutual constraints between parameters and the need of finding an adequate root-solver for defining quantiles, because no explicit solutions are available. A *regula falsi* algorithm was employed for this scope. However, as far as efficient could be an algorithm, computational times are higher with respect to distribution with an explicit expression for quantiles. Analytical implementation of TCEV parameters was conducted using a logarithm transformation of their parameters, and not their natural values. With this spite, new parameters were defined as:

$$Beta\ 1 = \ln(\Lambda_1) \quad Beta\ 2 = \ln(\theta_1) \quad Beta\ 3 = \ln(\Lambda_2) \quad Beta\ 4 = \ln(\theta_2)$$

An important step in obtaining a reliable tool is to fix it for diagnosing limiting cases. For example, in the case of TCEV, define a criterion for understanding when sample is Gumbel distributed is important. Simulating a Gumbel distribution time series with parameters $\zeta = 50$ and $\ln(\sigma) = 3$, computing TCEV distribution led to the following results (figs. 5.1 and 5.2):

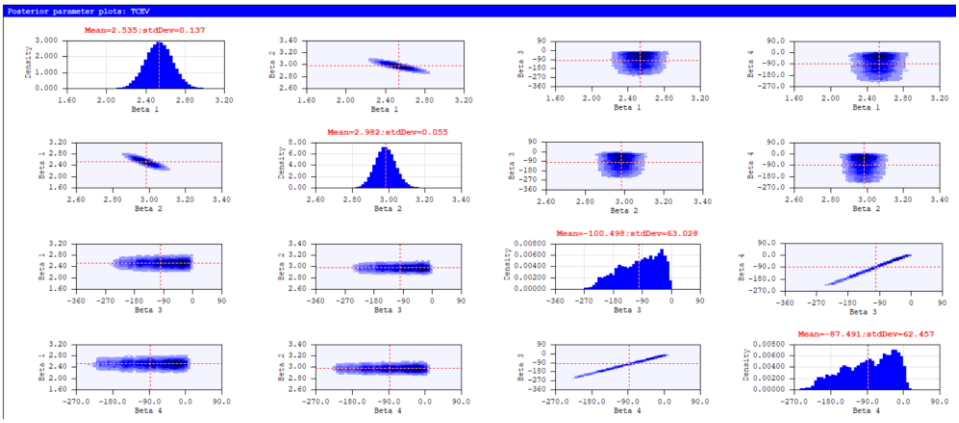

fig. 5.1 – Posterior parameters plot of TCEV (from a Gumbel distributed sample)

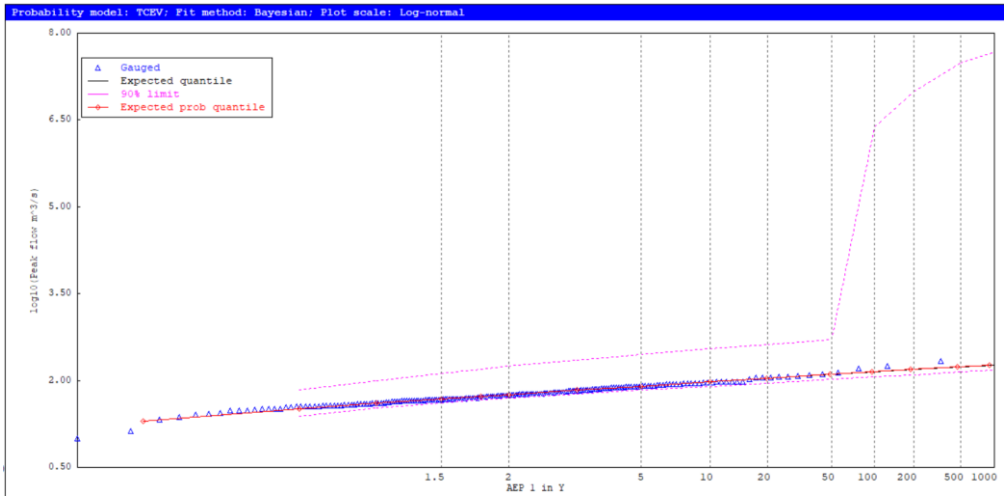

fig. 5.2 – Probability plot of TCEV (from a Gumbel distributed sample)

Parameter	Mean	Std dev	Correlation			
Beta 1	2.535	0.137	1			
Beta 2	2.982	0.055	-0.853	1		
Beta 3	-100.498	63.028	0.001	-0.006	1	
Beta 4	-87.491	62.457	0.001	-0.005	0.996	1

table 5.1 – A summary of Posterior Moments from Metropolis-Hastings for TCEV

Analysis of posterior plot is very interesting, because in can be noted that values of $Beta\ 3$ and $Beta\ 4$ assumes couples such as to make the second component considerably smaller than the first one. Overparametrization is reflected in fig. 5.2, where for higher quantiles the upper 95% confidence limit is extremely high. Furthermore, comparing traditional expression of Gumbel and TCEV (with only the first component) distributions, gives:

$$e^{-e^{-\left(\frac{x-\zeta}{\sigma}\right)}} = e^{\left[-\Lambda_1 e^{-\left(\frac{x}{\theta_1}\right)}\right]}$$

Considering that $Beta\ 2 = \theta_1 = \sigma$ and $\Lambda_1 = e^{\frac{\zeta}{\sigma}}$, can be noted that, from table 5.1:

$$Beta\ 2 = 2.982 \qquad \ln(\Lambda_1) = e^{\frac{\zeta}{\sigma}} = 2.526$$

5.3.3 – Applicative example: Kappa distribution

Unlike TCEV, Kappa has a more tractable analytical expression. This reflect in a faster computational step and a clearer interpretation of results.
As shown in paragraph 4.2.1, Kappa has different limit distributions. Reporting results for all cases was retained not necessary, and the more evident degenerate case is reported, i.e. when the parent is a GEV distribution.
Assuming a sample GEV distributed with parameters $\zeta = 50$, $\ln(\sigma) = 2$ and $\varepsilon = 0.3$, its implementation in FLIKE gave these results:

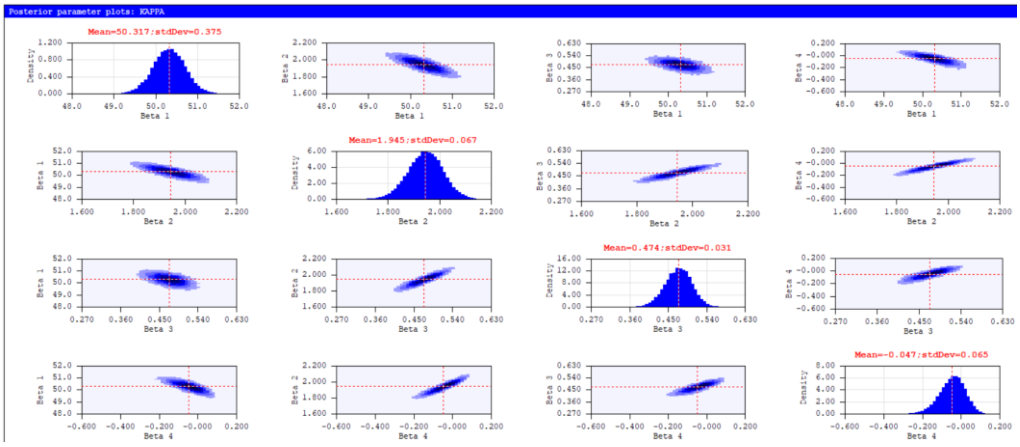

fig. 5.3 – Posterior parameters plot of Kappa (from a GEV distributed sample)

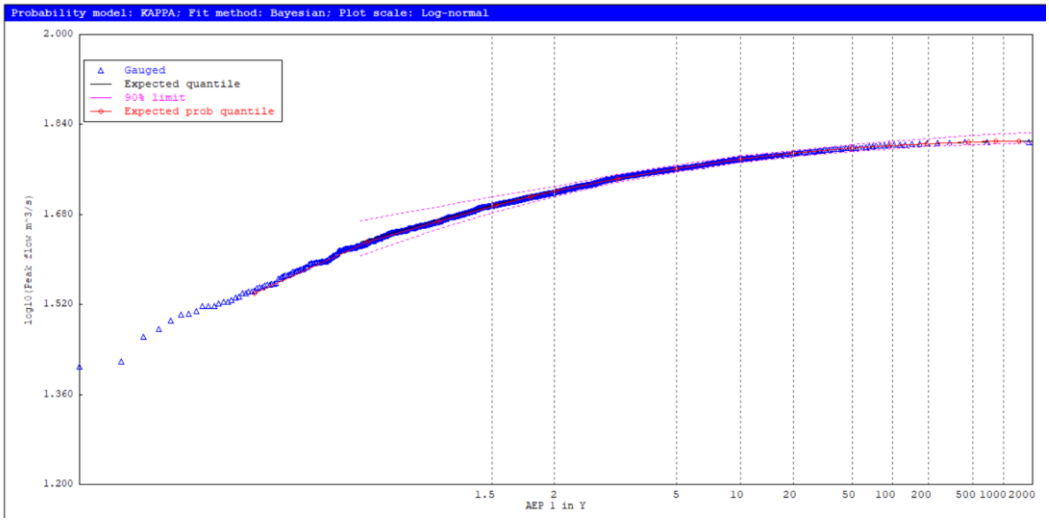

fig. 5.4 – Probability plot of Kappa (from a GEV distributed sample)

Parameter	Mean	Std dev	Correlation			
Location ζ	50.317	0.375	1			
$\ln(\sigma)$, scale	1.945	0.067	-0.801	1		
Shape ε	0.474	0.031	-0.55	0.914	1	
h	-0.047	0.065	-0.781	0.928	0.84	1

table 5.2 – A summary of Posterior Moments from Metropolis-Hastings for Kappa

Figs. 5.3 and 5.4 shows a well-behaved posterior and probability plots. Table 5.2 gave a correct confirmation about the nature of the parent distribution.

Examples of paragraphs 5.3.3.1 and 5.3.3.2 were firstly reported for illustrating the operative of the work on FLIKE software and describing the obtainable outputs. Secondarily, some applicative cases were shown, that allows to appreciate the amount on information that can be provided by the posterior

distribution, especially in terms of quantification of uncertainty in parameters and quantiles estimates.

5.4 – A measure for discerning the presence of two populations

Two Component Extreme Value distribution has a very specific nature, due to its ability in contemplating different components of the same natural phenomenon (e.g. floods, rainfall). This asset represents a relevant resource in the topic of model selection. In fact, when different distributions are fitted to data, in the TCEV case it is not enough the simple output of the goodness-of-fit or criteria applied. A further step should require the detection of the presence of these two components.

Plot analysis of the posterior distribution of parameters can give some signs about the nature of the data (as shown in the previous section), but distinguishing between to distributions that have a similar fit requires more detailed investigations. A relevant consequence is a better comprehension of the underlying phenomena, which can be really decisive. This will be corroborated by some practical evidences that will be illustrated and discussed in the next chapter.

Moving from this background, a metric for detecting the presence of two component in a flood time series moving from the samples of posterior distribution has been developed, and is here illustrated.

TCEV distribution has this CDF (chapter 1):

$$F(x|\theta) = e^{\left[-\lambda_1 e^{\left(-\frac{x}{\theta_1}\right)} - \lambda_2 e^{\left(-\frac{x}{\theta_2}\right)}\right]}$$

The relative importance of two components has been discussed about the condition of degeneration of TCEV into a Gumbel distribution. Proposed metric, named TC_{ratio},is obtained by these steps:

1. define some 1 in years AEP, T (T = 1.05, 1.1, 1.25, 2,5,10,50,100);
2. evaluate related non-exceedance probability 1-1/T
3. for each of the sampled parameters evaluate quantiles corresponding to these probabilities
4. for each T, define for each set of sampled parameters:

   - $fv1 = \left| -\theta_1 e^{\left(-\frac{x}{\Lambda_1}\right)} \right|$
   - $fv2 = \left| -\theta_2 e^{\left(-\frac{x}{\Lambda_2}\right)} \right|$
   - $TC_{ratio} = \frac{fv1}{fv1+fv2}$

   where *x* is the quantile estimated at point 3
5. plot the histogram of frequencies of ratio for all T. If TC_{ratio} is close to one, then there is a marked influence of the first component; otherwise, second component is predominant.

An example of this application is reported in fig. 5.5.

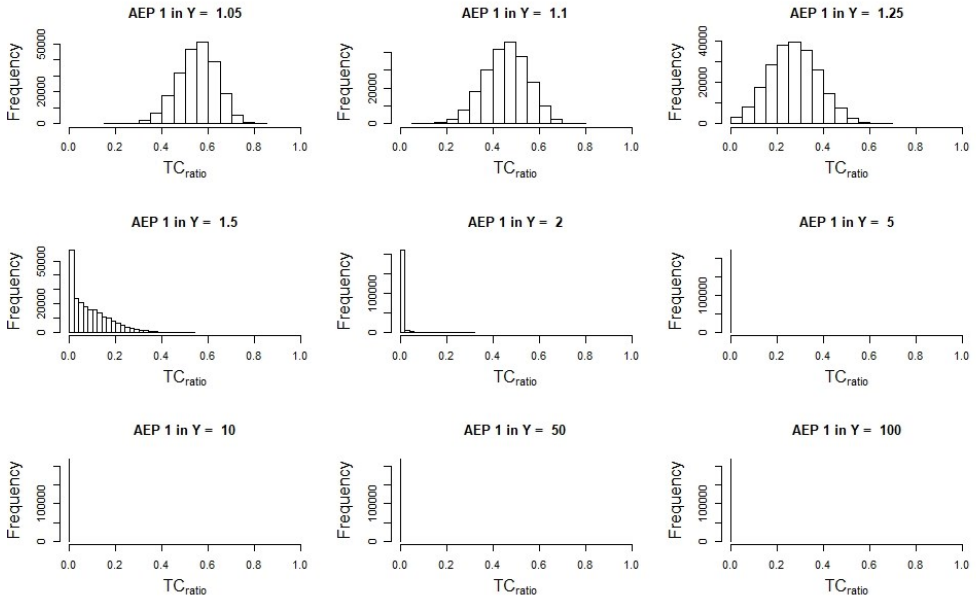

fig. 5.5 - TC_{ratio} plot with the presence of two components

In this case can be noted how histograms are modified moving from AEP 1 in Y equal to 1.05 to 100. In the first case, the presence of the first component (represented by $fv1$) can be easily detected and progressively disappear.
When, instead, the presence of only one component is noticed, plots are of the type illustrated in fig. 5.6.

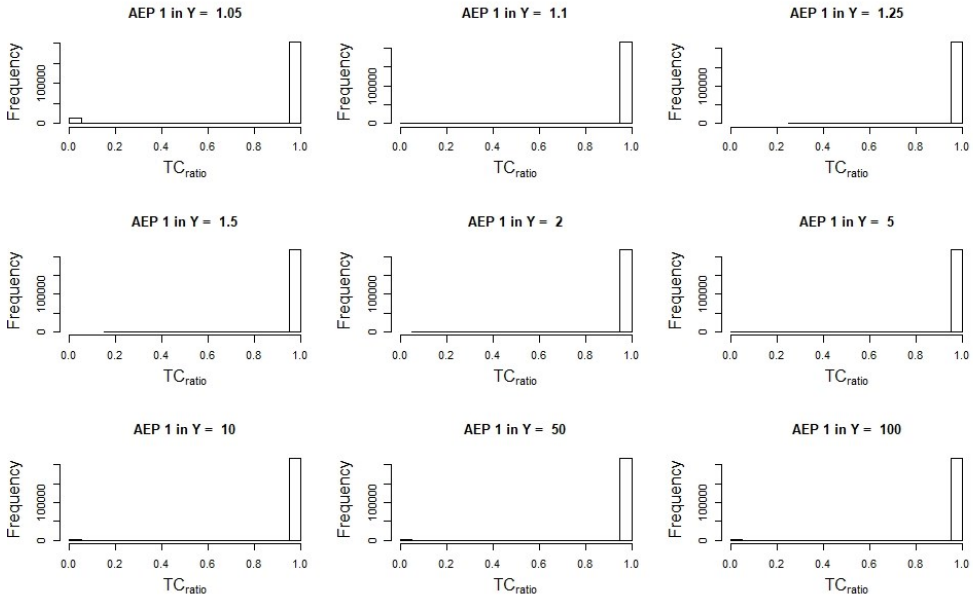

fig. 5.6 - TC_{ratio} plot with the presence one components

CHAPTER 6 – CASES STUDY

6.1 - Introduction

Investigations on the best fitting distribution for at-site flood data in Australia were conducted since '70s, when Conway (1970) in his master thesis studied floods regime for several coastal streams of New South Wales. Following studies were realized (McMahon and Srikanthan, 1981; Nathan and Weinmann, 1991; Vogel et al., 1993a) and concluded that not only a probability distribution can be advised for performing at-site FFA, but there are different candidate distributions that can give different fits according to the location of sites. This is definitely a meaningful finding, also because the extension and the different variety of Australian climates. Namely, Nathan and Weinmann (1991) and Vogel et al. (1993) found GEV and Generalized Pareto as the distributions with the best fit to their case studies. However, should be noted that Australian Rainfall Runoff (ARR), 1987, recommended Log Pearson Type 3 (LP3) distribution for at-site FFA in Australia. Haddad and Rahman (2008) analyzed 18 catchments in southeast Australia and found advisable the use of GEV. Same authors (2010) noted how two-parameter Log-Normal distribution gave the best performances for sites in Tasmania.

Rahman et al. (2013) realized a study for detecting the appropriate choice of a distribution of flood frequency analysis in the entire Australia, comparing several measures for testing goodness-of-fit of each distribution. Rahman et al. (2014) investigated the impact of outliers for flood frequency analysis in some sites of eastern Australia comparing original Grubbs and Beck and multiple Grubbs and Beck tests.

Goal of this chapter is to give a contribution to these previous studies introducing in the analysis two four-parameter distribution, i.e. TCEV and Kappa. These two distributions were compared with Log-Pearson type 3 and Generalized Extreme Value. A kind of hierarchical analysis for testing the suitability of these

distributions was performed, starting from a critical observation of the "full" LMRD and the use of the TC_{ratio} test. Therefore, visual analysis and some parametric tests were applied.
A focus on the role of uncertainty for model selection is then discussed. Finally, results of these studies are compared to previous literature findings. A single Italian case study is also proposed, in order to highlight the nature of TCEV distribution.
It is useful to remark that, in the spite of FLIKE options, all at-site Bayesian analysis were performed with no prior information. Furthermore, all quantiles are in m^3/s.

6.2 – Study area and dataset

Annual maximum flood data for New South Wales (NSW), Queensland (QLD) and Nothern Territory (NT) were employed in this study. Data were available in the framework of *Australian Rainfall and Runoff (ARR) Project 5 Regional Flood Methods*[9], are diffused online by the Bureau of Meteorology (BoM) of Australian Government and are freely available at the website http://www.bom.gov.au/waterdata/. Not all the records used in Project 5 (P5) were available. For New South Wales a shorter dataset was analyzed.
All available data are reported in Appendix B. Into this dataset, only sites with more than 40 records were selected. In the next three sub-paragraphs their location for each region will be described and displayed.

6.2.1 – New South Wales

For New South Wales records were available only for 88 sites over the 176 of P5 for NSW and ACT (Appendix B.1). Location of 7 selected sites is illustrated in fig.

[9] http://www.arr-software.org/pdfs/ARR_Project05_Stage3_Database%20report.pdf

6.1. In table 1, Station ID[10], Station and River name, Latitude and Longitude, Period of Record and Size for each site are reported. In the first column, a number for the identification of the site both on the map and on LMRD is introduced.

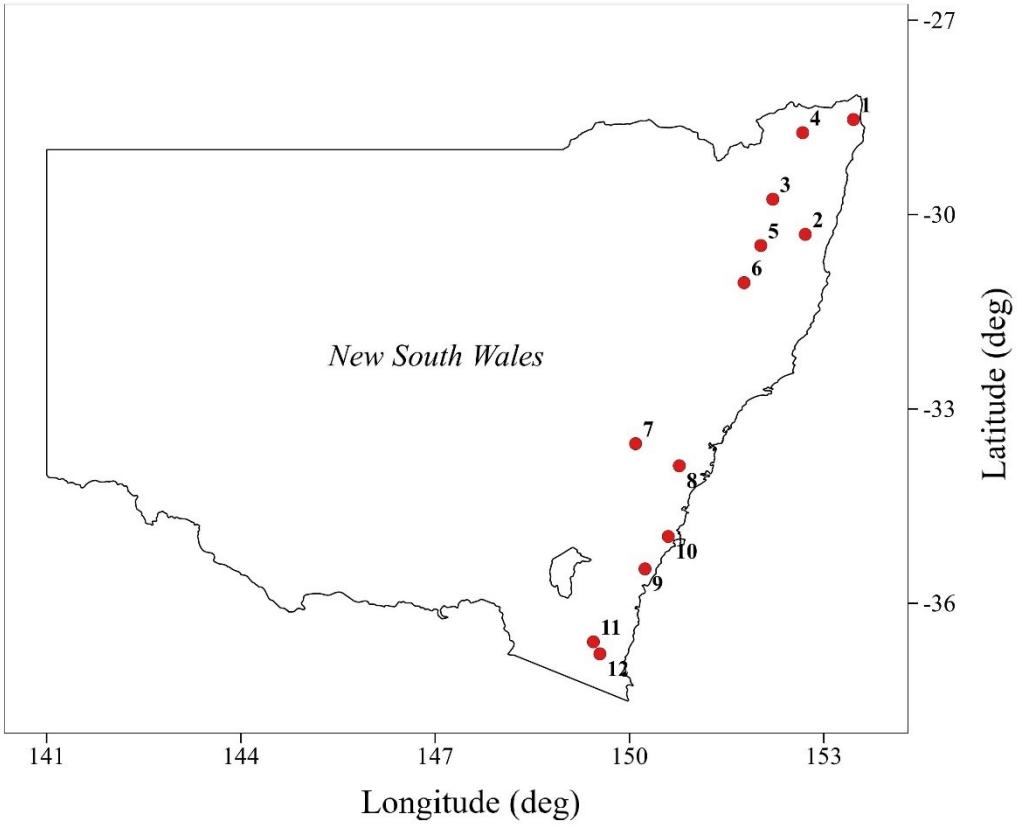

fig. 6.1 – Location of selected sites for NSW

[10] The same in BoM.

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
1	202001	Durrumbul (Sherrys Crossing)	Brunswick	-28.5333	153.4567	1972-2011	40
2	204017	Dorrigo No.2 & No.3	Bielsdown Ck	-30.3067	152.7133	1972-2011	40
3	204034	Newton Boyd	Henry	-29.7633	152.2117	1972-2011	40
4	204043	Bonalbo	Peacock Ck	-28.7367	152.6733	1961-2011	51
5	206014	Coninside	Wollomombi	-30.4783	152.0267	1955-2011	57
6	206018	Apsley Falls	Apsley	-31.0517	151.7683	1961-2011	51
7	212011	Lithgow	Coxs	-33.5367	150.0933	1962-2011	50
8	212320	Mulgoa Rd	South Ck	-33.8783	150.7683	1972-2011	40
9	216002	Brooman	Clyde	-35.4700	150.2383	1961-2011	51
10	216004	Falls Ck	Currambene Ck	-34.9700	150.5983	1971-2010	40
11	219001	Brown Mountain	Rutherford Ck	-36.5967	149.4417	1949-2010	62
12	219006	Tantawangalo Mountain (Dam)	Tantawangalo Ck	-36.7817	149.5417	1952-2010	59

table 6.1 – Selected sites for NSW

6.2.2 – Queensland

For Queensland, records were available for all 176 sites (Appendix B.2). As for New South Wales, their position and information are reported in fig. 6.2 and in table 6.2 respectively.

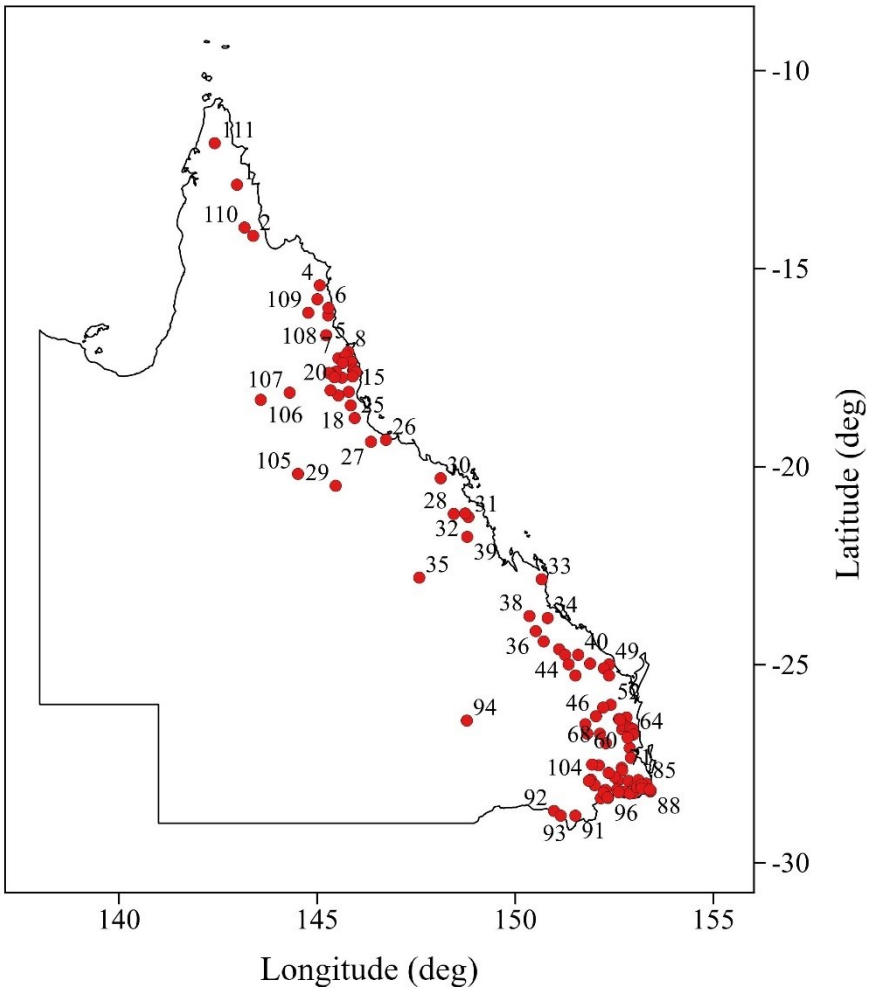

fig. 6.2 – Location of selected sites for Queensland

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
1	102101	Fall Ck	Pascoe	-12.88	142.98	1968-2011	44
2	104001	Telegraph Rd	Stewart	-14.17	143.39	1970-2011	42
3	105105	Developmental Rd	East Normanby	-15.77	145.01	1970-2011	42
4	107001	Flaggy	Endeavour	-15.42	145.07	1959-2011	53
5	108002	Bairds	Daintree	-16.18	145.28	1969-2011	43
6	108003	China Camp	Bloomfield	-15.99	145.29	1971-2011	41
7	110003	Picnic Crossing	Barron	-17.26	145.54	1926-2011	86
8	111001	Gordonvale	Mulgrave	-17.10	145.79	1917-1972	43
9	111005	The Fisheries	Mulgrave	-17.19	145.72	1967-2011	45
10	111105	The Boulders	Babinda Ck	-17.35	145.87	1967-2011	45
11	112002	Nerada	Fisher Ck	-17.57	145.91	1929-2011	83
12	112003	Glen Allyn	North Johnstone	-17.38	145.65	1959-2011	53
13	112004	Tung Oil	North Johnstone	-17.55	145.93	1967-2011	45
14	112101	U/S Central Mill	South Johnstone	-17.61	145.98	1917-2011	95
15	112102	Upper Japoonvale	Liverpool Ck	-17.72	145.90	1971-2012	42
16	113004	Powerline	Cochable Ck	-17.75	145.63	1967-2011	45
17	114001	Upper Murray	Murray	-18.11	145.80	1971-2011	41
18	116008	Abergowrie	Gowrie Ck	-18.45	145.85	1954-2004	58
19	116010	Blencoe Falls	Blencoe Ck	-18.20	145.54	1961-2011	51
20	116011	Ravenshoe	Millstream	-17.60	145.48	1963-2011	49
21	116012	8.7KM	Cameron Ck	-18.07	145.34	1962-2011	50

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
22	116013	Archer Ck	Millstream	-17.65	145.34	1962-2011	50
23	116014	Silver Valley	Wild	-17.63	145.30	1962-2011	50
24	116015	Wooroora	Blunder Ck	-17.74	145.44	1967-2011	45
25	116017	Running Ck	Stone	-18.77	145.95	1971-2011	41
26	118101	Gleesons Weir	Ross	-19.32	146.74	1916-1960	45
27	120102	Keelbottom	Keelbottom Ck	-19.37	146.36	1968-2011	44
28	120216	Old Racecourse	Broken	-21.19	148.45	1970-2011	42
29	120307	Pentland	Cape	-20.48	145.47	1970-2011	42
30	121001	Ida Ck	Don	-20.29	148.12	1958-2011	54
31	125002	Sarich's	Pioneer	-21.27	148.82	1961-2011	51
32	125004	Gargett	Cattle Ck	-21.18	148.74	1968-2011	44
33	129001	Byfield	Waterpark Ck	-22.84	150.67	1953-2011	59
34	130004	Old Stn	Raglan Ck	-23.82	150.82	1964-2011	48
35	130207	Clermont	Sandy Ck	-22.80	147.58	1966-2011	46
36	130319	Craiglands	Bell Ck	-24.15	150.52	1961-2011	51
37	130321	Mt. Kroombit	Kroombit Ck	-24.41	150.72	1964-2004	41
38	130335	Wura	Dee	-23.77	150.36	1972-2011	40
39	130413	Braeside	Denison Ck	-21.77	148.79	1972-2011	40
40	135002	Springfield	Kolan	-24.75	151.59	1966-2011	46
41	135004	Dam Site	Gin Gin Ck	-24.97	151.89	1966-2011	46
42	136006	Dam Site	Reid Ck	-25.27	151.52	1966-2011	46
43	136108	Upper Monal	Monal Ck	-24.61	151.11	1963-2011	47
44	136111	Dakiel	Splinter Ck	-24.75	151.26	1966-2011	47

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
45	136112	Yarrol	Burnett	-24.99	151.35	1966-2011	46
46	136202	Litzows	Barambah Ck	-26.30	152.04	1921-2011	91
47	136203	Brooklands	Barker Ck	-26.74	151.82	1941-2011	71
48	136301	Weens Br	Stuart	-26.50	151.77	1936-2011	76
49	137001	Elliott	Elliott	-24.99	152.37	1949-2011	63
50	137101	Burrum HWY	Gregory	-25.09	152.24	1967-2011	45
51	137201	Bruce HWY	Isis	-25.27	152.37	1967-2011	45
52	138002	Brooyar	Wide Bay Ck	-26.01	152.41	1910-2011	102
53	138010	Kilkivan	Wide Bay Ck	-26.08	152.22	1910-2011	102
54	138101	Kenilworth	Mary	-26.60	152.73	1921-1973	53
55	138102	Zachariah	Amamoor Ck	-26.37	152.62	1921-2011	91
56	138104	Kidaman	Obi Obi Ck	-26.63	152.77	1921-1963	43
57	138106	Baroon Pocket	Obi Obi Ck	-26.71	152.86	1941-1986	46
58	138107	Cooran	Six Mile Ck	-26.33	152.81	1948-2011	64
59	138110	Bellbird Ck	Mary	-26.63	152.70	1960-2011	52
60	138111	Moy Pocket	Mary	-26.53	152.74	1964-2011	48
61	138113	Hygait	Kandanga Ck	-26.39	152.64	1972-2011	40
62	141001	Kiamba	South Maroochy	-26.59	152.90	1938-2011	73
63	141003	Warana Br	Petrie Ck	-26.62	152.96	1959-2011	53
64	141006	Mooloolah	Mooloolah	-26.76	152.98	1972-2011	40
65	142001	Upper Caboolture	Caboolture	-27.10	152.89	1966-2011	46
66	142202	Drapers Crossing	South Pine	-27.35	152.92	1966-2011	46
67	143010	Boat Mountain	Emu Ck	-26.98	152.29	1967-2011	45

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
68	143015	Damsite	Cooyar Ck	-26.74	152.14	1969-2011	43
69	143102	Kalbar No.2	Warrill Ck	-27.92	152.60	1913-1970	55
70	143107	Walloon	Bremer	-27.60	152.69	1962-2011	50
71	143108	Amberley	Warrill Ck	-27.67	152.70	1962-2011	50
72	143110	Adams Br	Bremer	-27.83	152.51	1972-2011	40
73	143203	Helidon Number 3	Lockyer Ck	-27.54	152.11	1927-2011	84
74	143209	Mulgowie2	Laidley Ck	-27.73	152.36	1958-2011	44
75	143303	Peachester	Stanley	-26.84	152.84	1928-2011	84
76	145002	Lamington No.1	Christmas Ck	-28.24	152.99	1910-1954	45
77	145003	Forest Home	Logan	-28.20	152.77	1918-2011	60
78	145010	5.8KM Deickmans Br	Running Ckreek	-28.25	152.89	1966-2011	46
79	145011	Croftby	Teviot Brook	-28.15	152.57	1967-2011	45
80	145012	The Overflow	Teviot Brook	-27.93	152.86	1967-2009	43
81	145018	Up Stream Maroon Dam	Burnett Ck	-28.22	152.61	1971-2011	41
82	145101	Lumeah Number 2	Albert	-28.06	153.04	1911-2011	100
83	145102	Bromfleet	Albert	-27.91	153.11	1919-2011	93
84	145103	Good Dam Site	Cainbable Ck	-28.09	153.08	1963-2011	49
85	146002	Glenhurst	Nerang	-28.00	153.31	1920-2011	92
86	146003	Camberra Number 2	Currumbin Ck	-28.20	153.41	1928-1982	55
87	146010	Army Camp	Coomera	-28.03	153.19	1963-2011	49
88	146012	Nicolls Br	Currumbin Ck	-28.18	153.42	1971-2011	41
89	146014	Beechmont	Back Ck	-28.12	153.19	1972-2011	40
90	146095	Tallebudgera Ck Rd	Tallebudgera Ck	-28.15	153.40	1971-2011	41

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
91	416303	Clearview	Pike Ck	-28.81	151.52	1935-1987	48
92	416305	Beebo	Brush Ck	-28.69	150.98	1969-2011	43
93	416312	Texas	Oaky Ck	-28.81	151.15	1970-2011	42
94	422210	Tabers	Bungil Ck	-26.41	148.78	1967-2011	45
95	422302	Killarney	Spring Ck	-28.35	152.34	1910-1955	46
96	422303	Killarney	Spring Ck South	-28.36	152.34	1910-1955	46
97	422304	Elbow Valley	Condamine	-28.37	152.16	1916-1972	57
98	422306	Swanfels	Swan Ck	-28.16	152.28	1920-2011	92
99	422307	Kings Ck	Kings Ck	-27.90	151.91	1921-1966	43
100	422313	Emu Vale	Emu Ck	-28.23	152.23	1948-2011	64
101	422319	Allora	Dalrymple Ck	-28.04	152.01	1969-2011	43
102	422321	Killarney	Spring Ck	-28.35	152.33	1960-2011	52
103	422326	Cranley	Gowrie Ck	-27.52	151.94	1970-2011	42
104	422334	Aides Br	Kings Ck	-27.93	151.86	1970-2011	42
105	915011	Mt Emu Plains	Porcupine Ck	-20.18	144.52	1972-2011	40
106	917104	Roseglen	Etheridge	-18.31	143.58	1967-2011	45
107	917107	Mount Surprise	Elizabeth Ck	-18.13	144.31	1969-2011	43
108	919005	Fonthill	Rifle Ck	-16.68	145.23	1969-2011	42
109	919201	Goldfields	Palmer	-16.11	144.78	1968-2011	44
110	922101	Racecourse	Coen	-13.96	143.17	1968-2011	43
111	926002	Dougs Pad	Dulhunty	-11.83	142.42	1971-2011	41

table 6.2 - *Selected sites for Queensland*

6.2.3 – Northern Territory

Database for Northern Territory was composed by 50 sites. (Appendix B.3). In fig. 6.3 is reported their location on the map, while in table 6.3 information about are provided.

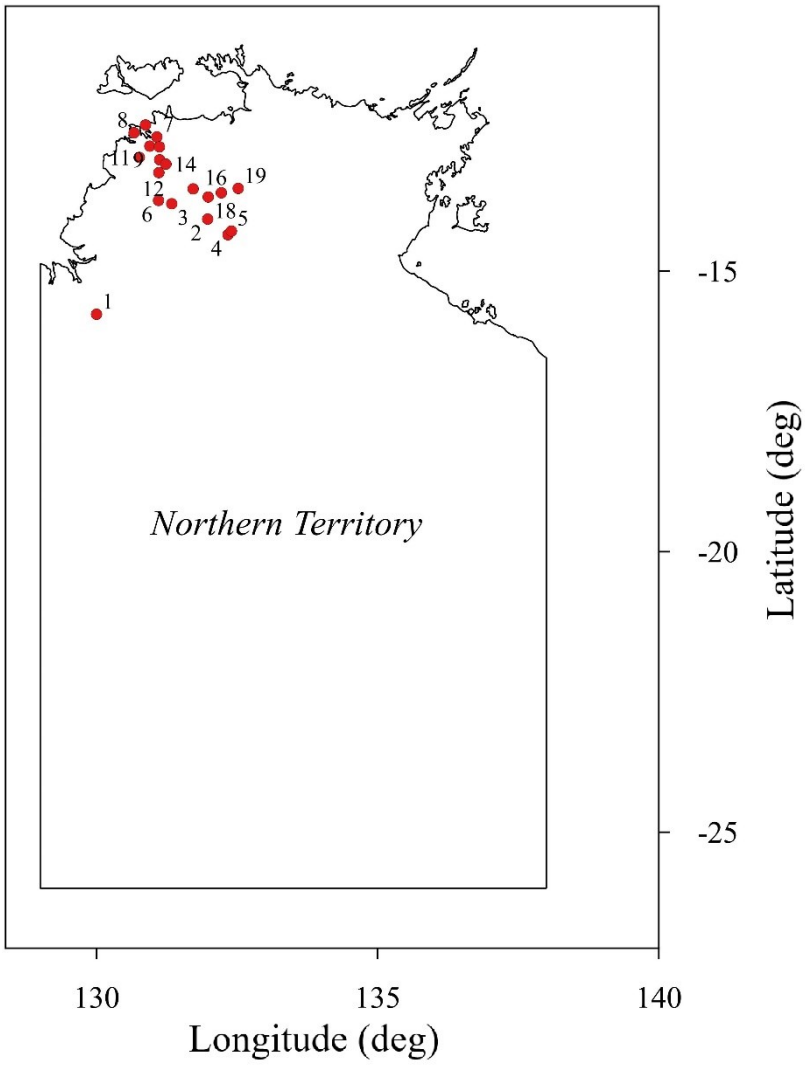

fig. 6.3 – Location of selected sites for Northern Territory

	Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record	Size
1	G8110004	Victoria HWY	East Baines	-15.7667	130	1963 - 2008	49
2	G8140008	Old Railway Br	Fergusson	-14.07	131.9767	1958 - 2011	53
3	G8140063	D/S Old Douglas H/S	Douglas	-13.7967	131.3383	1958 - 2011	54
4	G8140158	Dam Site	McAdden Ck	-14.3483	132.3383	1964 - 2011	45
5	G8140159	Waterfall View	Seventeen Mile C	-14.2833	132.4	1963 - 2008	46
6	G8140161	Tipperary	Green Ant Ck	-13.7383	131.1033	1966 - 2011	46
7	G8150018	Stuart HWY	Elizabeth	-12.605	131.0733	1955 - 2011	57
8	G8150096	Cox Peninsula	Carawarra Ck	-12.5317	130.6683	1966 - 2011	45
9	G8150098	Tumbling Waters	Blackmore	-12.77	130.9483	1960 - 2010	51
10	G8150127	D/S McMillans Rd	Rapid Ck	-12.3933	130.8717	1964 - 2011	48
11	G8150180	Gitchams	Finniss	-12.97	130.7617	1961 - 2007	47
12	G8170002	Railway Br	Adelaide	-13.2417	131.1083	1954 - 2007	54
13	G8170066	Stuart HWY	Coomalie Ck	-13.0133	131.1233	1958 - 2010	53
14	G8170084	Tortilla Flats	Adelaide	-13.09	131.235	1960 - 2011	52
15	G8170085	Stuart HWY	Acacia Ck	-12.7833	131.12	1964 - 2011	48
16	G8180026	El Sherana Rd Crossing	Mary	-13.6017	132.22	1962 - 2011	48
17	G8180069	near Burrundie	McKinlay	-13.5317	131.7183	1959 - 2009	51
18	G8180252	D/S El Sherana Rd	Harriet Ck	-13.6767	131.9867	1965 - 2010	46
19	G8200045	El Sherana (C)	South Alligator	-13.5233	132.52	1958 - 2009	52

table 6.3 - *Selected sites for Northern Territory*

6.2.4 – Climate

Australia is a wide country, in which there is a coexistence of several climate conditions. This spatial heterogeneity can be found also inside the single states. In this paragraph, a short description of the main features of climate for the three investigated states is reported.

Climate across New South Wales ranges from the hot, dry, continental conditions in the west, to the subtropical, wet conditions in the northeast, to the alpine cold of the southeast. A marked variability can be found also in the average annual rainfall, which can assume values more than five times as much rain falling along the coast as in the west. Climate, furthermore, is affected by the influence of the Great Dividing Range, whose impacts influences mainly the distribution of rainfall and results in four distinct climate zones:

i. the Coast: is characterized by a mild climate, which ranges from subtropical near the Queensland border to the cool temperate climate of the south.
ii. The Ranges and Tablelands of the Great Dividing Range: climate is temperate in the north and cool temperate in the south. Areas above 1200 m have an Alpine climate.
iii. The Western Slopes and Plains: climate is generally hot and dry with a cooler winter.
iv. The Arid Plains: continuing west into the central and far northwest regions of the State, the climate becomes hotter and semi-arid to arid.

Queensland state is characterized by a considerable climate variation. In the inland west there are low rainfall and hot summers, a monsoon season in the north, and warm temperate conditions along the coastal strip contrast. Coastal strip climate is influenced by Coral and Tasman Seas, which in general, keep the region free from extremes of temperature and provide moisture for rainfall. In this region the annual median rainfall ranges between 1000 and 1600 mm. increasing to over 3200 mm along parts of the north Queensland coast.

In the Northern Territory, instead, two climate zones can be distinguished:

i. the northern end: is characterized by a tropical climate and a wet (with tropical cyclones and monsoon rains) and a dry season.
ii. the central region: which is located in the desertic center of Australia. There is a semi-arid climate with little rain.

6.3 – Applications and Results

6.3.1 – Preliminary results

In order to exploit the reliability of incorporating the use of TCEV and Kappa distributions in at-site flood frequency analysis, some preliminary investigations were made.
The first, described in sub-paragraph 6.3.1.1, involves the use of L-Moments Ratio Diagram for a visual comparison between sample and theoretical L-Moments for candidate distributions. For a better visualization, Kappa area was neglected in all figures. However, due to the relationships with other three-parameter distribution represented on the plot, it will be possible to discuss the suitability of this distribution for at-site analysis.
In sub-paragraph 6.3.1.2 the outcome of TC_{ratio} test is illustrated, while in 6.3.1.3 the outcome of the visual analysis of results is commented. Finally, in last sub-paragraph the application of AIC and BIC parametric tests is reported.

6.3.1.1 – L-Moments Ratio Diagrams

The use of L-Moments Ratio Diagram is considered an affordable preliminary assessment for evaluating the suitability of different distributions to sample data. Because of the large number of sites, results will be displayed in fig. 6.4, 6.6 and 6.8 for each region and only for investigated sites. In order to achieve a comparison with LP3, in figures 6.5, 6.7 and 6.9 τ_3 and τ_4 of the logarithms of annual maximum are plotted with theoretical curve for P3 distribution (Vogel et al., 1993a).

In appendix B LMRDs for all available sites in respective regions are plotted.

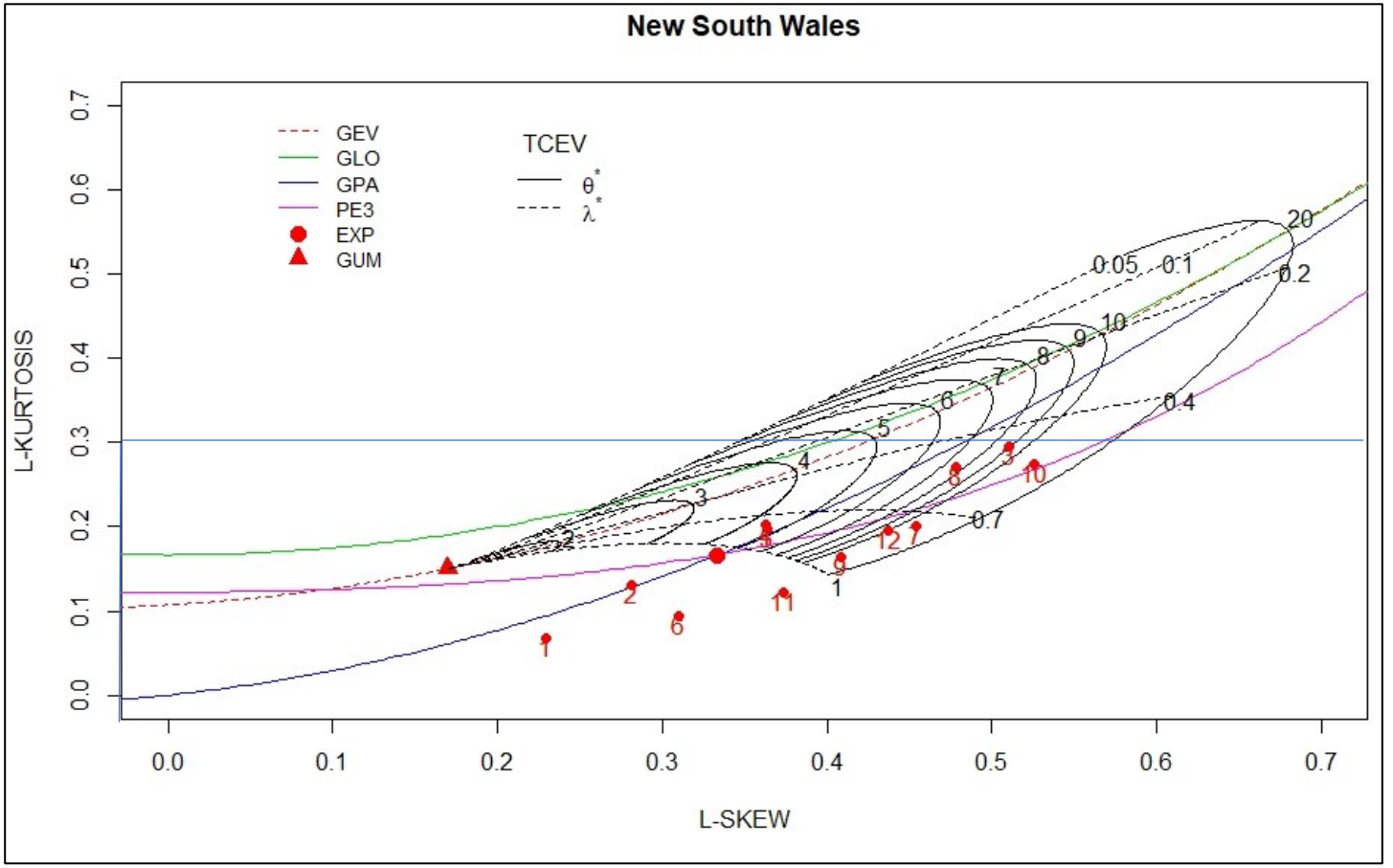

fig. 6.4 – L-Moments Ratio Diagram for selected sites in New South Wales

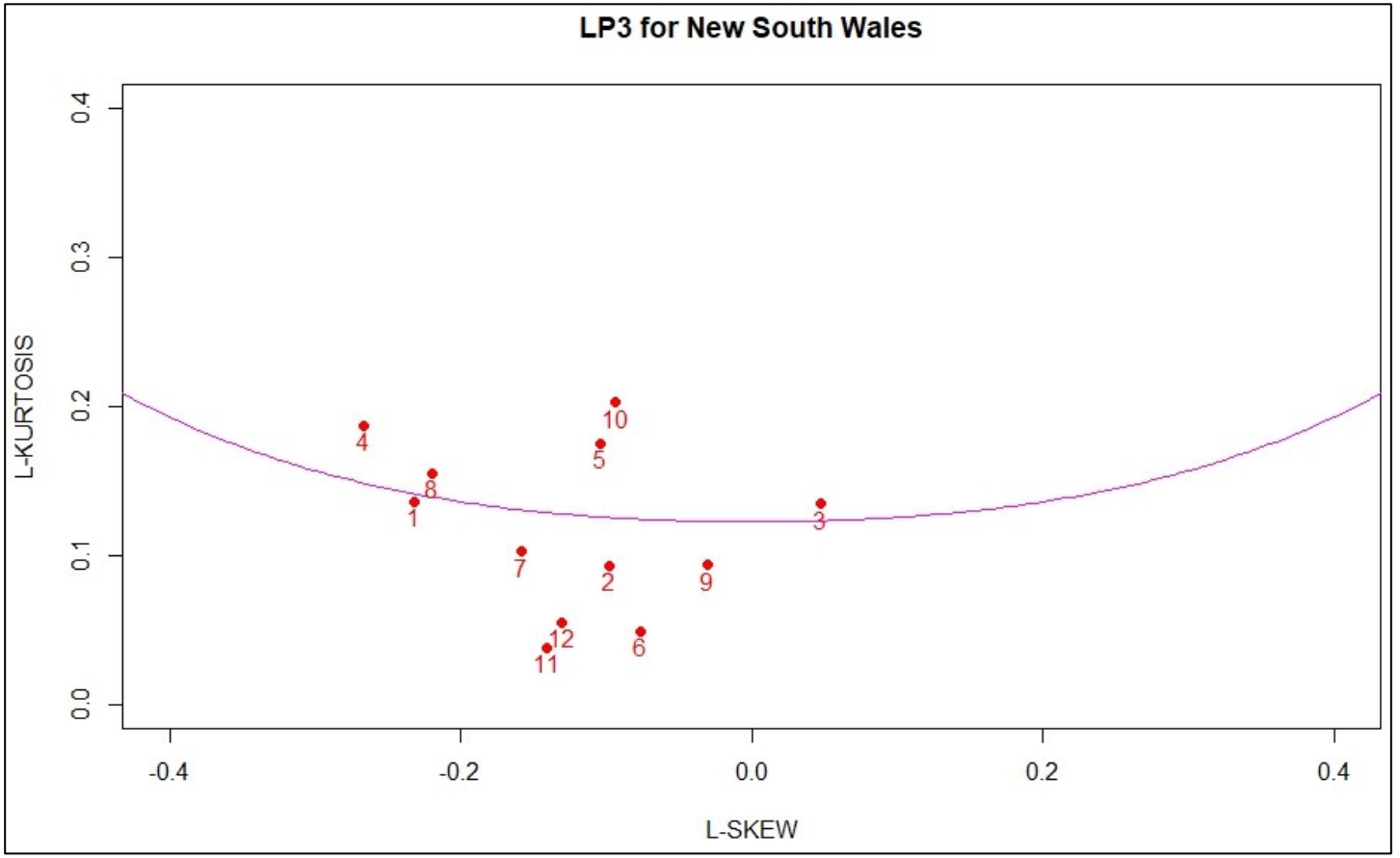

fig. 6.5 – L-Moments Ratio Diagram for Log-Pearson Type 3 for selected sites in New South Wales

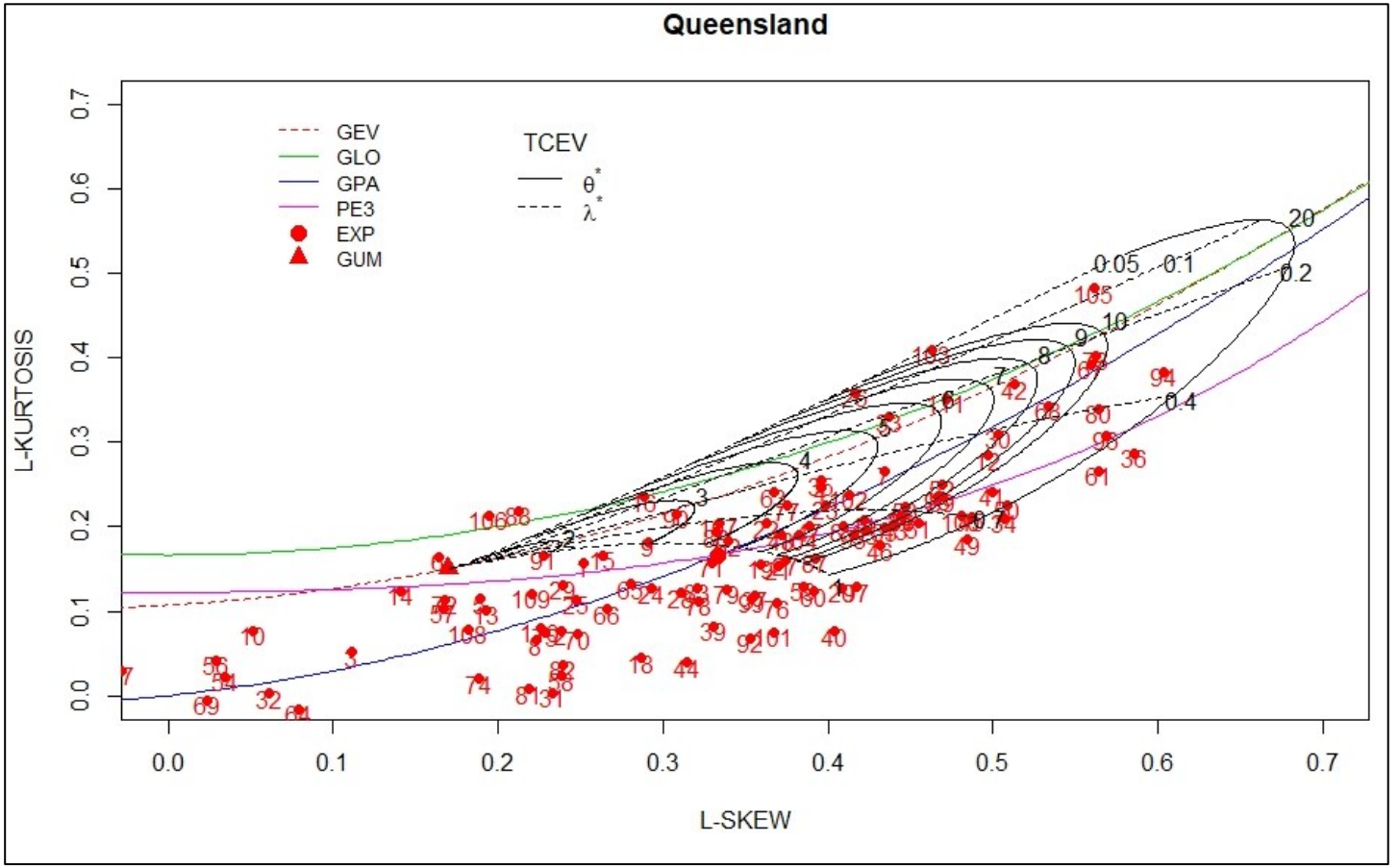

fig. 6.6 – L-Moments Ratio Diagram for selected sites in Queensland

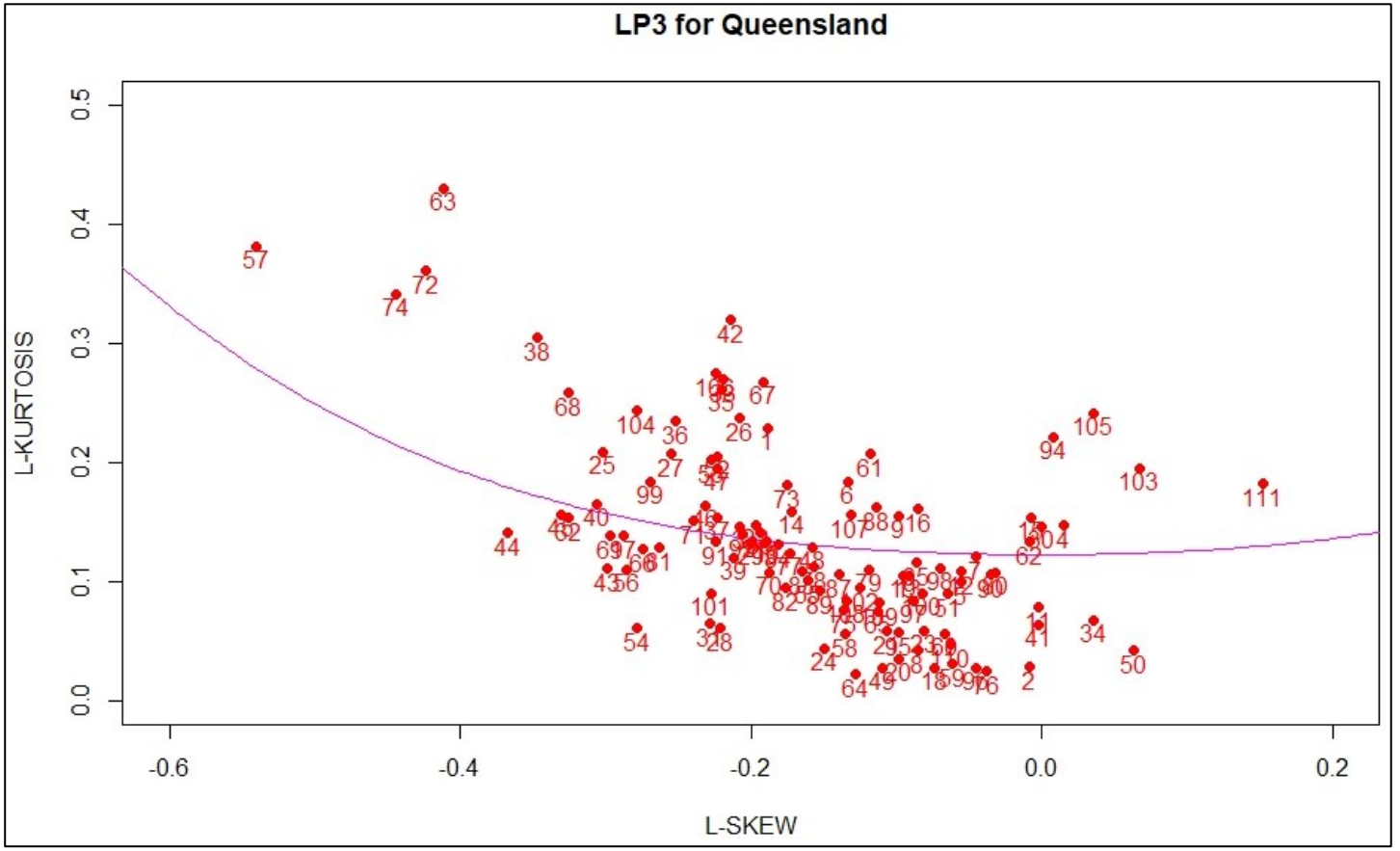

fig. 6.7 – L-Moments Ratio Diagram for Log-Pearson Type 3 for selected sites in Queensland

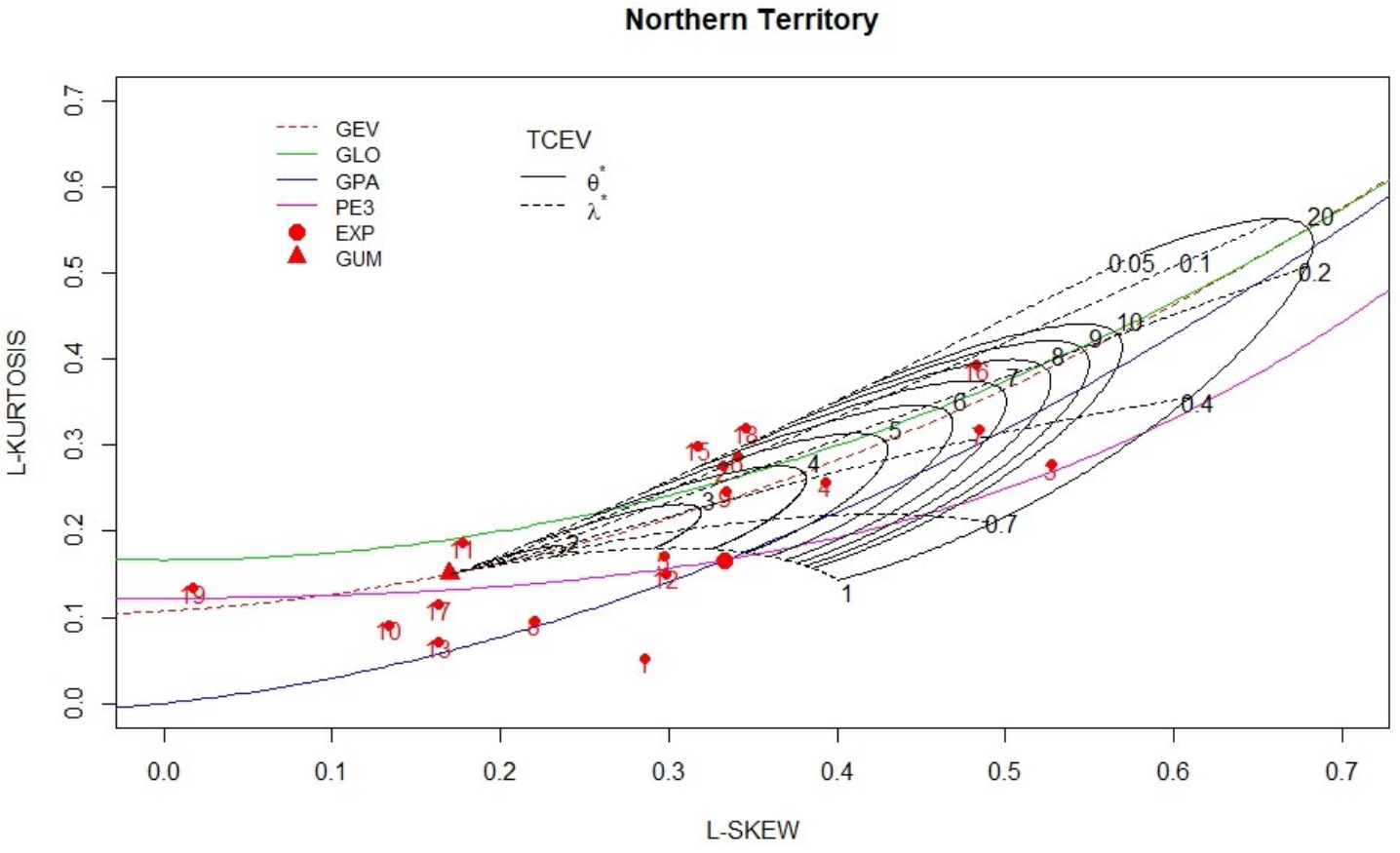

fig. 6.8 – L-Moments Ratio Diagram for selected sites in Northern Territory

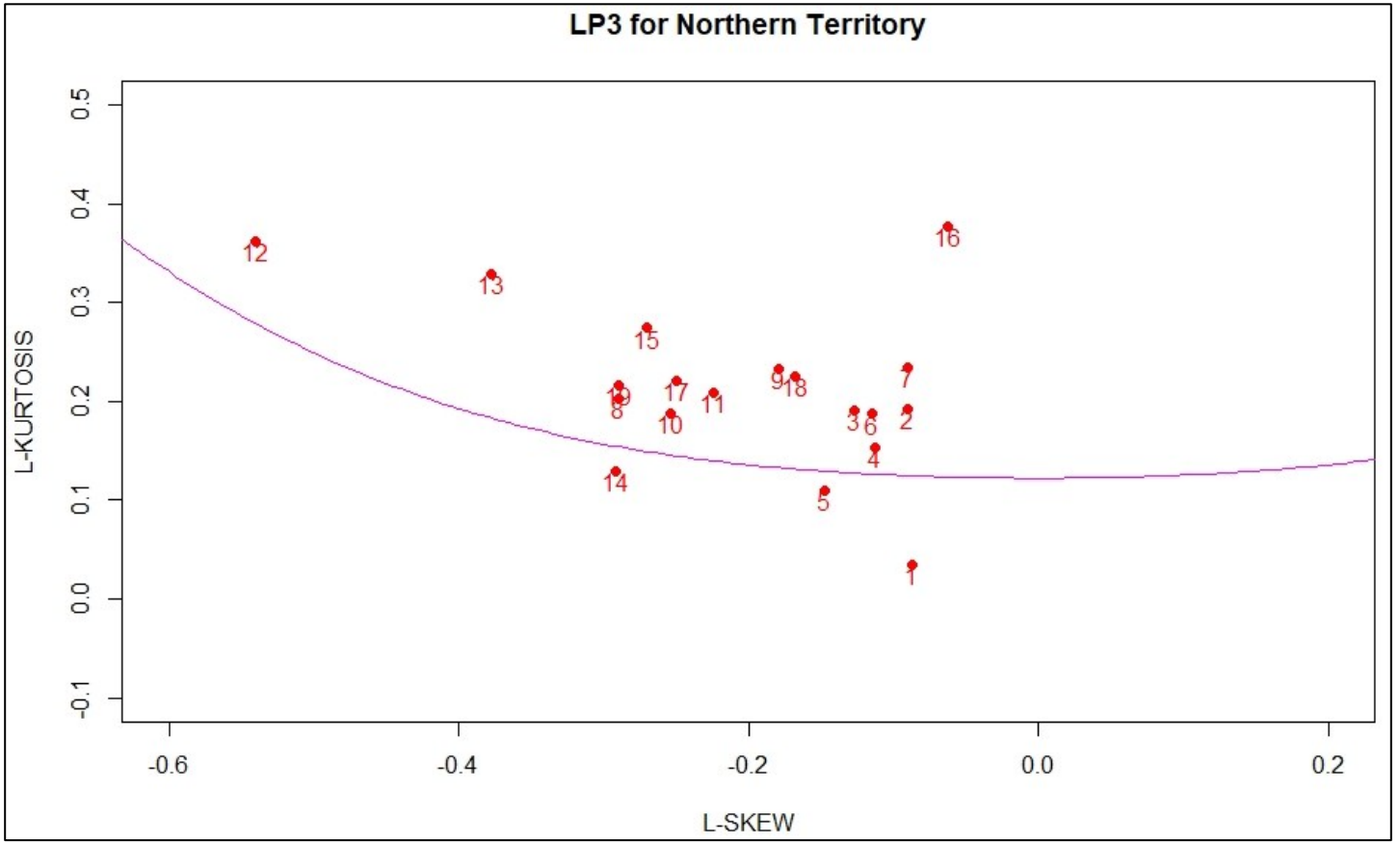

fig. 6.9 – L-Moments Ratio Diagram for Log-Pearson Type 3 for selected sites in Northern Territory

First, it should be noted that the limited number of sites is a limit for drawn definitive conclusions about New South Wales, while for Queensland and Northern Territory there is a clearer context.

It is interesting to note how TCEV lines can be a helpful tool for discriminating between candidate distributions. Looking at fig. 6.4 and 6.6 for New South Wales and Queensland, can be recognized the presence of a cloud of points in the area between the TCEV curves for $\Lambda^* = 1$ and $\theta^* = 20$ and Generalized Pareto line, and that this area is crossed by the P3 distribution line. Queensland plot, furthermore, reports a wider cloud below the $\Lambda^* = 1$ line characterized by values of τ_4 between 0.0 and 0.1. Northern Territory, instead, denotes a lack of points in the above-mentioned area in the TCEV space (except for one site).

Observed dispersion in sample points in fig. 6.6 and 6.8 suggests that differences can be observed when performing FFA for those sites. Heterogeneity of region is confirmed by the evaluation of the Heterogeneity measure (Hosking and Wallis, 1993), that assumes the values of 23.97 for Queensland and of 5.95 for Northern Territory. Both measures were evaluated using the *R* package *lmomRFA* (Hosking, 2019) with 1000 simulations.

LMRDs showing only LP3 distribution, instead, seems to confirm the note ductility of this distribution in giving often a good fit to eastern Australia flood data for at-site analyses. Is should be remarked that LP3 distribution was recommended for at-site analyses by Australian Rainfall and Runoff (ARR) 1987 (I. E. Aust., 1987).

This preliminary analysis confirms TCEV as a good candidate distribution for at-site FFA in Queensland and New South Wales, while Kappa distribution (albeit not plotted in previous diagrams) by an initial analysis is suspected to be not adequate for New South Wales and of interest in some cases in Queensland and Northern Territory.

6.3.1.2 - TC_{ratio} analysis

TC_{ratio} is a measure for discerning the presence of components of different nature in a sample. It is a measure based on the sampled posterior distribution of TCEV in a Bayesian framework. In paragraph 5.4 was illustrated in detail. Fig. 6.10 contains the binary outcome of this investigation for selected sites.

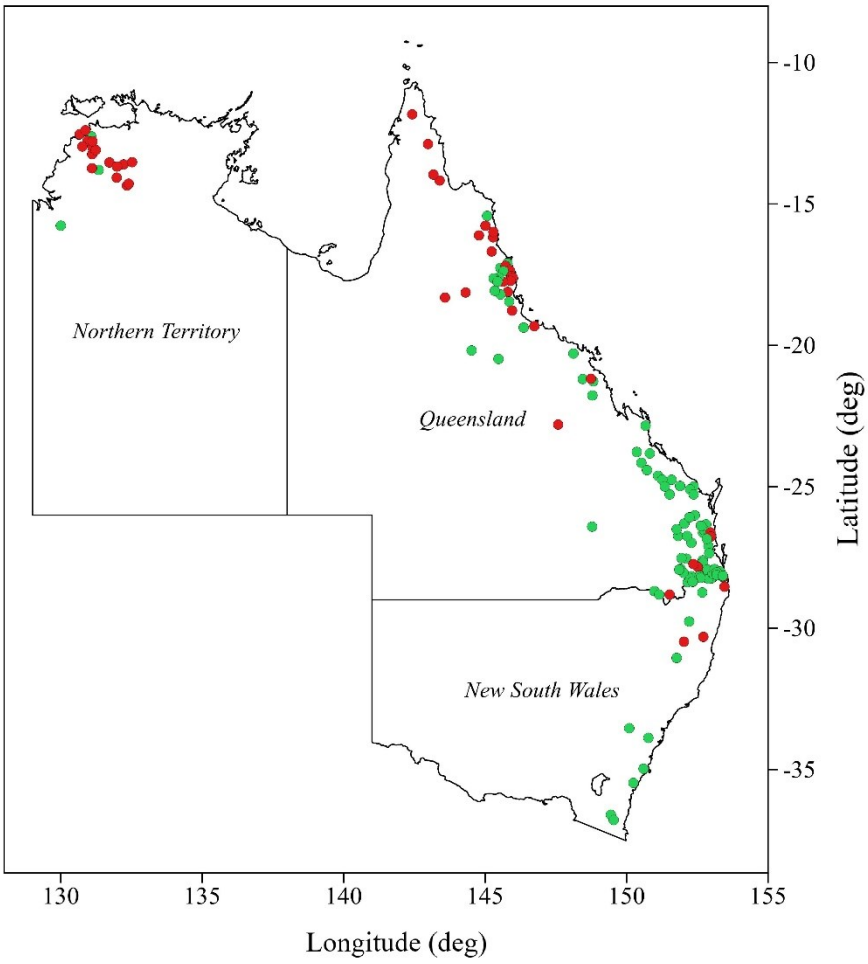

fig. 6.10 – Outcome of TC_{ratio} test: green points indicate the presence of two-components, red the absence

This figure shows surprisingly a spatial correlation of the outcome of TC_{ratio} test. Presence of two different populations seems to be a common characteristic in the Eastern Coast of Australia below the latitude of 23° south. This result seems to be in the spite of those found by Micevski et al. (2006, fig. 8), which reported a similar behavior of their *flood ratio* measure.

Conversely, in Northern Territory only three sites show a positive outcome for the presence of a double component.

Some good-behaved TC_{ratio} diagrams are illustrated in figs. 6.11-6.13.

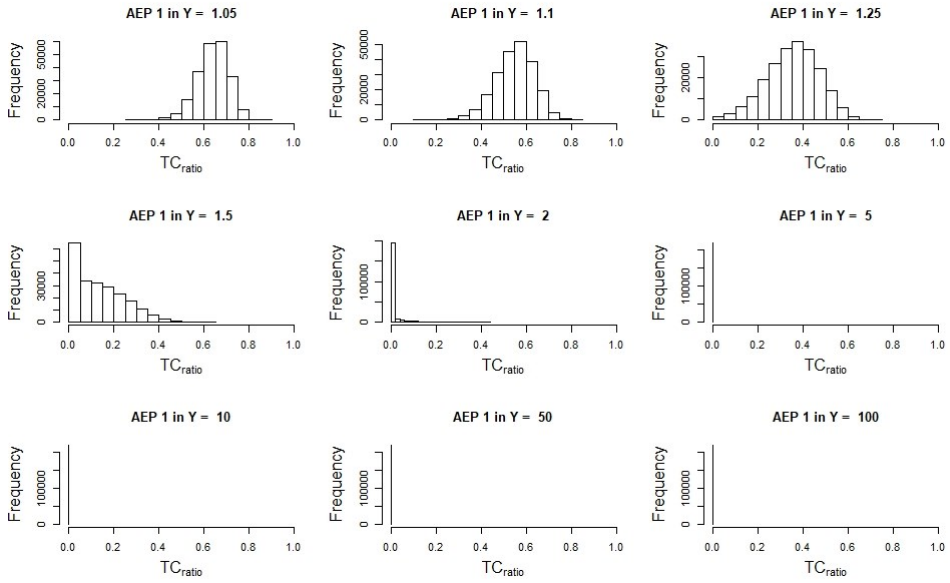

fig. 6.11 - TC_{ratio} diagram for Tantawangalo Creek at Tantawangalo mountain (dam) (NSW)

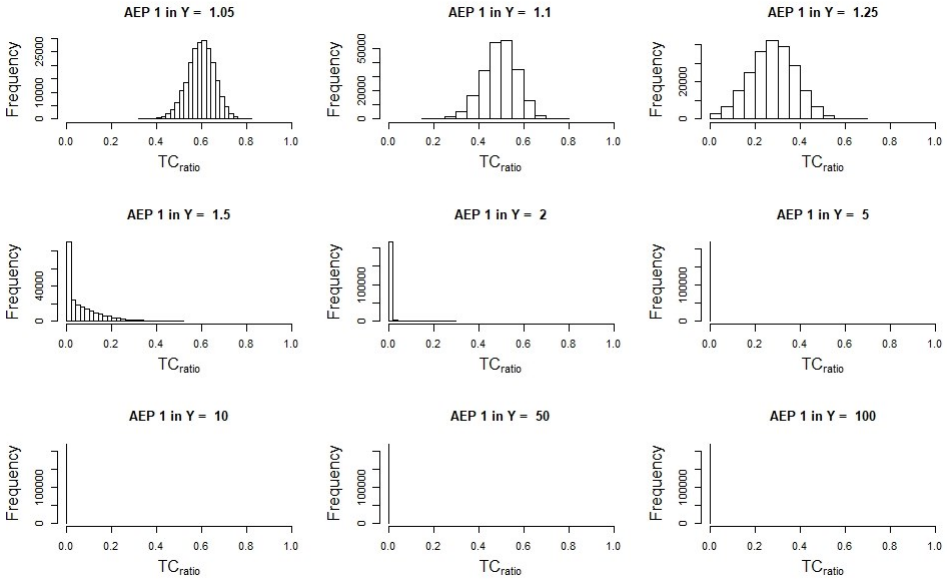

fig. 6.12 - TC_{ratio} diagram for Amamoor Creek at Zachariah (Qld)

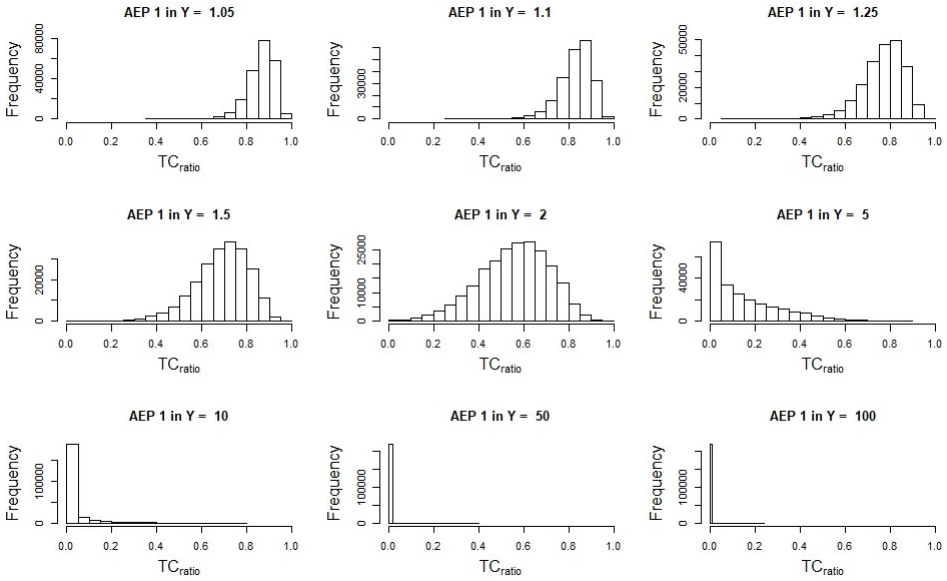

fig. 6.13 - TC_{ratio} diagram for Elizabeth River - Stuart Hwy (NT)

Instead, in figs. 6.14-15 TC_{ratio} diagrams that exclude the presence of a second components are plotted.

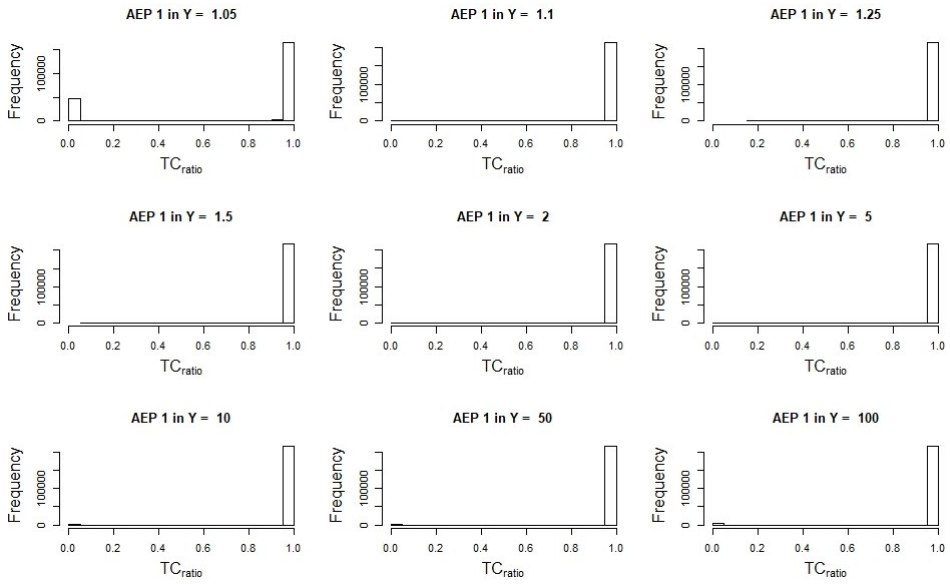

fig. 6.14 - TC_{ratio} diagram for Ross River at Gleesons Weir (Qld)

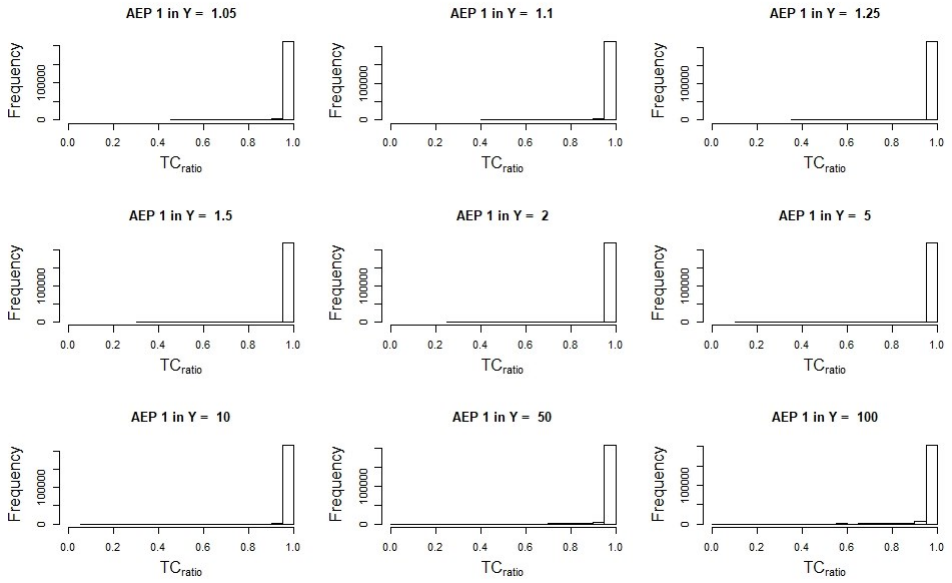

fig. 6.15 - TC_{ratio} diagram for South Alligator River - El Sherana (NT)

6.3.1.3 – Visual Inspection

A visual inspection of probability plot is always necessary for understanding the goodness of fit of a probability distribution to observed data. Application of traditional measures of goodness of fit or model selection without any screening of probability plots can lead to misleading results. However, combining different tests for detecting the best performant between several candidate distributions is a good strategy, because of the increasing in information content available to the analyst.

In sub-paragraph 6.3.1.1 a classical measure for an initial screening of data such as L-moments Ratio Diagram was employed. As results, it showed that TCEV distribution could give a good fit for sites in New South Wales and Queensland, but not in Northern Territory. In 6.3.1.2, a new measure called TC_{ratio} was employed for suspecting the presence of two component of different nature in a

time series. Spatial distribution of results of this test showed clusters similar to those illustrated in Micevski et al. (2006).

Next step in that sort of hierarchical procedure for detection of the best fitting distribution, is to investigate the behavior of probability plots for three different distributions: TCEV, LP3 and GEV. Kappa distribution was omitted also because of the frequent issues arose in the computational step. In most cases, it was not possible start in sampling the posterior distribution. In all computations, Bayesian analysis with no prior information (Kuczera, 1999) was performed.

Several probability plots are reported in the following figures. Their choice was led by the need of providing the main behaviors verified in the study area.

Furthermore, for illustrating the uncertainty in the estimate of TCEV parameters, their parameters probability plots from the posterior distribution are also illustrated.

New South Wales: Apsley River at Apsley Falls

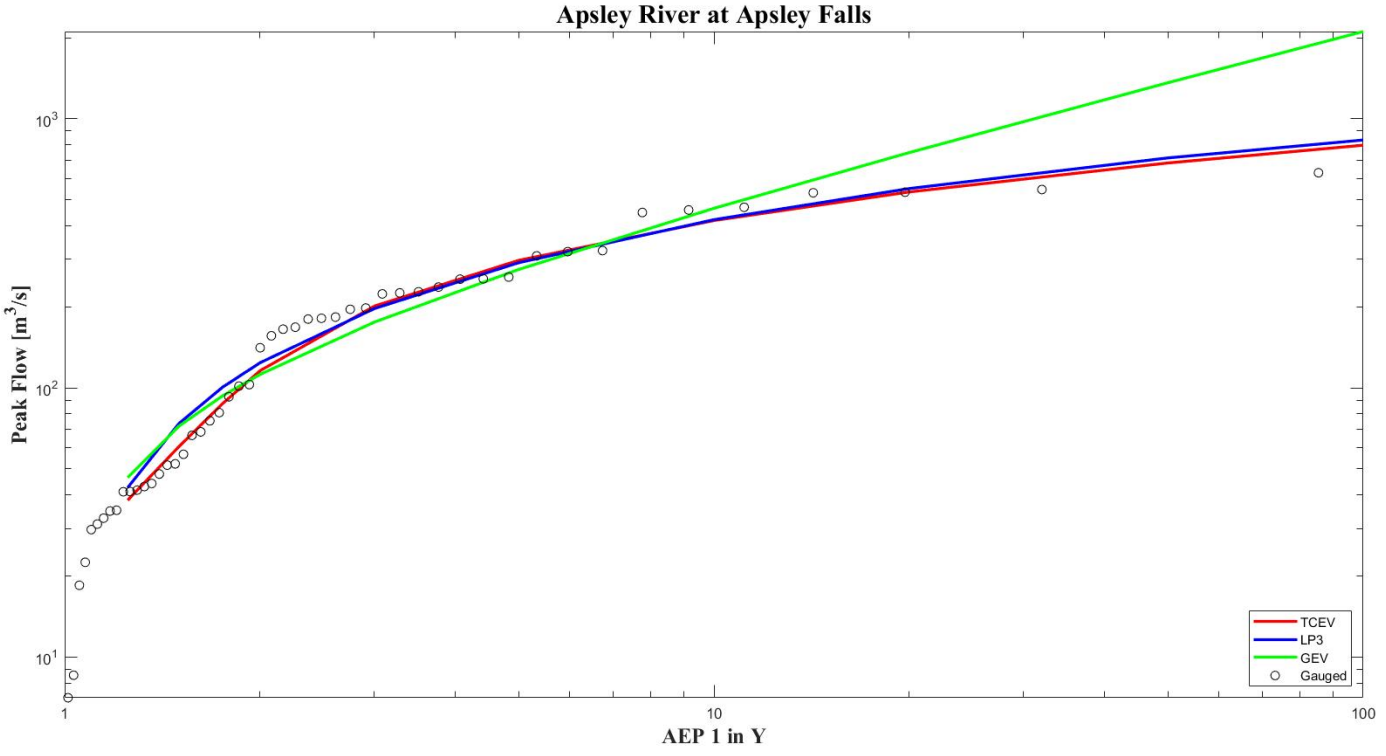

figure 6.16 – TCEV, LP3 and GEV expected quantiles probability plots for Apsley River at Apsley Falls

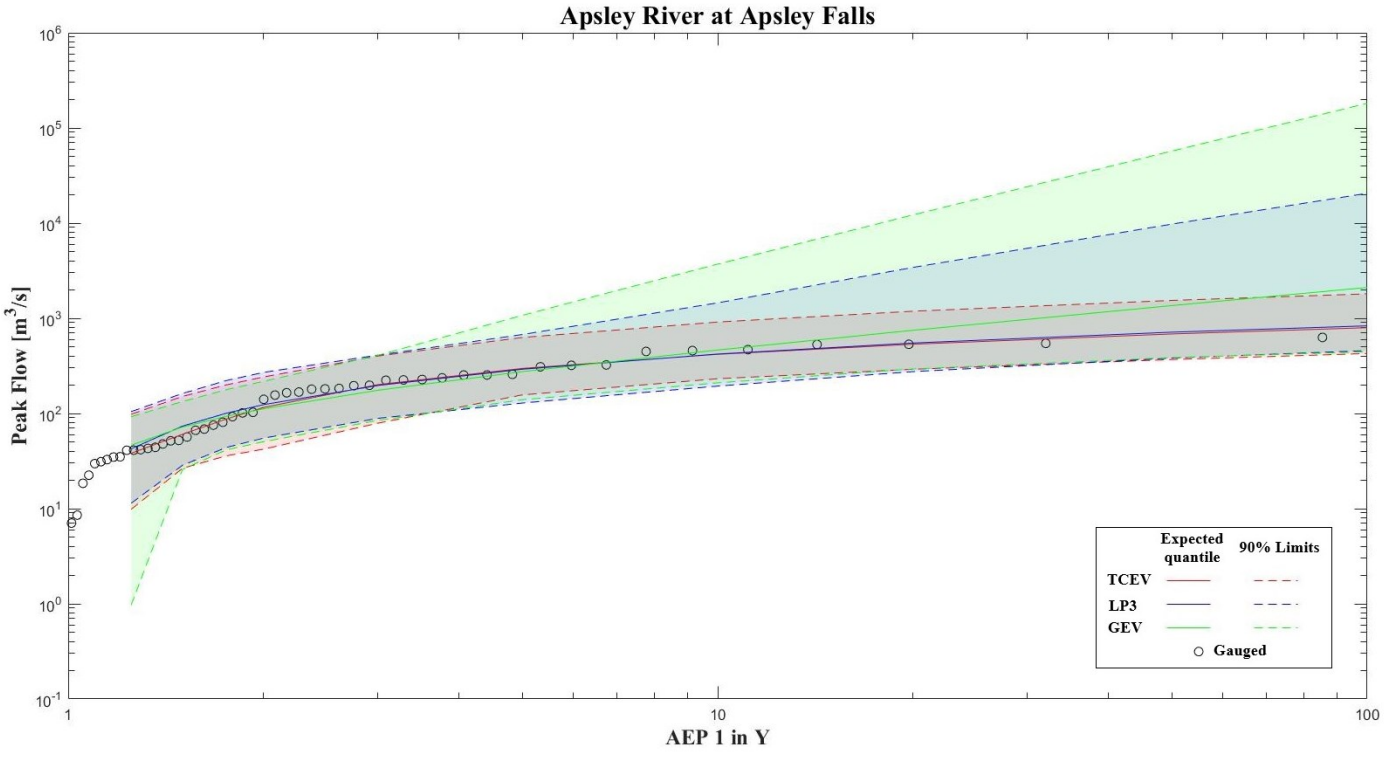

figure 6.17 – TCEV, LP3 and GEV probability plots for Apsley River at Apsley Falls

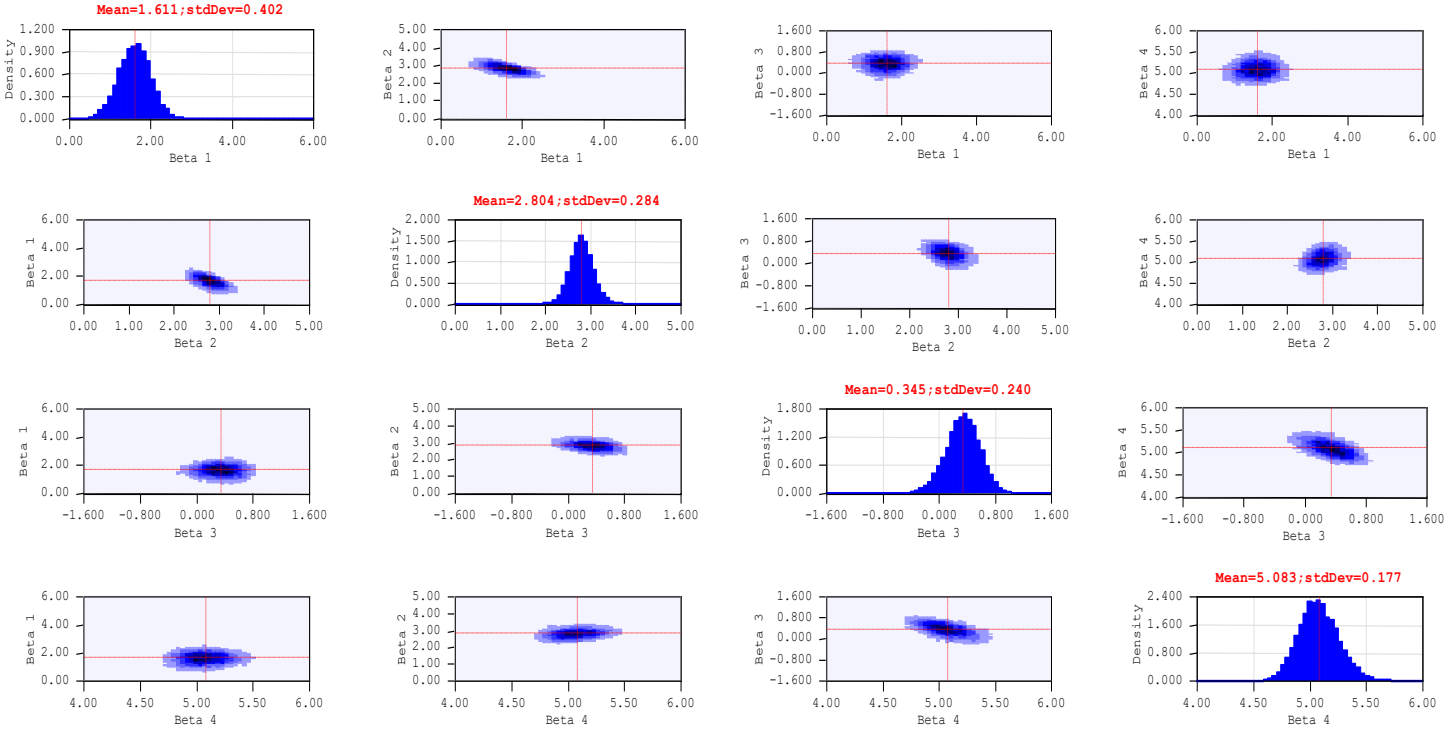

figure 6.18 – TCEV parameters probability plot for Apsley River at Apsley Falls

1 in Y AEP	Distribution	Expected Parameter Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	115.81	42.21	242.2
	LP3	123.96	55.2	270.4
	GEV	112.18	50.57	217.9
10	TCEV	418.44	231.92	914.1
	LP3	421.49	194.09	1454
	GEV	464.58	209.74	3714.2
50	TCEV	684.68	368.74	1542
	LP3	714.89	381.98	9762.4
	GEV	1359.99	386.17	57139
100	TCEV	797.24	426.58	1807.5
	LP3	833.38	457.97	20580.6
	GEV	2108.11	445.44	181500

Table 6.4 - Comparison of Selected Quantiles with 90% Confidence Limits for Apsley River at Apsley Falls

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.611	2.804	0.345	5.083
Std Dev	0.402	0.284	0.240	0.177

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	4.670	0.161	-0.776
Std Dev	0.168	0.127	0.428

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	79.383	4.380	-0.609
Std Dev	14.143	0.184	0.202

Tables 6.5 – Compsrison of estimates parameters for Apsley River at Apsley Falls

New South Wales: Rutherford Creek at Brown Mountain

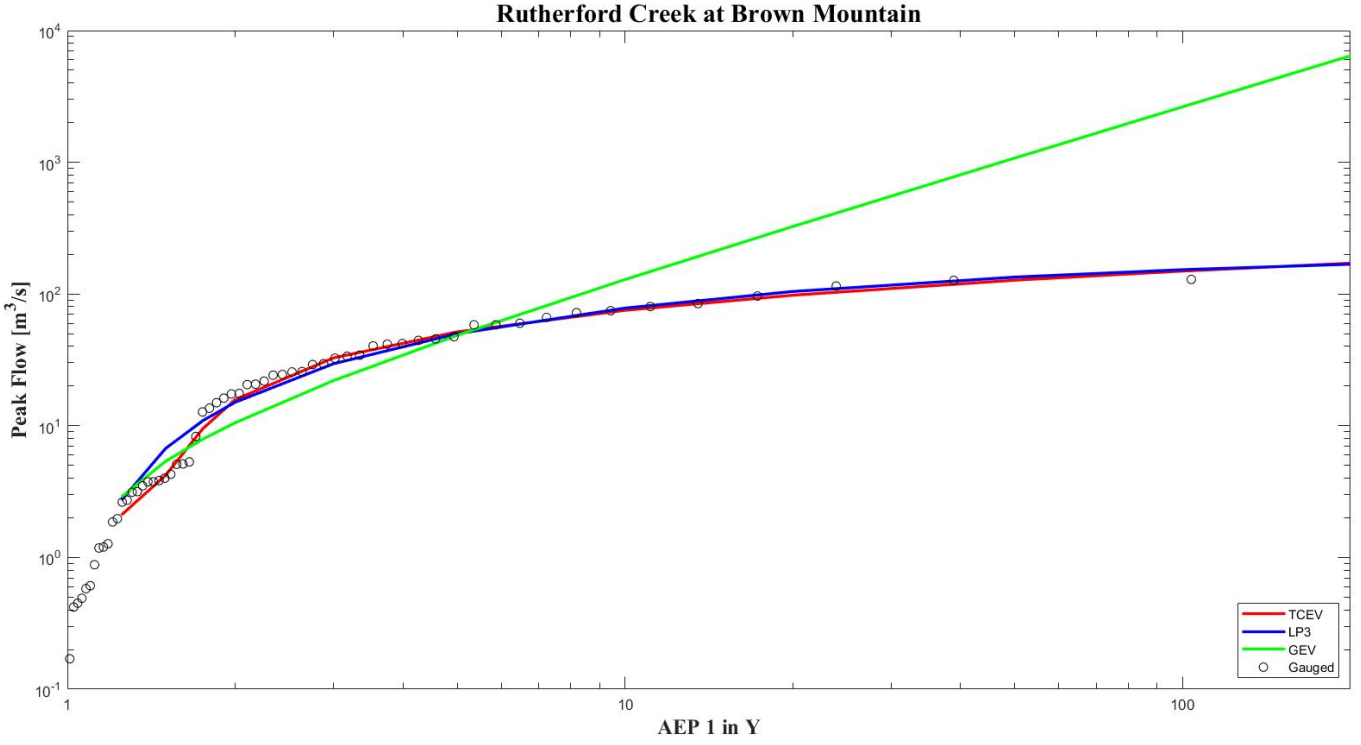

figure 6.19 – TCEV, LP3 and GEV expected quantiles probability plots for Rutherford Creek at Brown Mountain

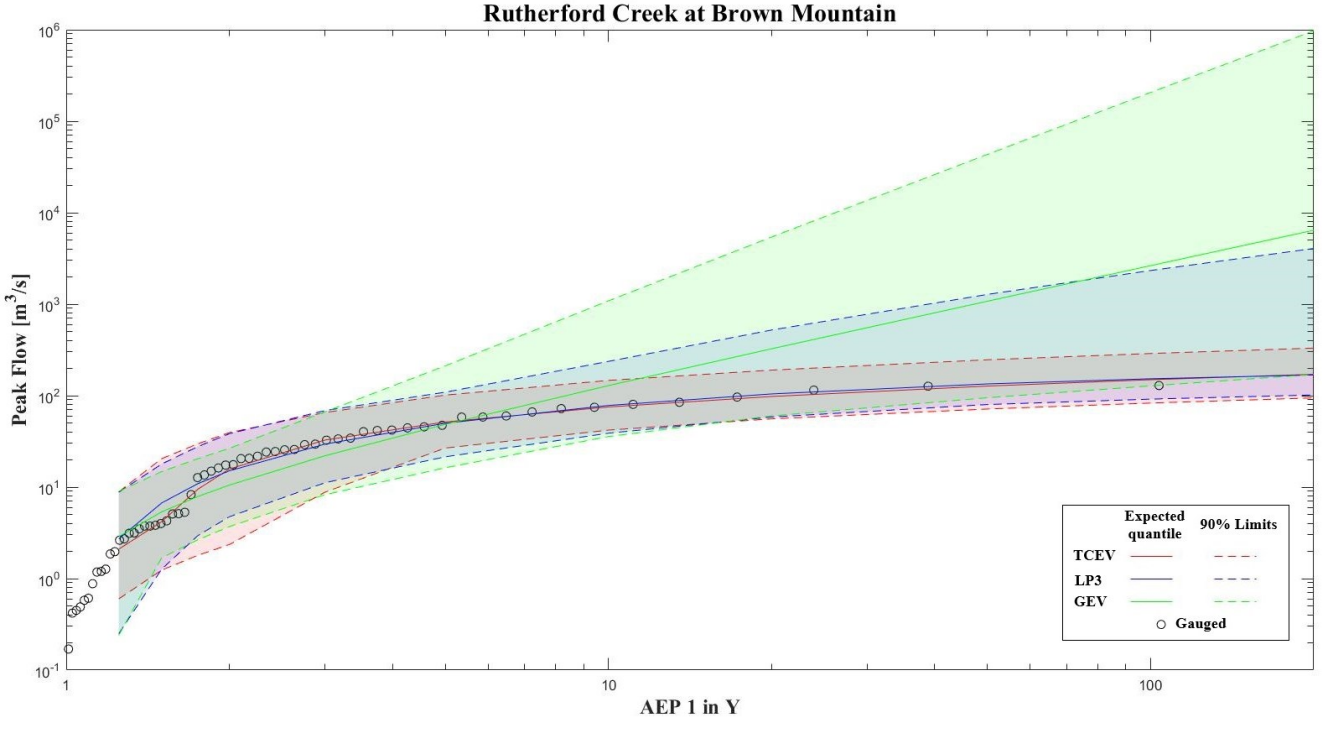

figure 6.20 – TCEV, LP3 and GEV probability plots for Rutherford Creek at Brown Mountain

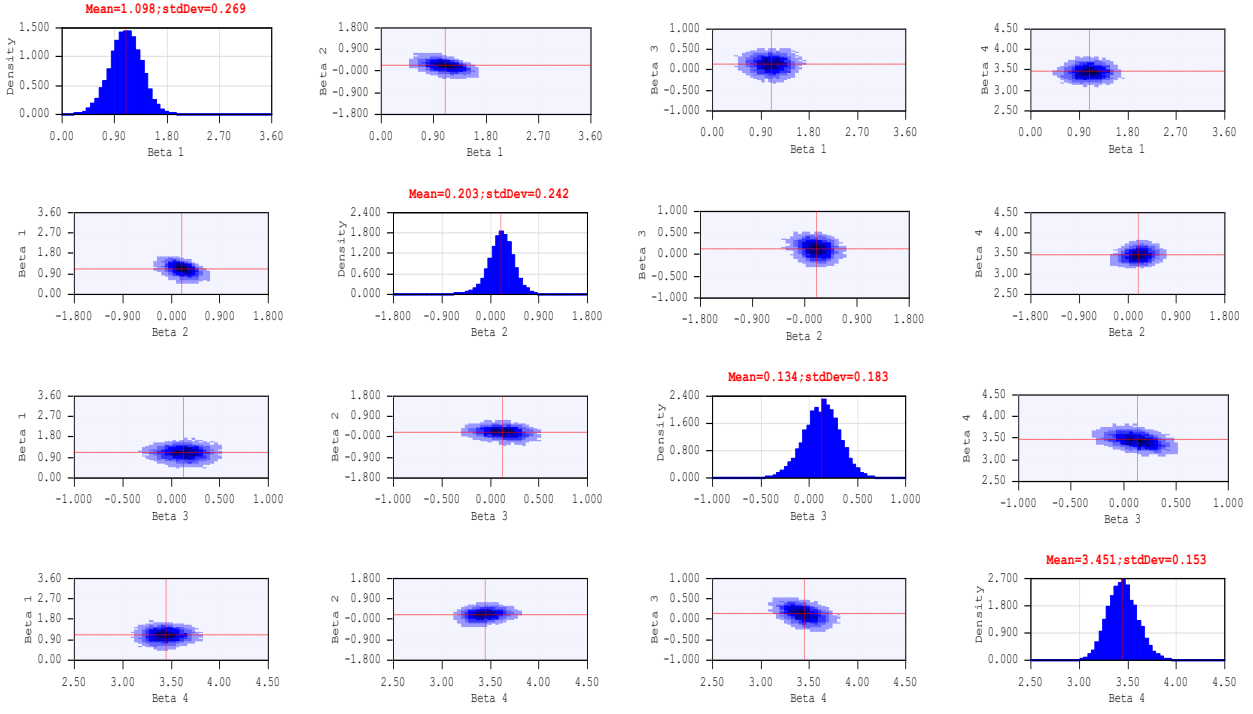

figure 6.21 – TCEV parameters probability plot for Rutherford Creek at Brown Mountain

1 in Y AEP	Distribution	Expected Parameter Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	115.81	42.21	242.2
	LP3	123.96	55.2	270.4
	GEV	112.18	50.57	217.9
10	TCEV	418.44	231.92	914.1
	LP3	421.49	194.09	1454
	GEV	464.58	209.74	3714.2
50	TCEV	684.68	368.74	1542
	LP3	714.89	381.98	9762.4
	GEV	1359.99	386.17	57139
100	TCEV	797.24	426.58	1807.5
	LP3	833.38	457.97	20580.6
	GEV	2108.11	445.44	181500

Table 6.6 - Comparison of Selected Quantiles with 90% Confidence Limits for Rutherford Creek at Brown Mountain

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.611	2.804	0.345	5.083
Std Dev	0.402	0.284	0.240	0.177

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	4.670	0.161	-0.776
Std Dev	0.168	0.127	0.428

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	79.383	4.380	-0.609
Std Dev	14.143	0.184	0.202

Tables 6.7 – Comparison of estimates parameters for Rutherford Creek at Brown Mountain

New South Wales: Tantawangalo Creek at Tantawangalo Mountain (dam)

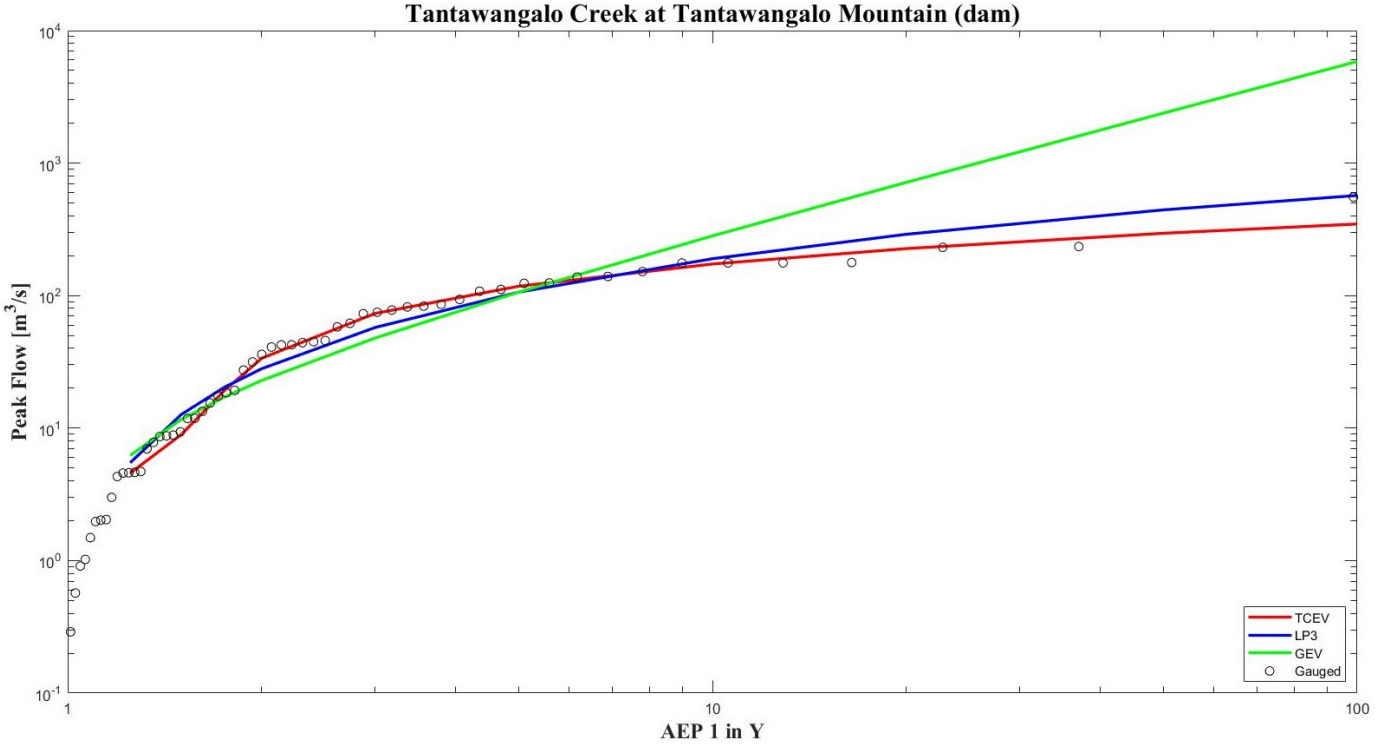

figure 6.22– TCEV, LP3 and GEV expected quantiles probability plots for Tantawangalo Creek at Tantawangalo Mountain (dam)

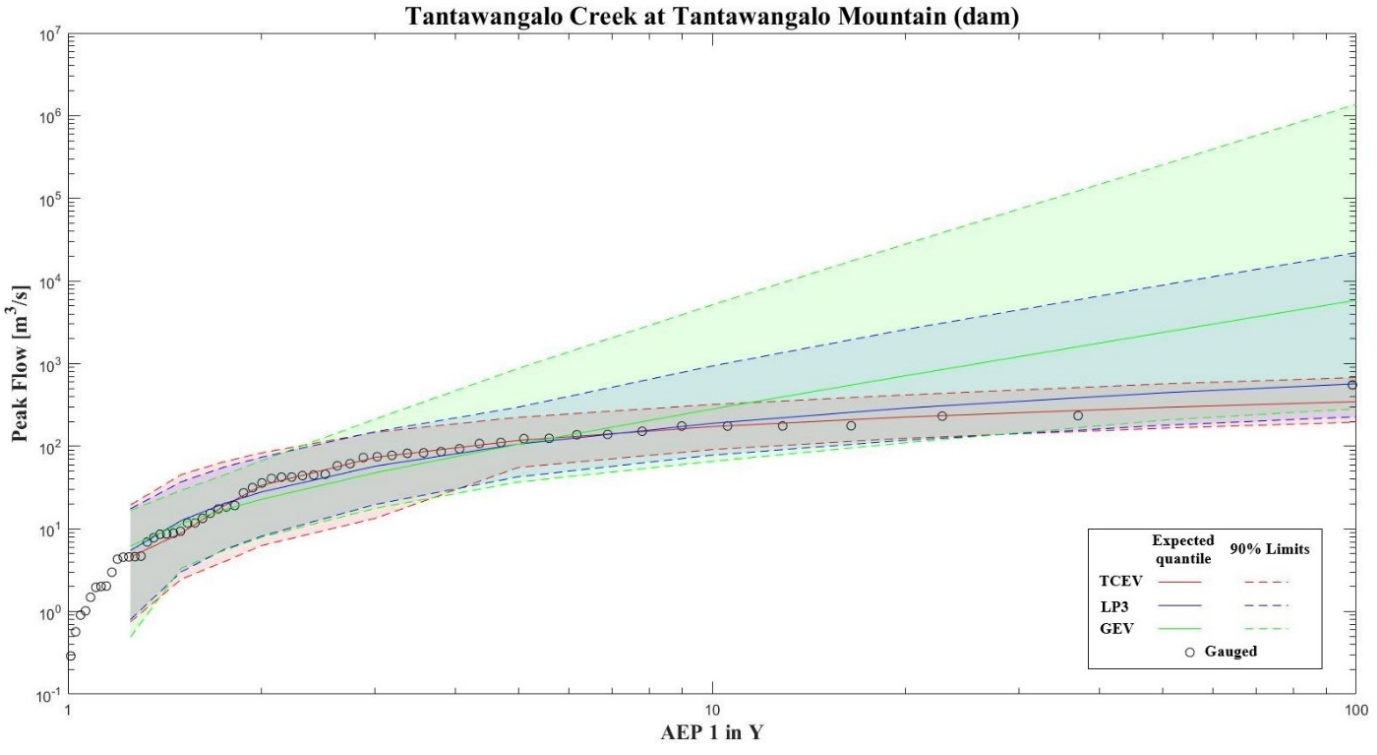

figure 6.23 – TCEV, LP3 and GEV probability plots for Tantawangalo Creek at Tantawangalo Mountain (dam)

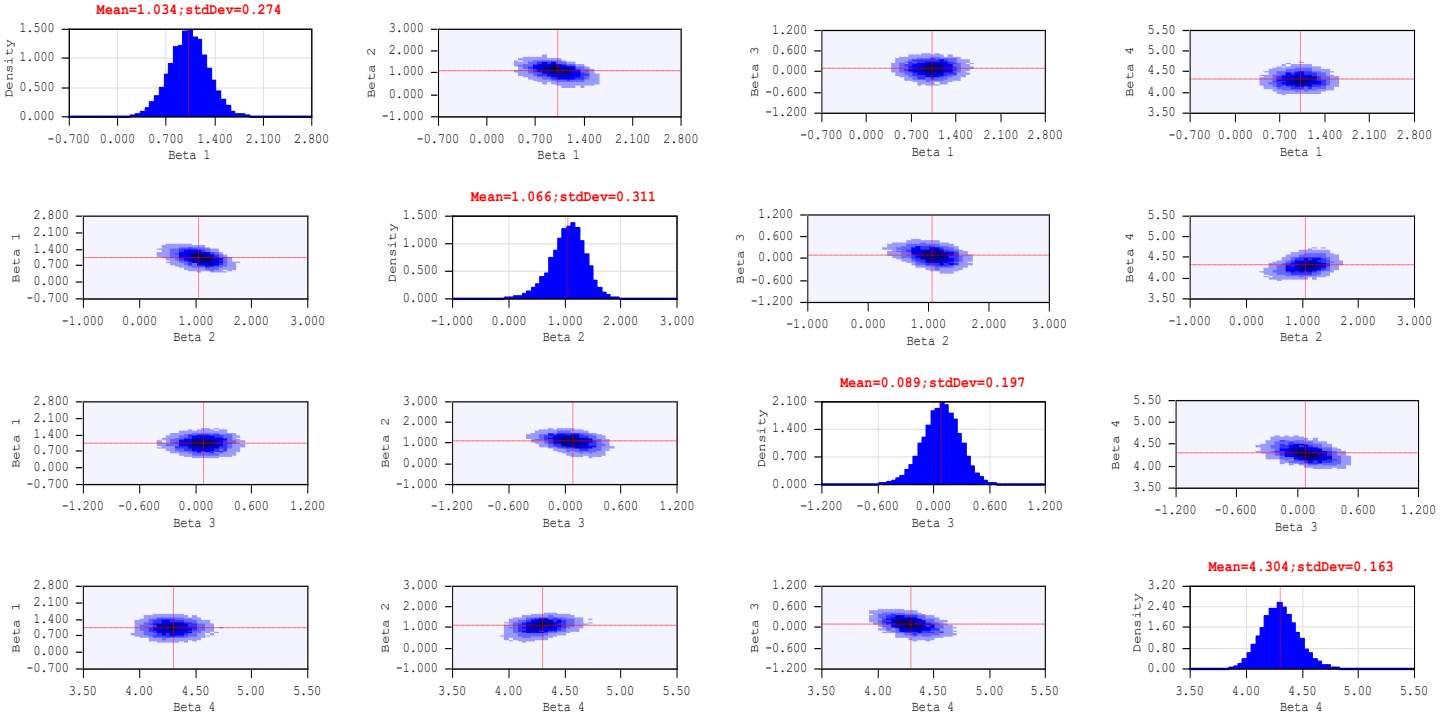

figure 6.24 – TCEV parameters probability plot for Tantawangalo Creek at Tantawangalo Mountain (dam)

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	33.71	6.31	83.4
	LP3	28.09	8.21	73.8
	GEV	22.89	7.95	67.4
10	TCEV	173.09	91.16	320.1
	LP3	189.74	78.07	937.8
	GEV	282.15	65.63	5138.1
50	TCEV	295.28	165.59	567.8
	LP3	443.39	179.47	8902.4
	GEV	2379.63	206.23	253310.7
100	TCEV	346.93	195.74	679.4
	LP3	569.93	227.48	22021
	GEV	5846.92	284.52	1373512

Table 6.8 - Comparison of Selected Quantiles with 90% Confidence Limits for Tantawangalo Creek at Tantawangalo Mountain (dam)

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.03366	1.06605	0.08914	4.30386
Std Dev	0.27446	0.31098	0.19683	0.16346

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	3.12041	0.58862	-0.72165
Std Dev	0.23891	0.10915	0.29827

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	13.416	3.007	-1.287
Std Dev	3.516	0.253	0.261

Tables 6.9 – Comparison of estimates parameters for Tantawangalo Creek

Queensland: Bell Creek at Craiglands

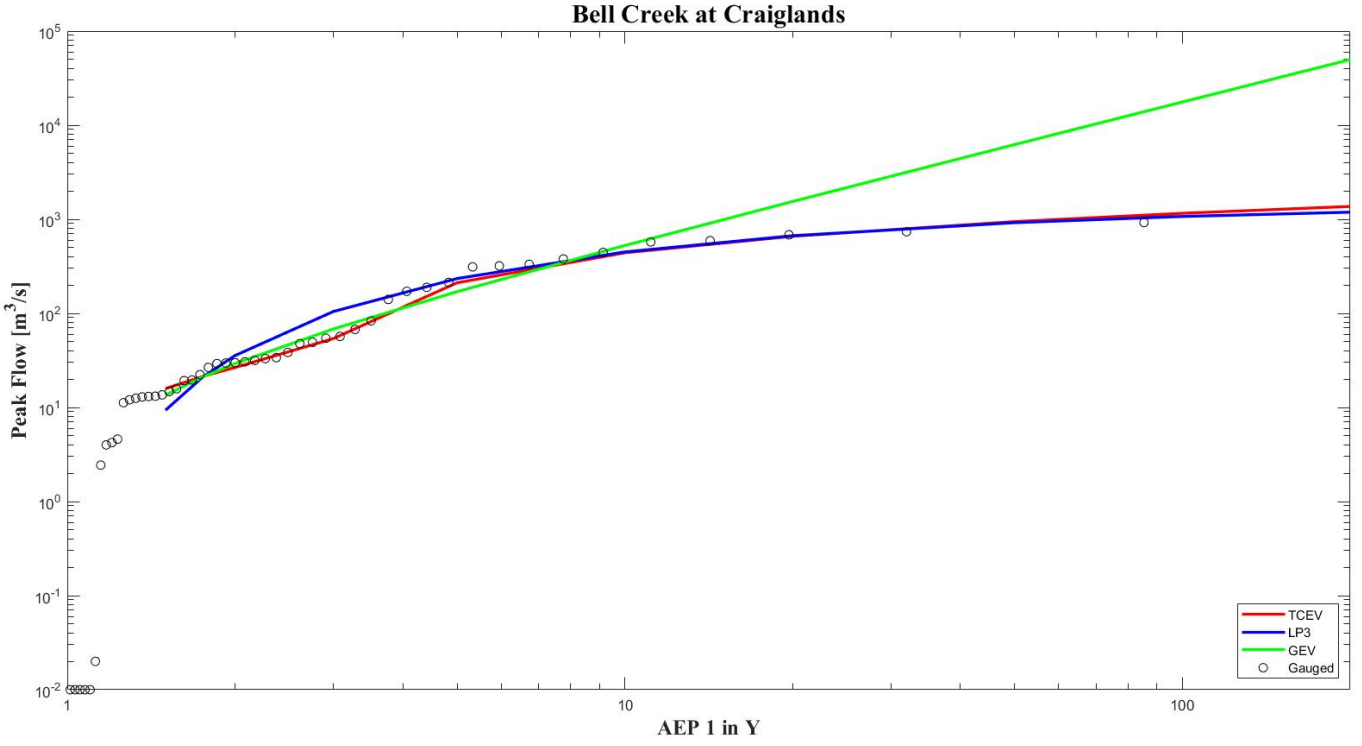

figure 6.25– TCEV, LP3 and GEV expected quantiles probability plots for Bell Creek at Craiglands

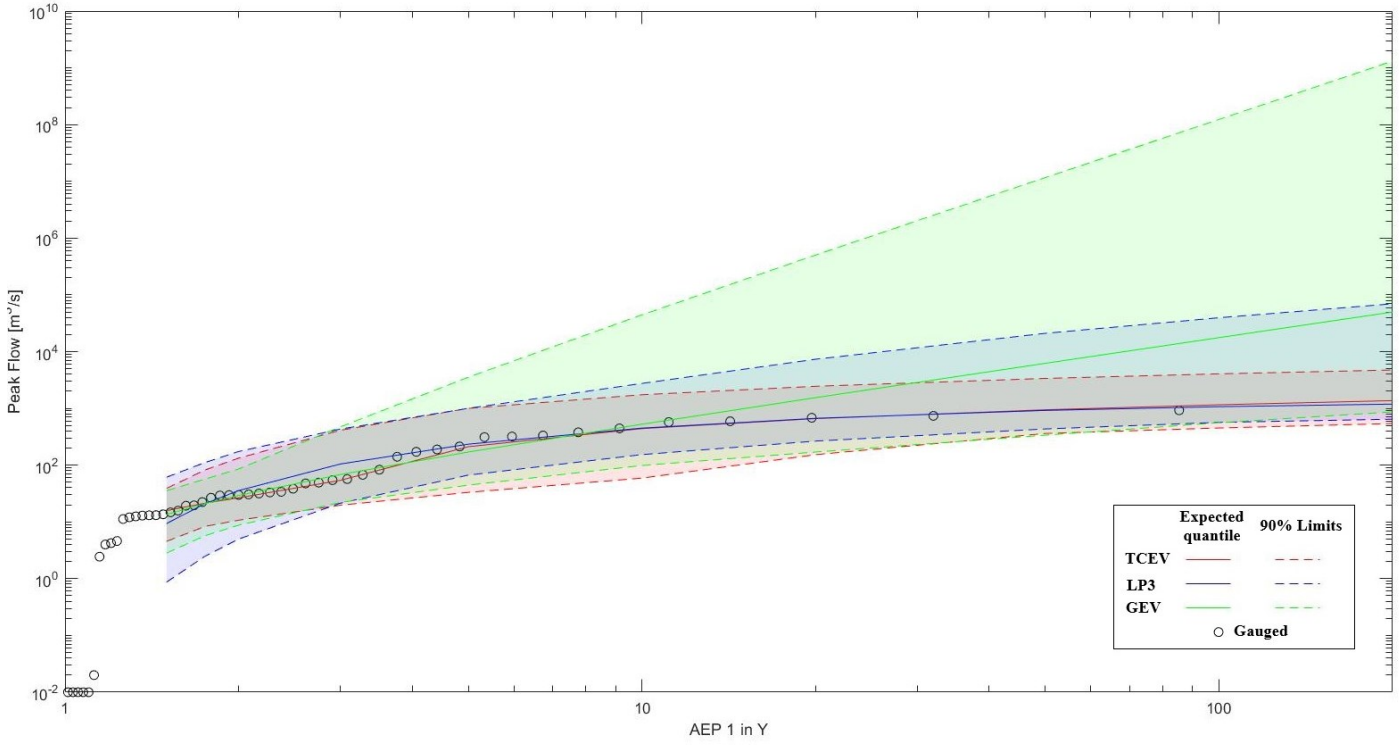

figure 6.26 – TCEV, LP3 and GEV probability plots for Bell Creek at Craiglands

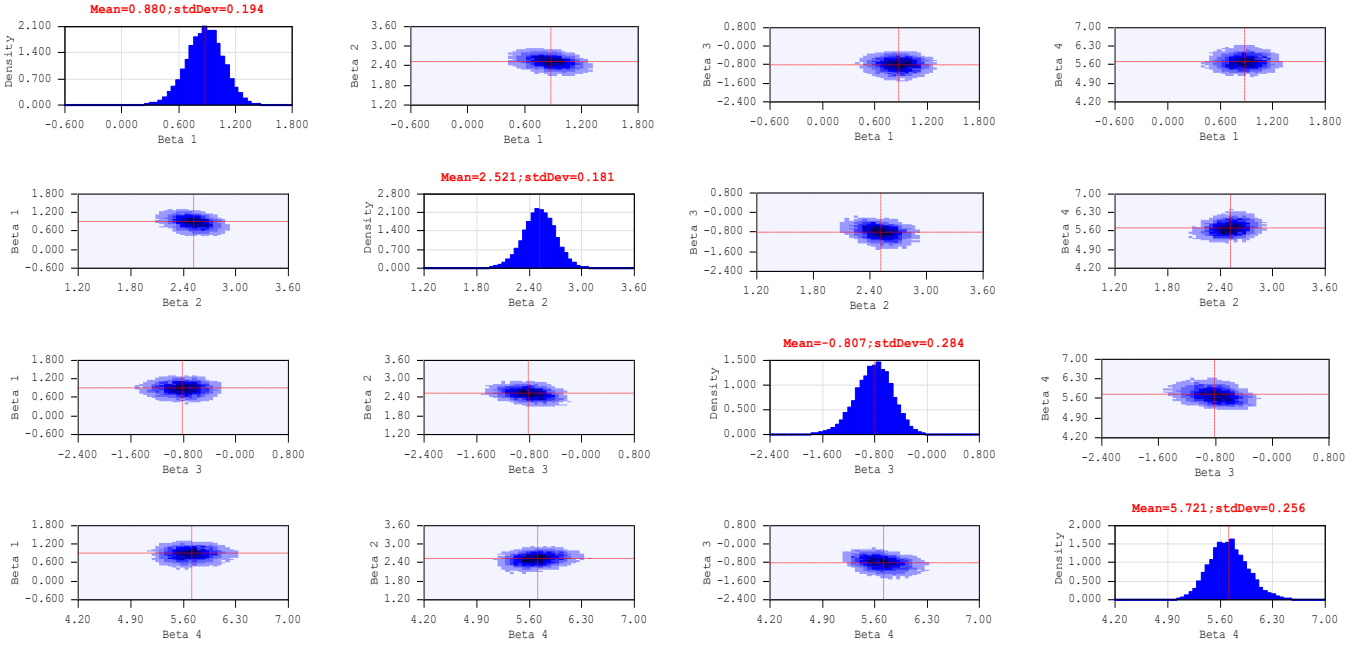

figure 6.27 – TCEV posterior parameters plots for Bell Creek at Craiglands

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	26.6	10.71	133.1
	LP3	35.66	4.98	173.9
	GEV	29.32	8.77	85.6
10	TCEV	440.42	59.1	1741.5
	LP3	449.76	153.51	2754.6
	GEV	526.26	98.62	44462.5
50	TCEV	944.4	361.57	3371.8
	LP3	920.54	434.43	21008
	GEV	6220	337.47	11662496
100	TCEV	1157.47	451.23	4061
	LP3	1071.66	567.09	39524.4
	GEV	17642.79	558.29	122842976

Table 6.10 - Comparison of Selected Quantiles with 90% Confidence Limits for Bell Creek at Craiglands

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	0.88025	2.5211	-0.80711	5.72082
Std Dev	0.19384	0.18106	0.28366	0.25625

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	2.88243	1.12842	-1.38957
Std Dev	0.44031	0.14375	0.26173

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	15.934	3.312	-1.493
Std Dev	5.022	0.295	0.324

Tables 6.11 – Comparison of estimates parameters for Bell Creek at Craiglands

Queensland: Albert River at Bromfleet

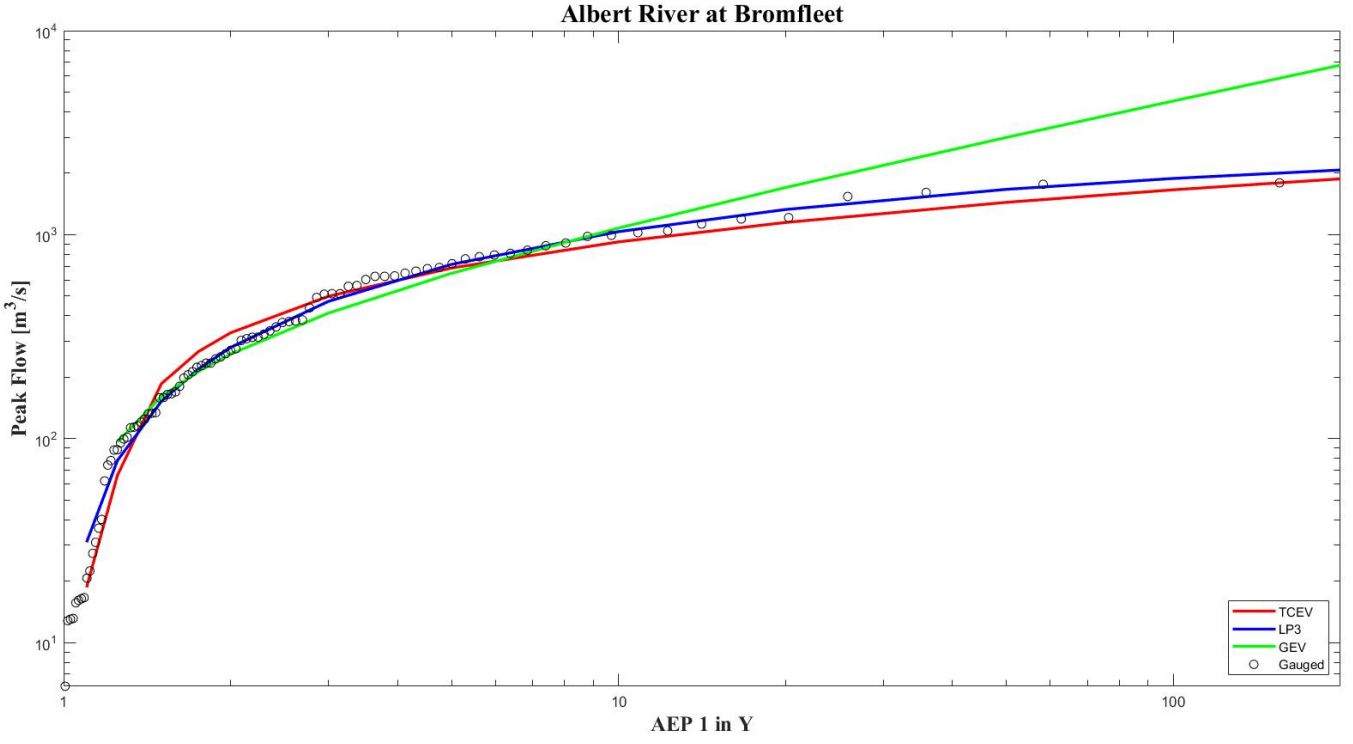

figure 6.28– TCEV, LP3 and GEV expected quantiles probability plots for Albert River at Bromfleet

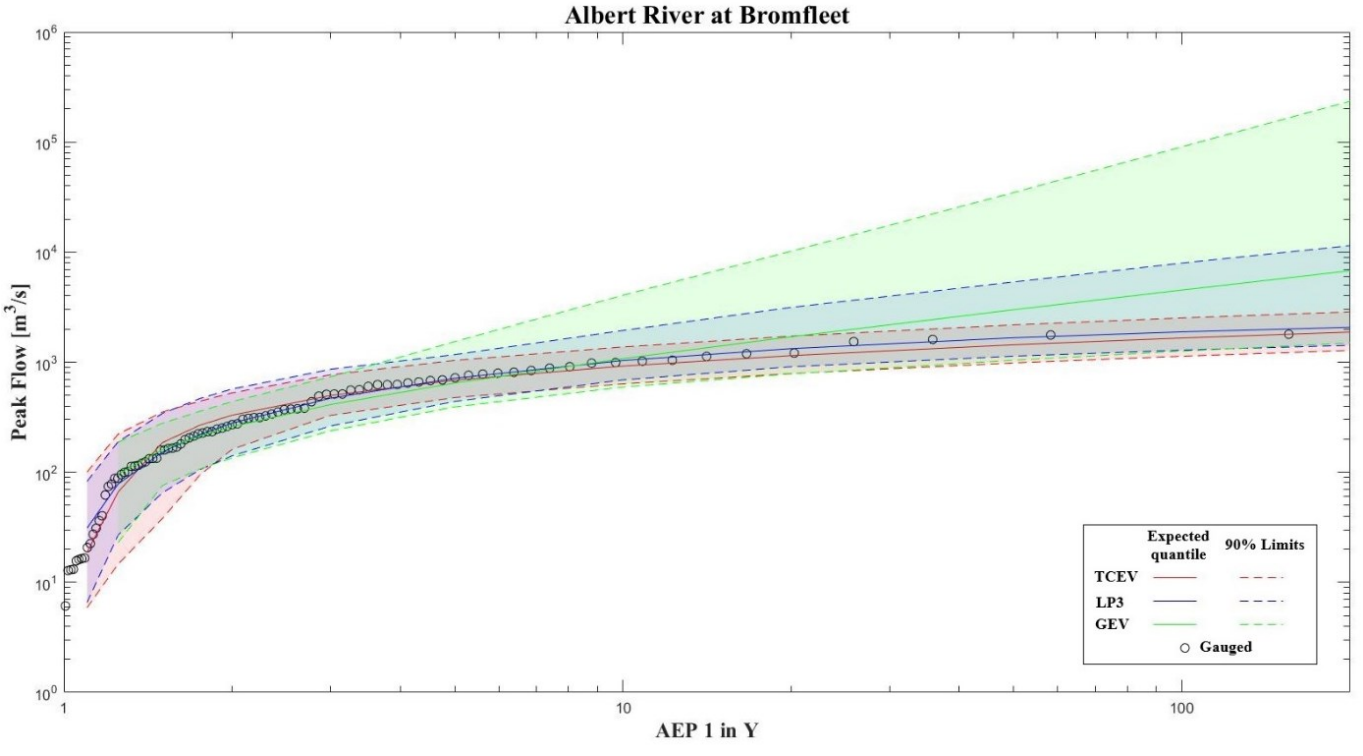

figure 6.29 – TCEV, LP3 and GEV probability plots for Albert River at Bromfleet

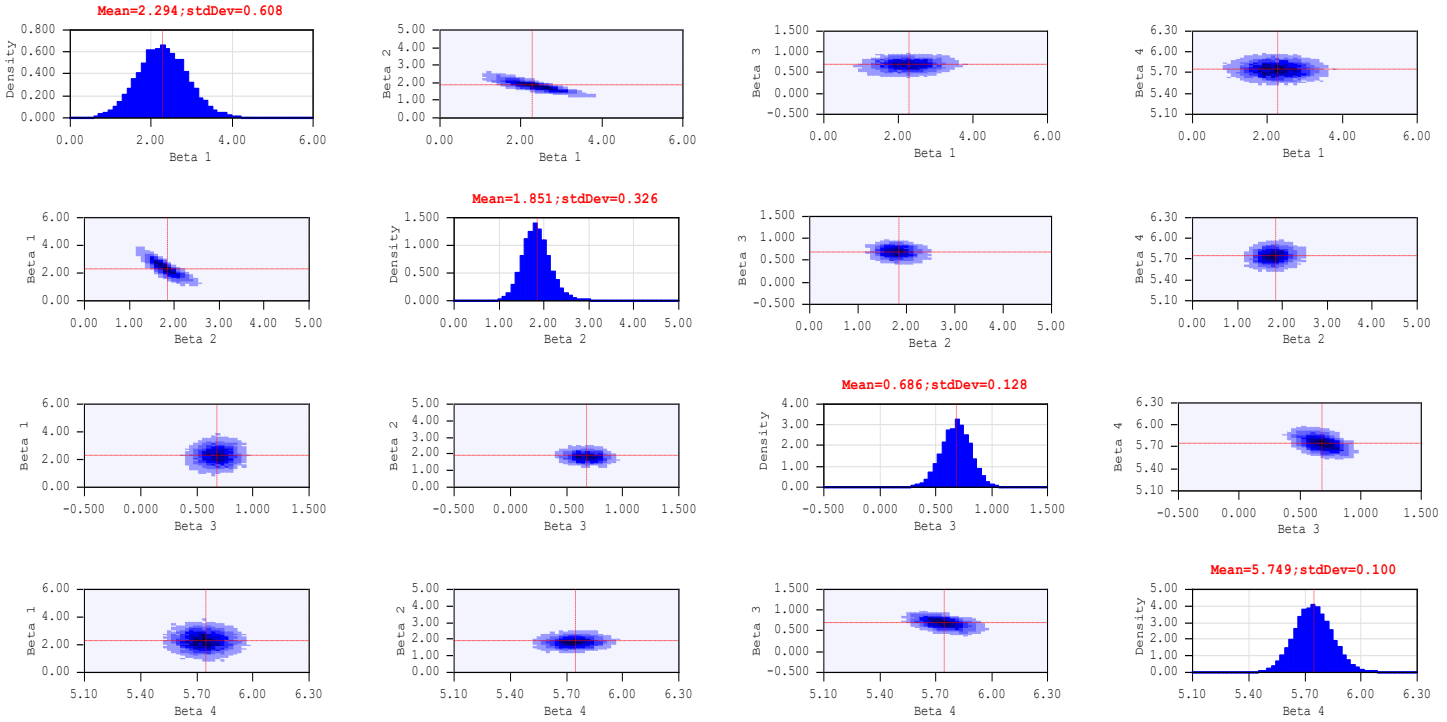

figure 6.30 – TCEV posterior parameters plots for Albert River at Bromfleet

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	330.38	161.69	527.2
	LP3	280.08	141.17	571.4
	GEV	259.73	135.32	438.6
10	TCEV	921.65	635.74	1376.2
	LP3	1034.6	690.58	1941.7
	GEV	1077.17	594.77	4058.4
50	TCEV	1440.01	985.01	2182.5
	LP3	1667.38	1134.72	5358.4
	GEV	2995.95	1041.19	34782.9
100	TCEV	1659.15	1132.67	2524.6
	LP3	1886.23	1287.22	7958.4
	GEV	4522.54	1258.92	90590

Table 6.12 - Comparison of Selected Quantiles with 90% CL for Albert River at Bromfleet

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	2.29427	1.85065	0.68613	5.74895
Std Dev	0.6083	0.32589	0.12776	0.09951

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	5.39685	0.3243	-1.05255
Std Dev	0.14672	0.09799	0.2214

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	179.028	5.290	-0.563
Std Dev	26.272	0.132	0.156

Tables 6.13 – Comparison of estimates parameters for Bell Creek at Albert River at Bromfleet

Queensland: Liverpool Creek at Upper Japoonvale

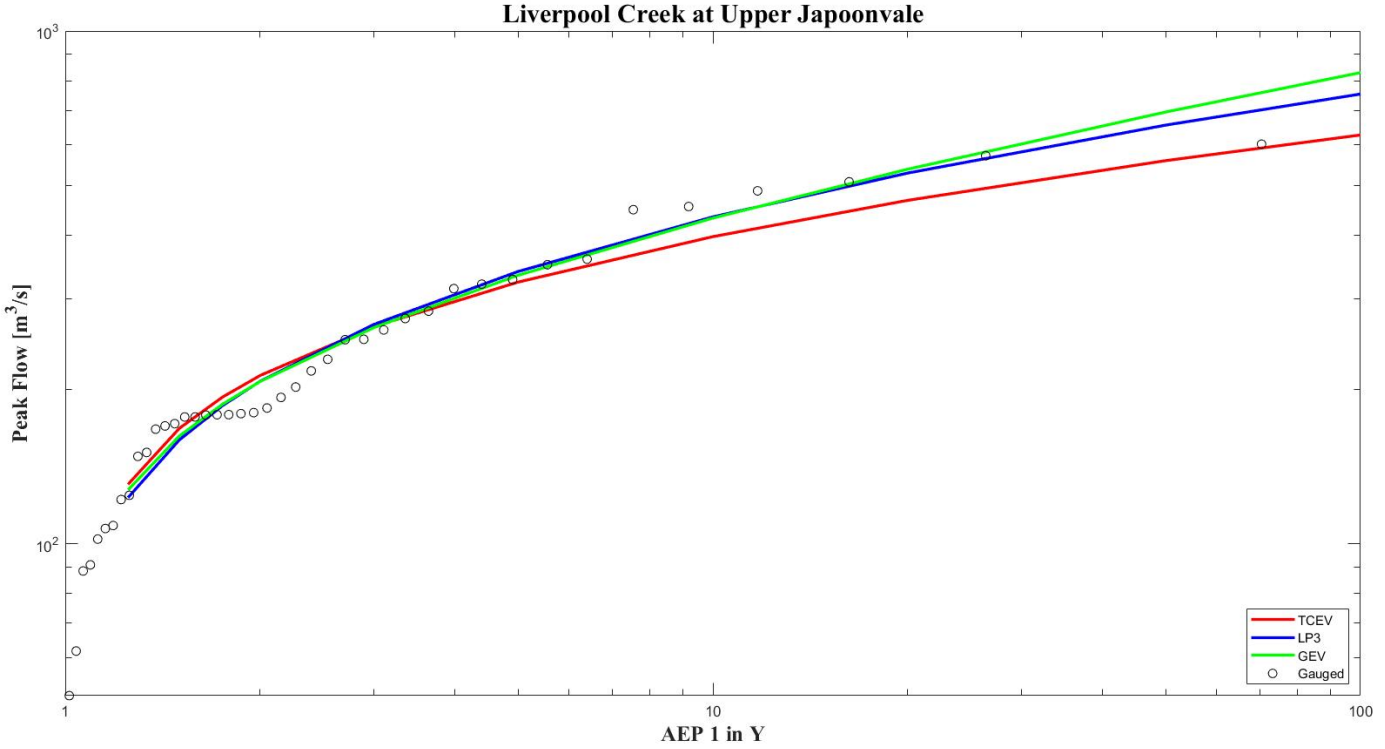

figure 6.31– TCEV, LP3 and GEV expected quantiles probability plots for Liverpool Creek at Upper Japoonvale

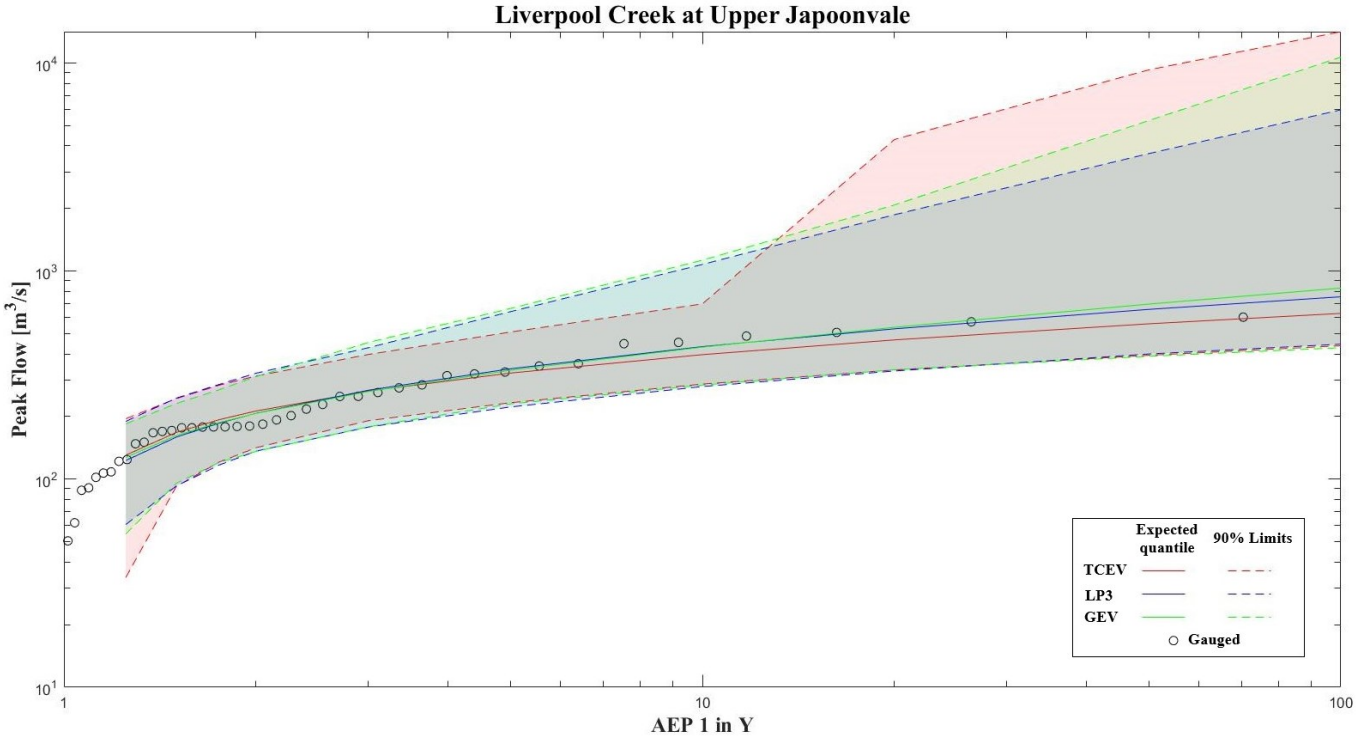

figure 6.32 – TCEV, LP3 and GEV probability plots for Liverpool Creek at Upper Japoonvale

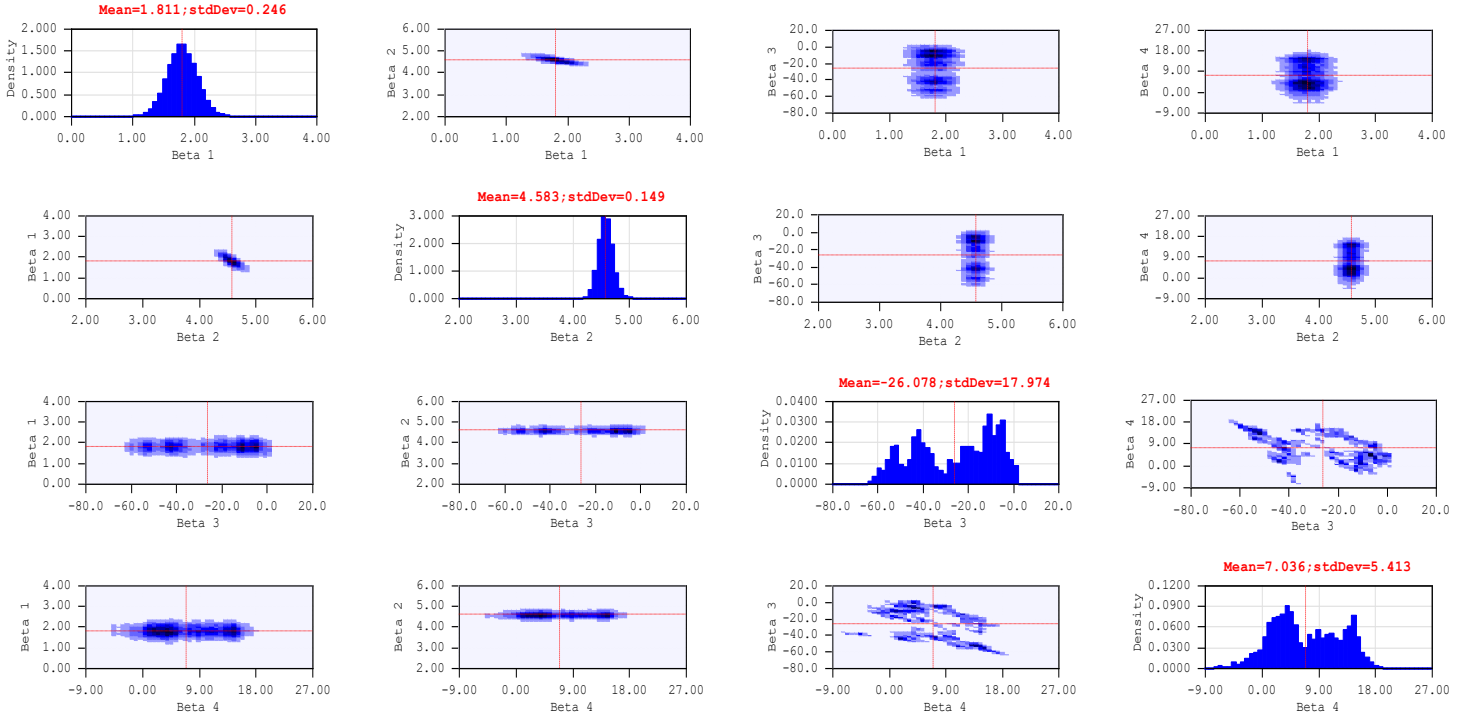

figure 6.33 – TCEV posterior plots for Liverpool Creek at Upper Japoonvale

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	212.91	141.77	313.4
	LP3	207.47	136.45	323.5
	GEV	207.23	136.03	311.6
10	TCEV	397.08	286.71	696.4
	LP3	433.97	278.92	1077.7
	GEV	431.64	283.64	1129.5
50	TCEV	558.54	394.18	9280.7
	LP3	655.15	399.89	3673.7
	GEV	694.78	389.81	5292.4
100	TCEV	626.8	438.16	14167.5
	LP3	753.33	445.91	5963.8
	GEV	829.49	428.69	10683.5

Table 6.14 - Comparison of Selected Quantiles with 90% CL for Liverpool Creek at Upper Japoonvale

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.81129	4.58255	-26.0782	7.03624
Std Dev	0.24569	0.14919	17.97398	5.41288

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	5.31364	-0.5038	-0.21199
Std Dev	0.09498	0.12219	0.39099

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	171.109	4.560	-0.165
Std Dev	17.101	0.145	0.141

Tables 6.15 – Comparison of estimates parameters for Liverpool Creek at Upper Japoonvale

Northern Territory: Elizabeth River - Stuart Hwy

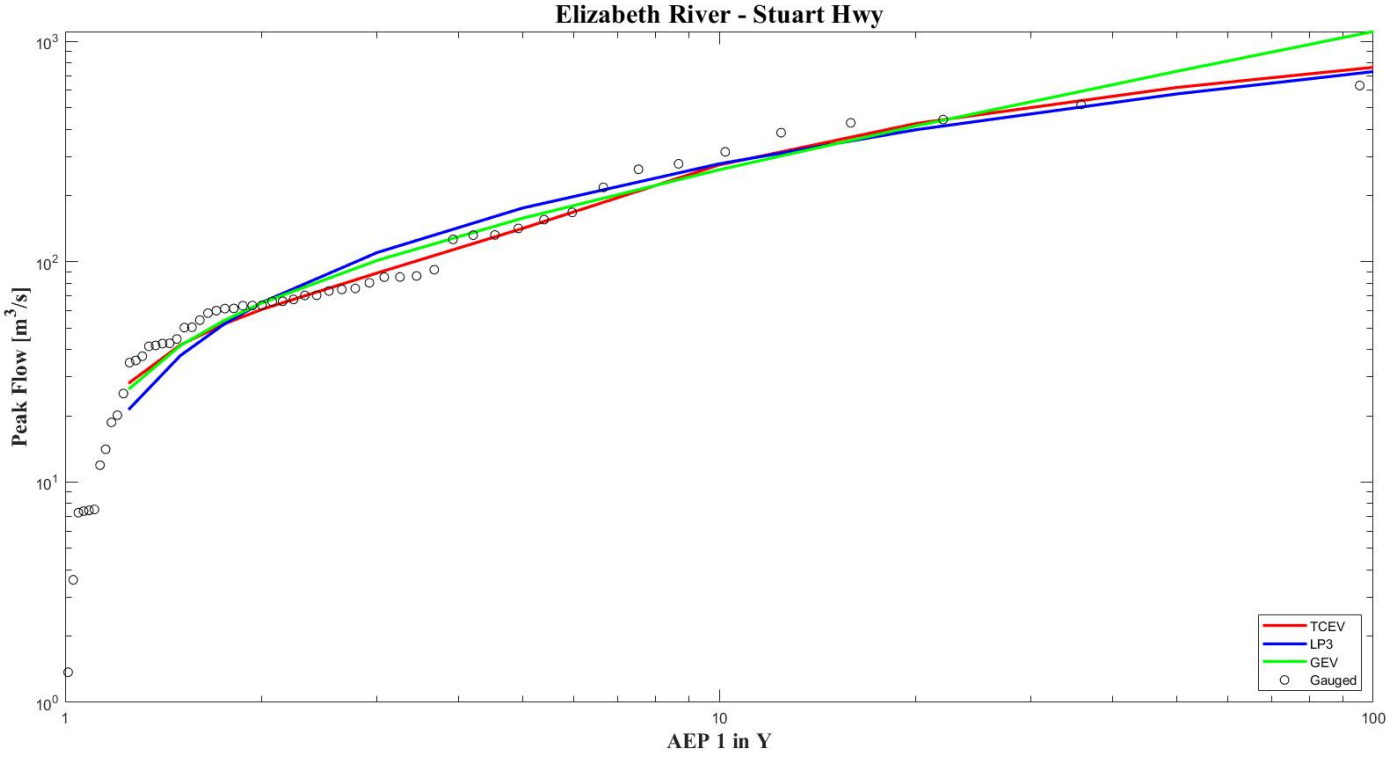

figure 6.34– TCEV, LP3 and GEV expected quantiles probability plots for Elizabeth River - Stuart Hwy

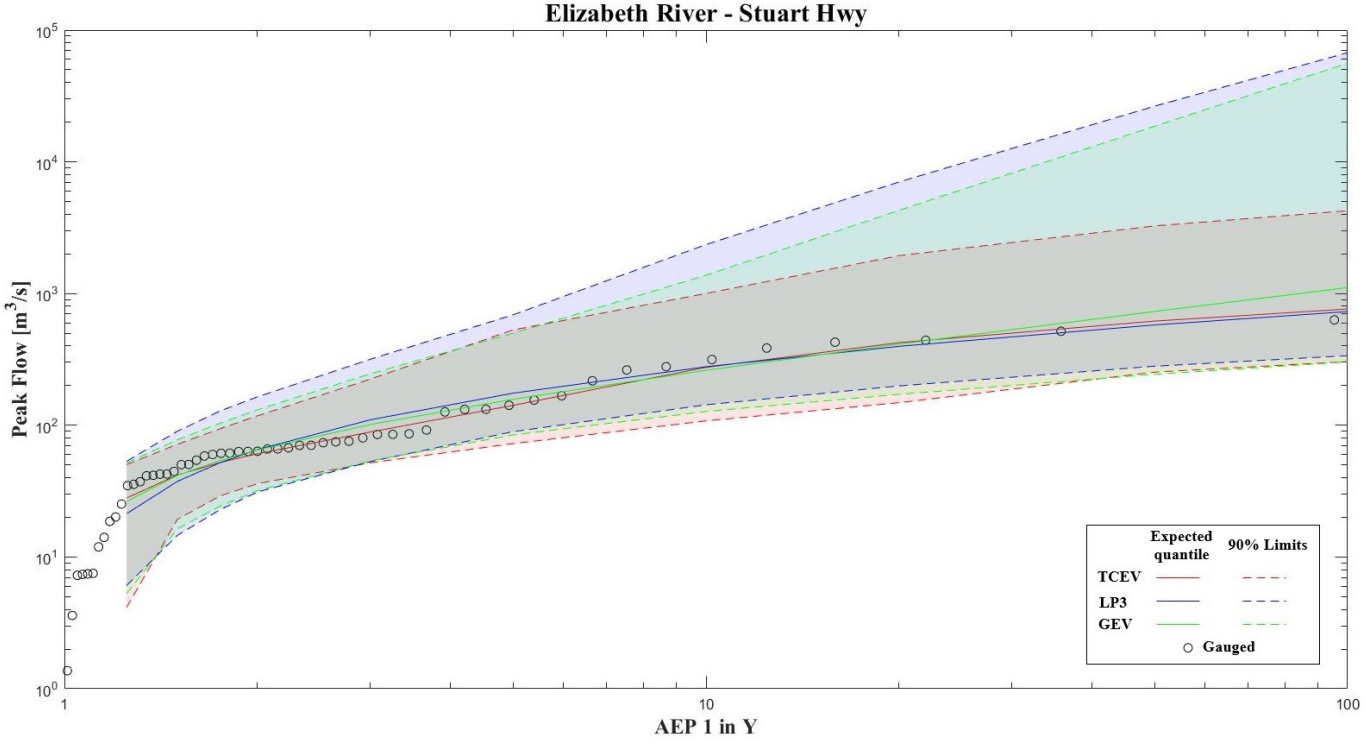

figure 6.35– TCEV, LP3 and GEV probability plot for Elizabeth River - Stuart Hwy

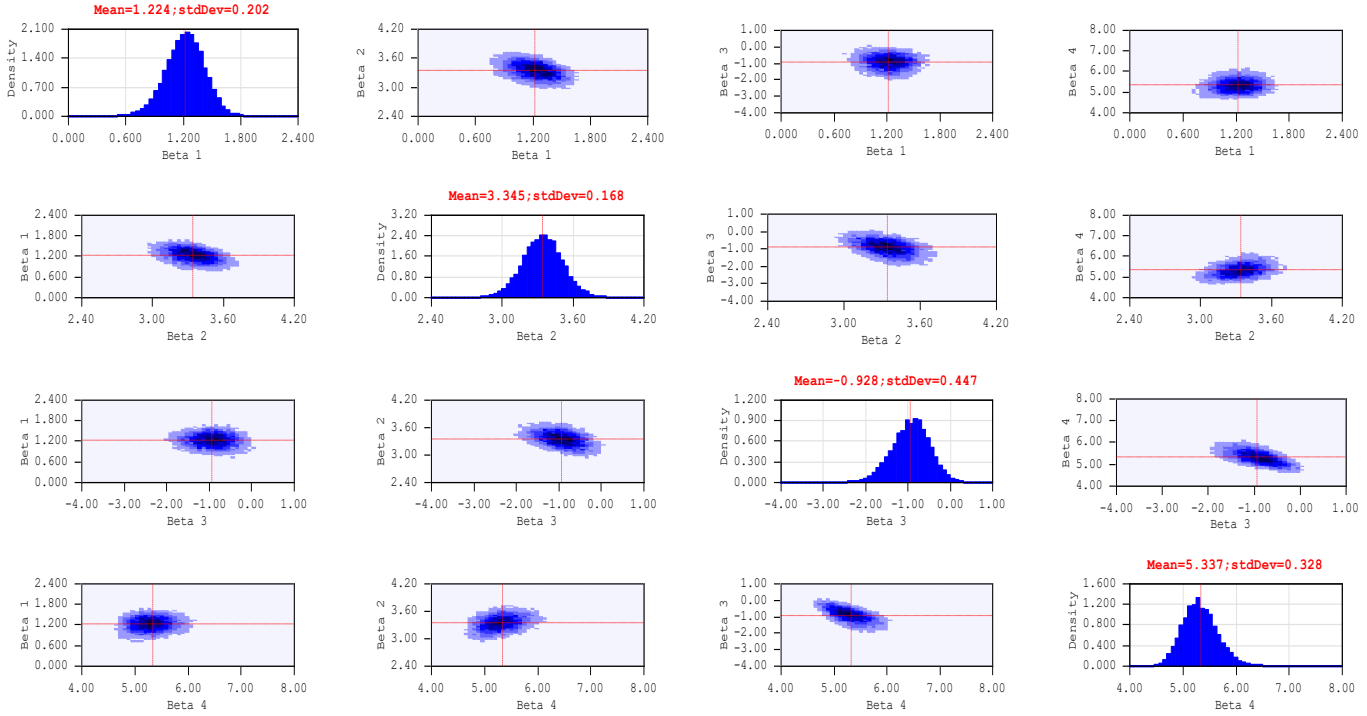

figure 6.36– TCEV posterior plots for Elizabeth River - Stuart Hwy

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	60.82	36.05	118
	LP3	65.43	31.15	164.5
	GEV	65.08	31.95	131.1
10	TCEV	275.14	107.91	1001.6
	LP3	278.34	143.06	2352.9
	GEV	261.64	127.21	1376.8
50	TCEV	617.93	253.24	3261.8
	LP3	577.6	280.7	26441.1
	GEV	731.41	243.73	18635.2
100	TCEV	763.02	304.42	4252
	LP3	729.59	337.65	67407.2
	GEV	1109.35	302.81	56052.7

Table 6.16 - Comparison of Selected Quantiles with 90% CL for Elizabeth River - Stuart Hwy

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.22352	3.34517	-0.9283	5.33658
Std Dev	0.20165	0.16801	0.44713	0.32791

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	4.08458	0.23374	-0.45916
Std Dev	0.16823	0.10482	0.2826

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	45.929	3.849	-0.574
Std Dev	7.347	0.158	0.154

Tables 6.17 – Comparison of estimates parameters for Elizabeth River - Stuart Hwy

Northern Territory: Finniss River - Gitchams Xng

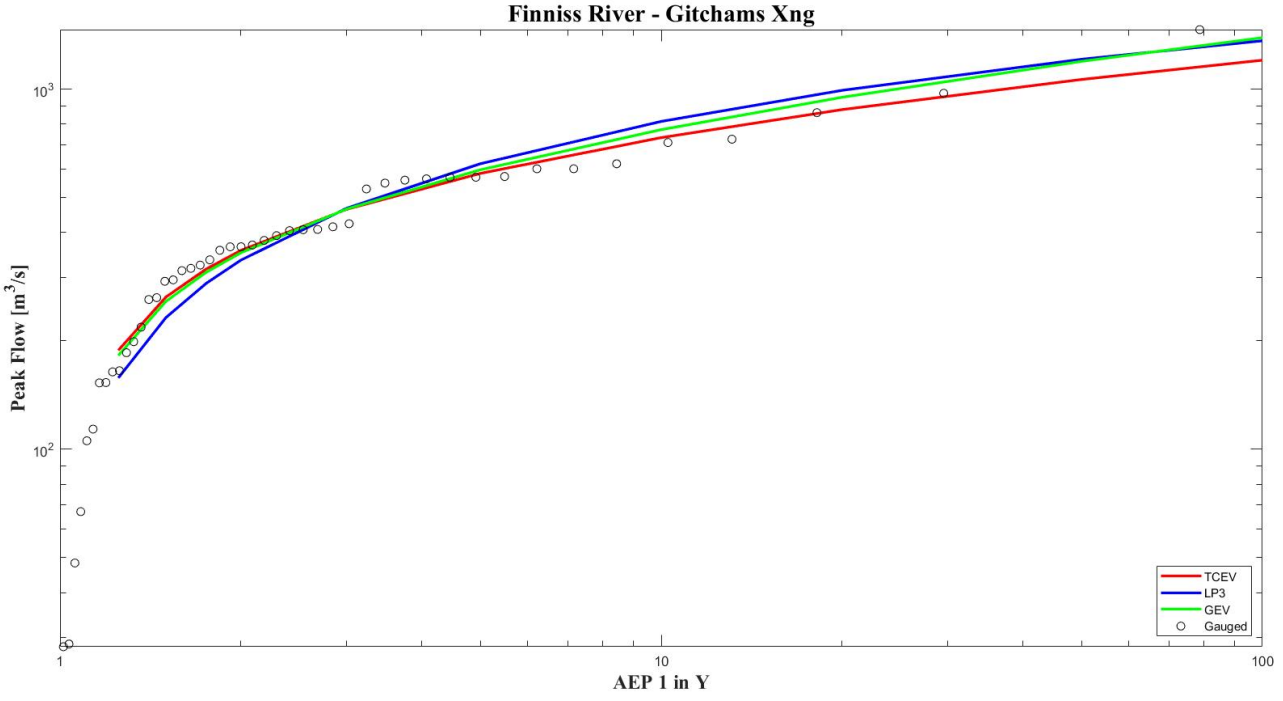

figure 6.37– TCEV, LP3 and GEV expected quantiles probability plots for Finniss River - Gitchams Xng

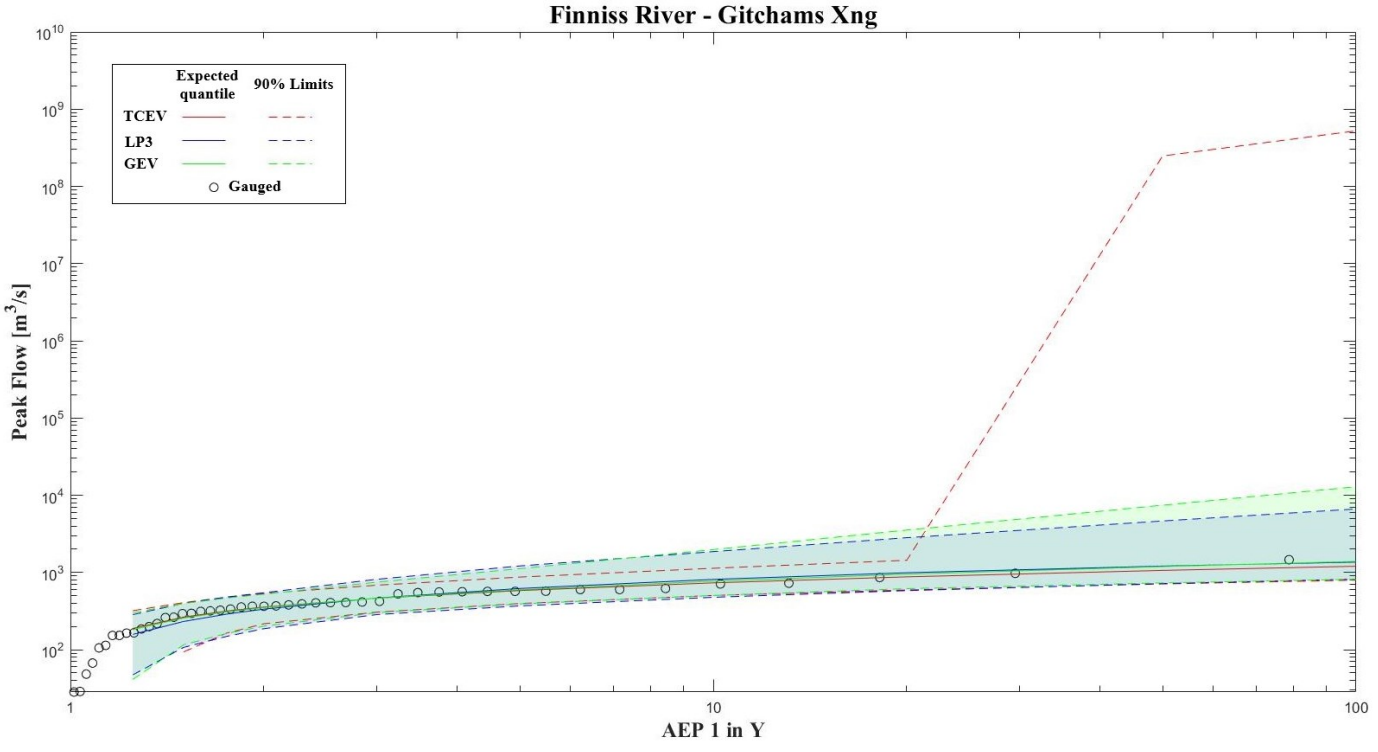

figure 6.38– TCEV, LP3 and GEV probability plot for Finniss River - Gitchams Xng

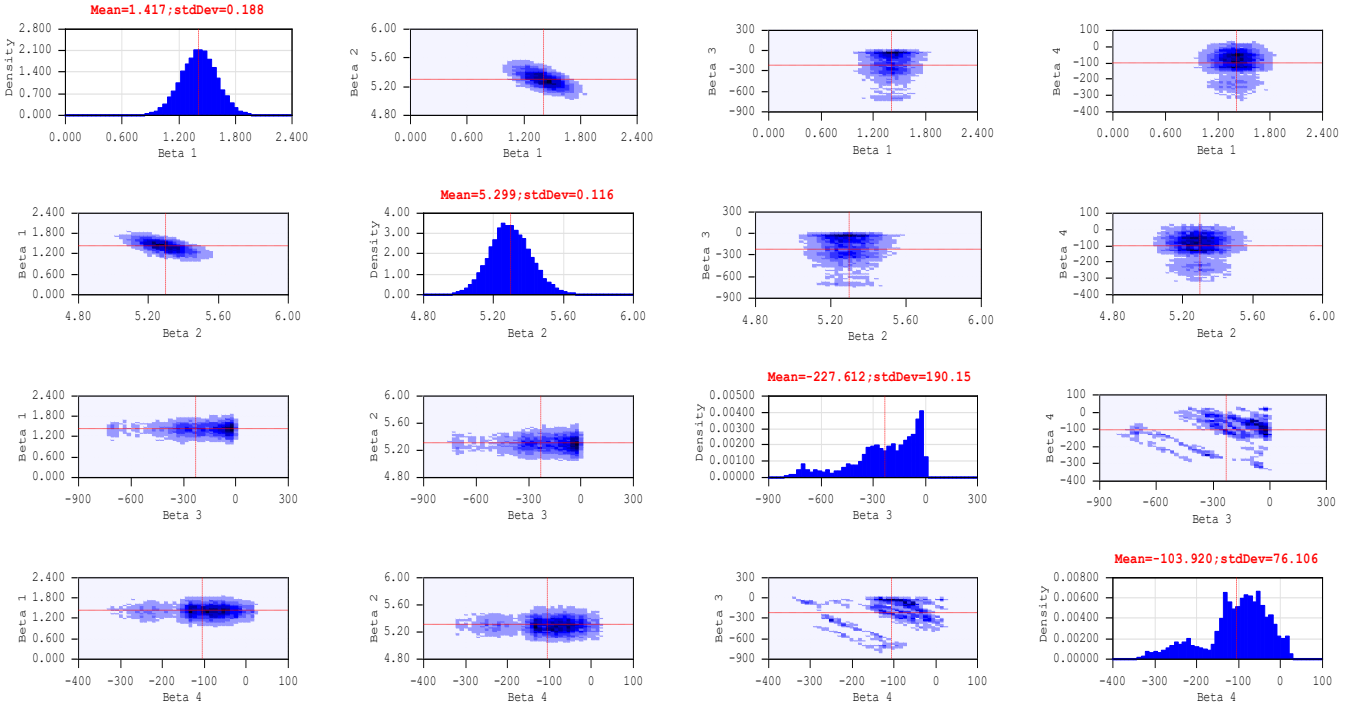

figure 6.39– TCEV posterior plot for Finniss River - Gitchams Xng

AEP 1 in Y	Distribution	Expected Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	356.81	216.85	535.4
	LP3	335.35	187.24	546.9
	GEV	351.43	201.68	523.8
10	TCEV	733.64	504.34	1134.3
	LP3	813.73	477.08	1863.1
	GEV	772.12	504.37	1979.4
50	TCEV	1064.01	709.15	247499232
	LP3	1210.71	718.41	4645.8
	GEV	1194.61	732.39	7434.5
100	TCEV	1203.68	795.73	523175296
	LP3	1362.62	820.25	6648.3
	GEV	1390.01	815.39	12895.5

Table 6.17 - Comparison of Selected Quantiles with 90% CL for Finniss River - Gitchams Xng

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.41724	5.29848	-227.612	-103.92
Std Dev	0.18827	0.11616	190.1551	76.1064

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	5.71822	-0.18447	-0.70488
Std Dev	0.1226	0.11941	0.23246

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	276.436	5.307	-0.077
Std Dev	33.212	0.127	0.114

Tables 6.18 – Comparison of estimates parameters for Finniss River - Gitchams Xng

Plots from 6.16 to 6.39 are a very precious source of information. Relatively to the three New South Wales sites reported (Apsley River at Apsley Falls, Rutherford Creek at Brown Mountain, Tantawangalo Creek at Tantawangalo Mountain, figs. 6.16-6.24) it should to be noted the extraordinary ability of TCEV distribution in fitting the particular "S-shape" of observed data. This behavior is also confirmed by the observation that there is an inversion in curvature in quantile curves when shifting from low to high AEP 1 in Y. LP3 distribution, which adequateness in describing flood data in NSW has been pointed out in some studies (Micevski et al., 2006; Franks and Kuczera, 2002), gives good fits too, but behavior is different if compared with TCEV's. Log-Pearson 3 distribution, in fact, denote an inadequate fit to data in the left tail, which is not able to model changes in shape from low to high values. However, 90% limits indicate that ability of TCEV in modulating the left tail is unbalanced by a relevant decreasing in uncertainty for AEP 1 in Y lower than 1.5. Instead, for high AEP 1 in Y (from 20 to 100) performances of TCEV and LP3 in expected quantiles are comparable, giving a good fit to data, but TCEV shows a relevant reduction in related uncertainty. This is clearly indicated by the divergence in 90% limits showed in LP3 cases.

For those sites, is of relevance noticing the poor fit of GEV. This distribution is, in fact, unable to fit the right tail, leading to a marked overestimation of quantiles.

A similar framework can be reported for sited in Queensland located below the latitude of 23° south (fig. 6.25-6.30). For Bell Creek at Craiglands (figs. 6.25-6.27) performances of TCEV are surprisingly better than those of LP3, but some investigations are needed to PILFs, that seems affecting more LP3 than TCEV.

Figs. 6.28-6.30 (Albert River at Bromfleet) show, instead, how tendency of TCEV in giving a double curvature to expected quantile, sometimes can led to misleading results. In this case, in fact, LP3 is the preferable distribution.

As for NSW, also in this zone of Queensland GEV gives poor fits to data.

However, the presence of an "S-shape" in data, that emphasizes skills of TCEV compared to those of LP3, is not always well pronounced, and is often influenced

by the presence of PILFs. In those cases, LP3 and TCEV give comparable results to expected quantiles fitting. Uncertainty in quantile estimates can be retained an affordable measure in model selection, and generally led to give preference to TCEV distribution.

Analysis of sites located in the northern part of Queensland confirms the general inappropriateness of TCEV in modeling their data (except for some exceptions). Liverpool Creek at Upper Japoonvale (figs. 6.31-6.33) is an interesting example. In fig. 6.26 can be observed that TCEV estimates in upper quantiles has high uncertainty, which reflect the structure of its posterior density (fig. 6.33). LP3 and GEV, instead, have comparable fits to higher quantiles, also when looking at 90% limit.

Two sites are, finally, reported for Northern Territory. In the first case (Elizabeth River - Stuart Hwy, figs. 6.34-6.36) a good fit of TCEV has to be noticed, also if low flows seem to affect in a significant part its ability to model the "S-shape". However, TCEV fits is much better than those of PL3 and GEV.

The second case (Finniss River - Gitchams Xng, figs. 6.37-6.39) is similar to Liverpool Creek at Upper Japoonvale, with TCEV showing am high uncertainty for high 1 in Y AEP. Can be observed the good fit of GEV, that shows a shape parameter close to zero: this can subtend the opportunity of using a two-parameter distribution (e.g. Gumbel).

Summarizing, visual inspection of probability plots confirmed previous findings arose from analysis of L-Moments Ratio Diagram and TC_{ratio} plots about the possibility of using a four-parameter TCEV distribution in modelling at-site time series of annual maximum of peak flows. In addition, the ability of Bayesian analysis in reliable quantification of connected uncertainty, provided additional support of TCEV in at-site flood frequency analysis for several sites in New South Wales and Queensland.

6.3.1.4 – AIC and BIC criteria

The use of these two criteria (described in paragraph 3.2.2.1) for model selection in Australia was proposed by Rahman et al. (2013). In this study were applied also for a comparison on previous work on the same subject and dataset.
Figs. from 6.40 to 6.42 show, respectively, a graphical representation of AIC, AIC_c and BIC tests, while tables 6.19-21 present a summary of outputs of these tests.

	New South Wales		
	AIC	AIC_c	BIC
LP3	75.00%	58.33%	75.00%
TCEV	25.00%	41.67%	25.00%
GEV	0.00%	0.00%	0.00%
N.C.	0.00%	0.00%	0.00%

Table 6.19 – AIC, AIC_c and BIC results for NSW

	Queensland		
	AIC	AIC_c	BIC
LP3	66.67%	62.16%	65.77%
TCEV	23.42%	31.53%	24.32%
GEV	8.11%	4.50%	8.11%
N.C.	1.80%	1.80%	1.80%

Table 6.20– AIC, AIC_c and BIC results for Queensland

	Northern Territory		
	AIC	AIC_c	BIC
LP3	63.16%	62.16%	75.00%
TCEV	15.79%	31.53%	25.00%
GEV	10.53%	4.50%	0.00%
N.C.	10.53%	1.80%	0.00%

Table 6.21 – AIC, AIC_c and BIC results for Northern Territory

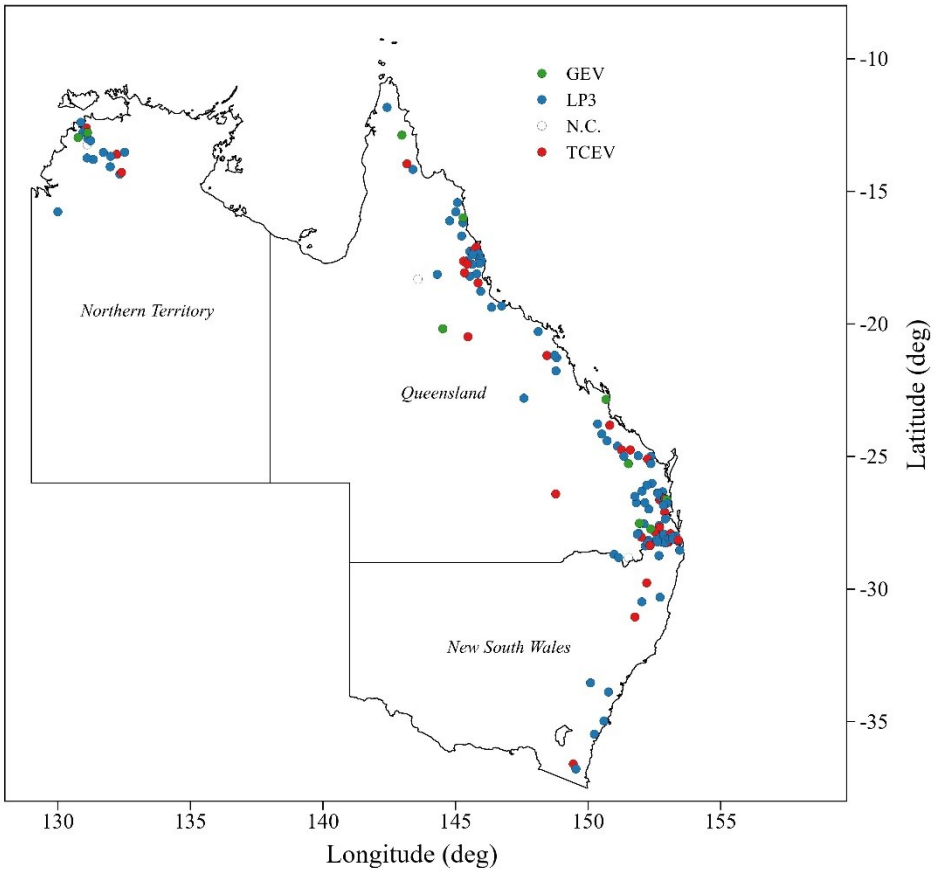

fig. 6.40 – Akaike Information Criterion results (N.C. stands for Not Computable)

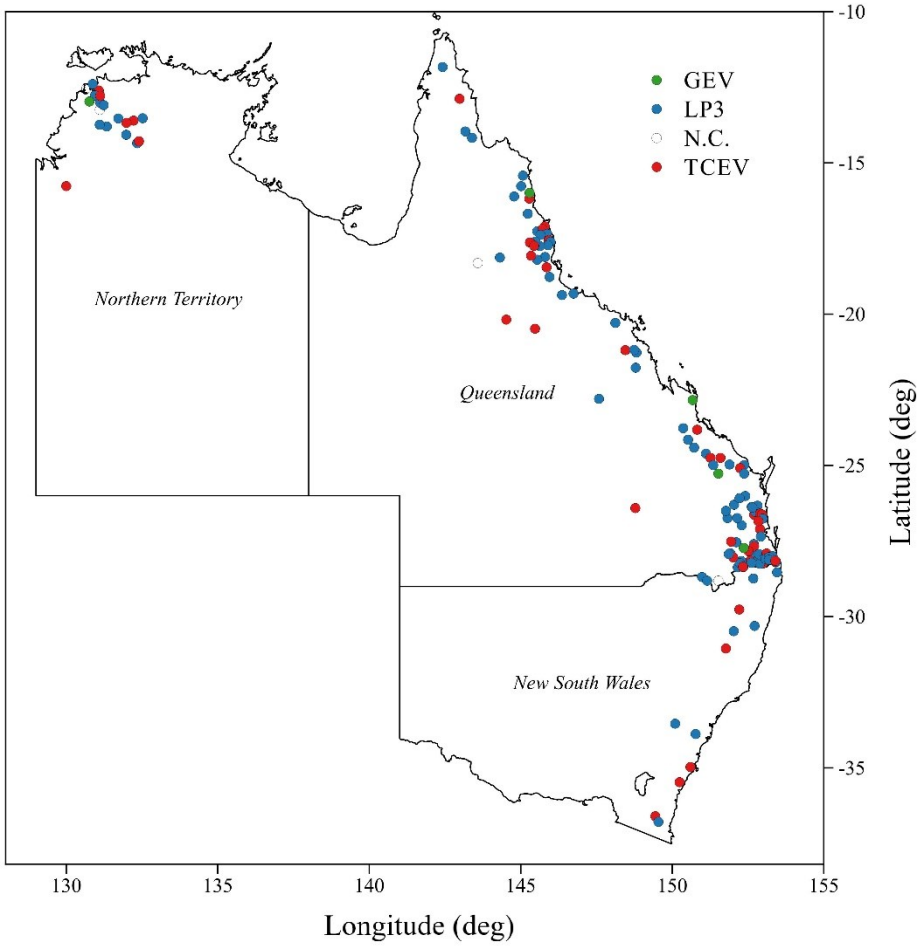

fig. 6.41 – Corrected version of Akaike Information Criterion results (N.C. stands for Not Computable)

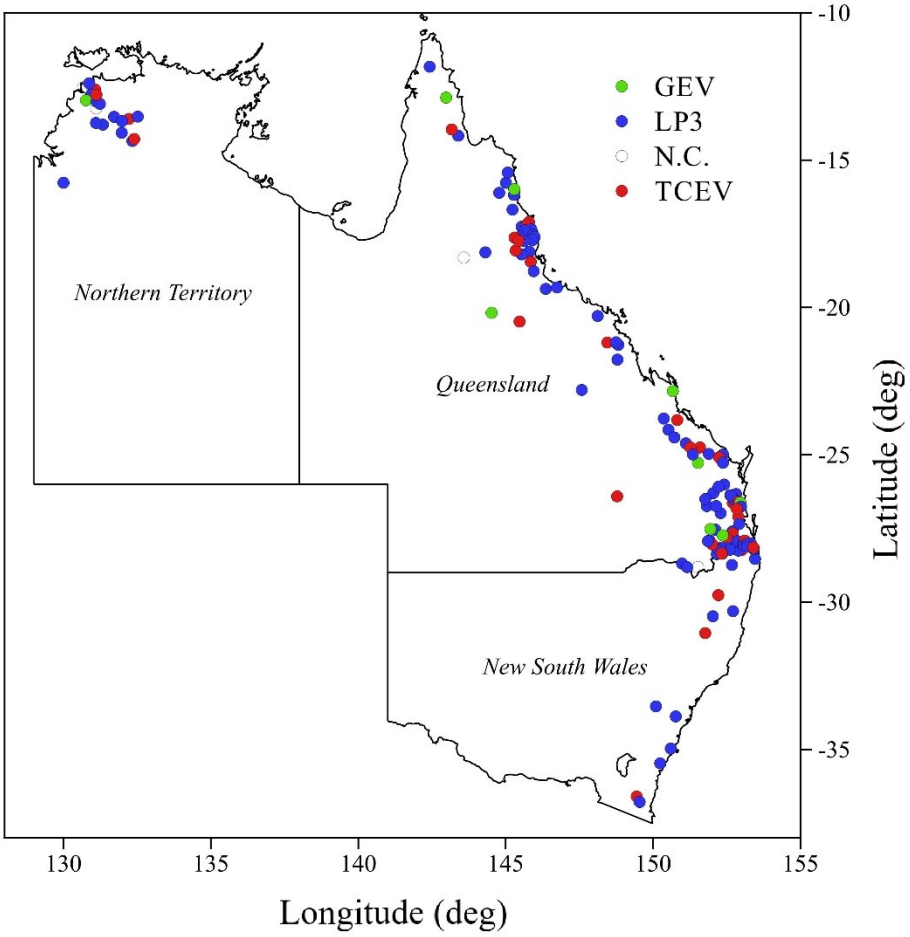

fig. 6.42 – Bayesian Information Criterion results (N.C. stands for Not Computable)

Results highlighted that AIC and BIC gave similar results, with a marked preference in the selection of LP3. Modified version of AIC, AIC_c, led to an increasing in the number of sites for which TCEV is the most suitable distribution. However, it is a not surprising result. Laio et al. (2009) documented that both AIC and BIC give preferences to log-transformed variables. This characteristic can affect the ability of these tests in detecting TCEV as the best performant, giving

preference to LP3. Finally, the role of parsimony principle must be taken into account for selecting a three-parameter distribution with respect to TCEV and Kappa.

6.3.2 – The role of uncertainty

Although important, a satisfactory fit to observed data can be regarded just a *conditio sine qua non* for judging adequate the choice of a certain distribution. One of the main advantages in performing a Bayesian analysis relies in its ability of accounting explicitly for uncertainty. In sub-paragraph 6.3.1.3 probability plots for different distribution were attached. Focusing the attention only on sites in the macro-region composed by New South Wales and the part of Queensland below the latitude of 23° south, it was noticed how TCEV and LP3 expected probability quantiles gave good results in fitting observed data. On the contrary, GEV distribution failed in interpreting upper quantiles. These statements can find confirmation by the figures 6.43-6.46, that illustrate boxplots of expected quantiles for different 1 in Y AEP for the above-mentioned macro region.
As emerges from figs. 6.43 and 6.44, the median of all boxplots does not show relevant mutual differences. When moving to 1 in Y AEP equal to 50 (fig. 6.55) and 100 (fig. 6.46) this quality is kept constant only by TCEV and LP3 distributions, while an overestimation for GEV is denounced.
These results give additional confirmations to the findings of previous paragraph.

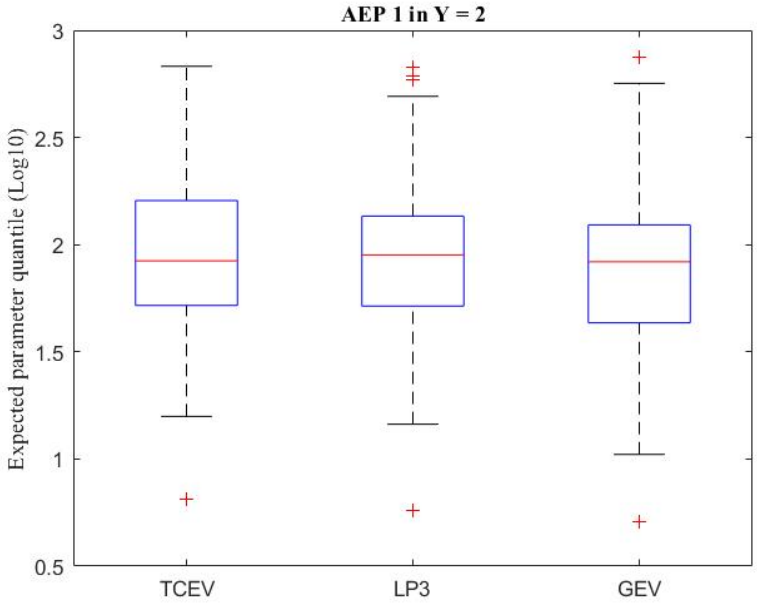

fig. 6.43 – Boxplot of expected parameter quantile for 1 in 2 AEP

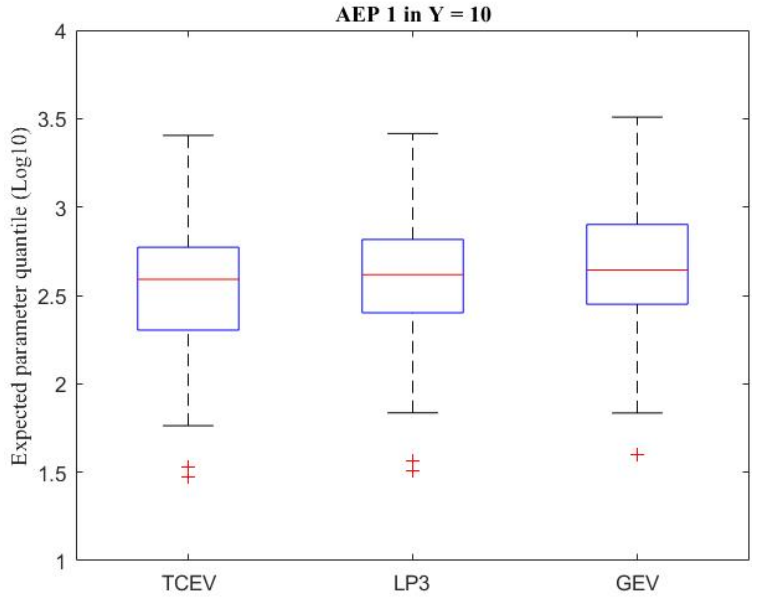

fig. 6.44 – Boxplot of expected parameter quantile for 1 in 10 AEP

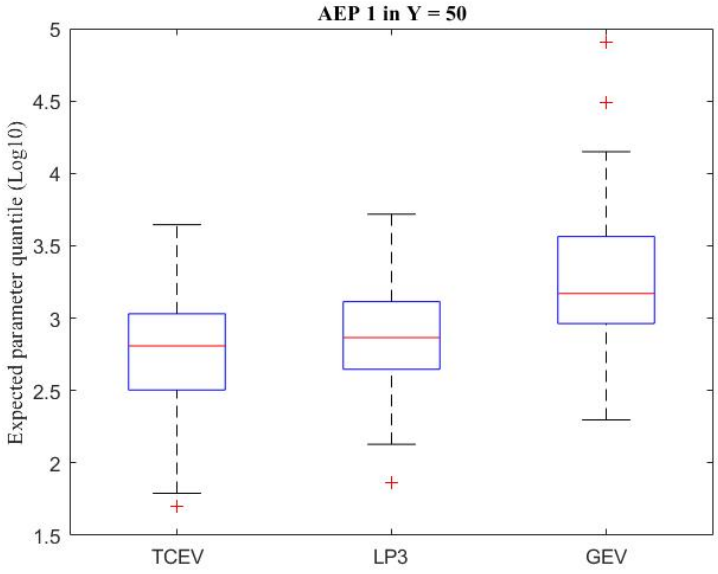

fig. 6.45 – Boxplot of expected parameter quantile for 1 in 50 AEP

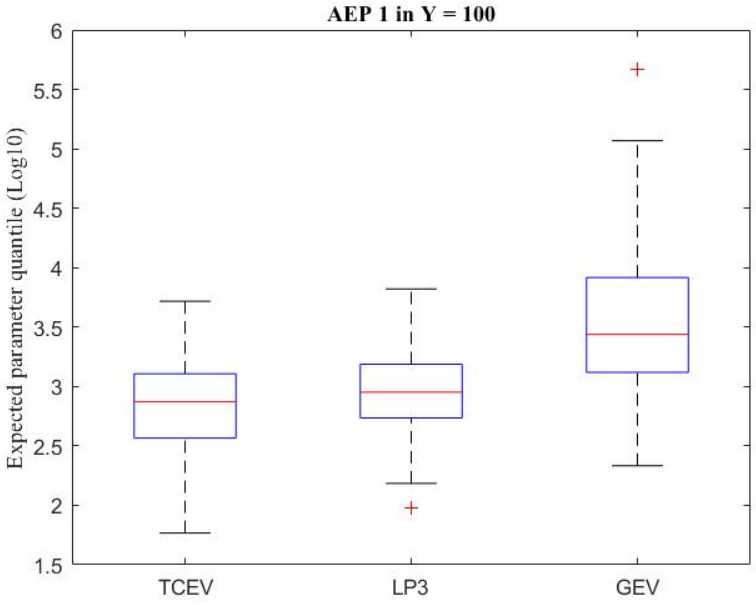

fig. 6.46– Boxplot of expected parameter quantile for 1 in 100 AEP

The substantially equivalence between TCEV and LP3 in fitting data in this macro region needs further investigations. This reflect all considerations about model selection, that in this case is reflected in the choice between a three- and a four-parameter probability distribution. Parsimony principle would suggest to opting for the less parametrized distribution, i.e. LP3. However, the role of uncertainty in estimates, above all of quantiles, cannot be neglected in this phase.
Quantifying uncertainty as the amplitude of the 90% confidence limits for each 1 in Y AEP, figs. 6.47-6.50 gave a clear summary its distribution.

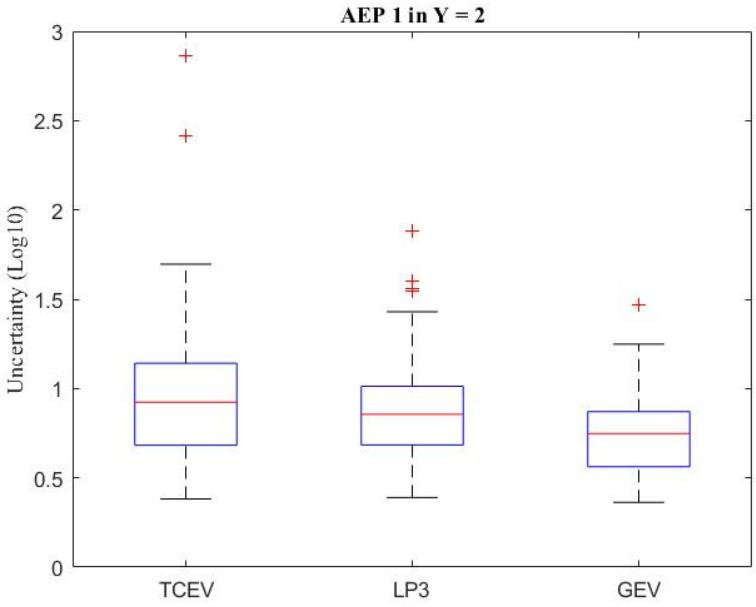

fig. 6.47 – Boxplot of quantile uncertainty for 1 in 2 AEP

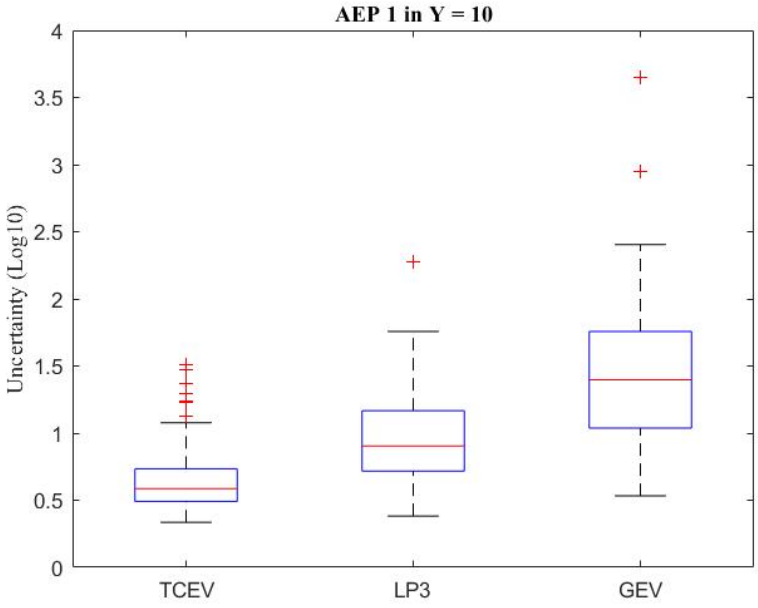

fig. 6.48 – Boxplot of quantile uncertainty for 1 in 10 AEP

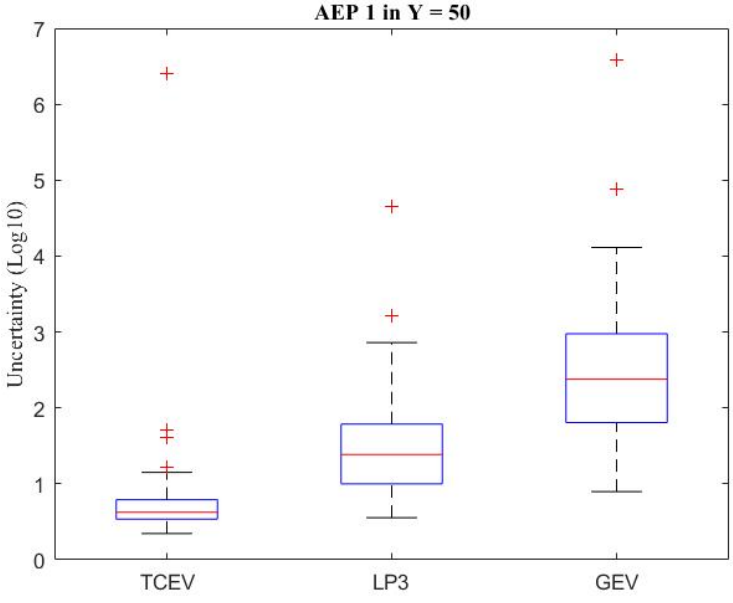

fig. 6.49 – Boxplot of quantile uncertainty for 1 in 50 AEP

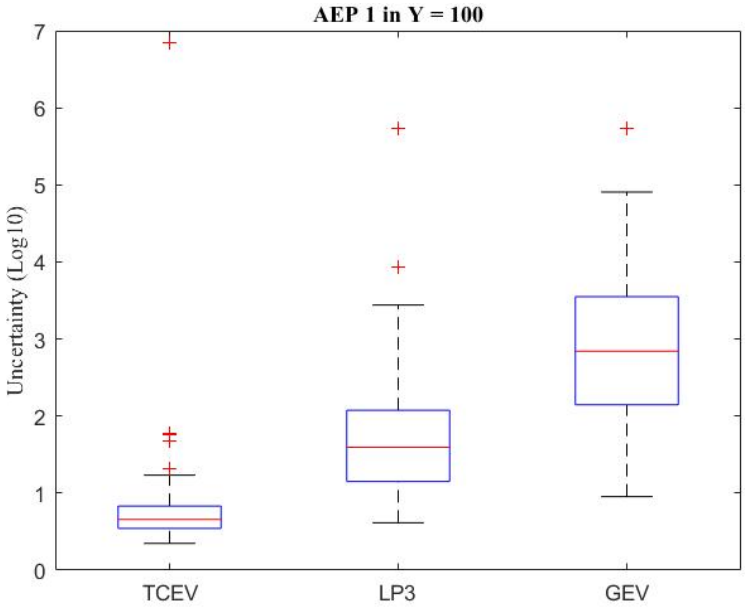

fig. 6.50 – Boxplot of quantile uncertainty for 1 in 100 AEP

These boxplots give a relevant contribute to the discussion. In fact, also in this case a confirmation of findings of visual analysis emerged. For small 1 in Y AEP, TCEV shows the highest uncertainty with respect to LP3 and GEV (fig. 6.47). This state dramatically changes moving from 10 to 100 1 in Y AEP, where TCEV is generally characterized by the lowest uncertainty.

6.4 – An Italian case study: Venosa River at Ponte Sant'Angelo (Basilicata region, Southern Italy)

Venosa River flows in Basilicata, a region of Southern Italy. Interest in application of the above illustrated procedure to this case study arise from some studies conducted on the topic of theoretically derived distributions of floods. In this respect, Gioia et al. (2008) proposed a scheme for flood frequency derived distributions that includes two different threshold driven mechanisms of runoff

generation. Each of these mechanisms is responsible for ordinary and extraordinary events: the first is characterized by frequent occurrences while the second one produces rare events and higher average of exceedances. These concepts were summarized by the authors in the TCIF (Two Component Iacobellis and Fiorentino, TCIF, distribution). One of the sites for with they found a good correspondence to this theoretical behavior was Venosa at Ponte Sant'Angelo.
It seems coherent with the aim of this thesis to investigate the fit of TCEV distribution. As for the Australian cases study, TCEV, Kappa, LP3 and GEV distributions were fitted. Results are shown in figs. 6.51-6.53, while fig. 6.54 reports the outcome of TC_{ratio} measure. Due to numerical complications, plots for Kappa distribution are omitted.

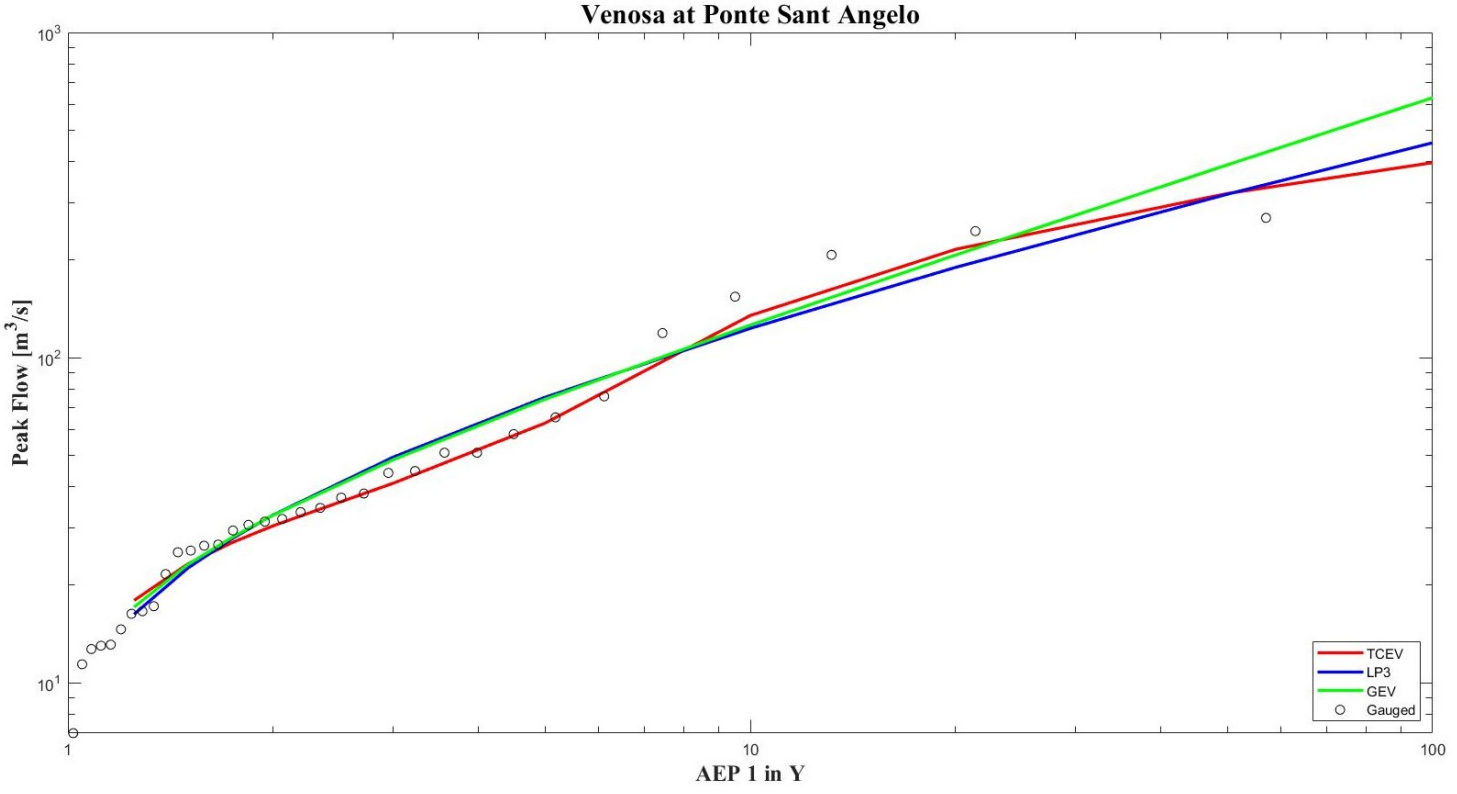

figure 6.51 – TCEV, LP3 and GEV expected quantiles probability plots for Venosa at Ponte Sant'Angelo

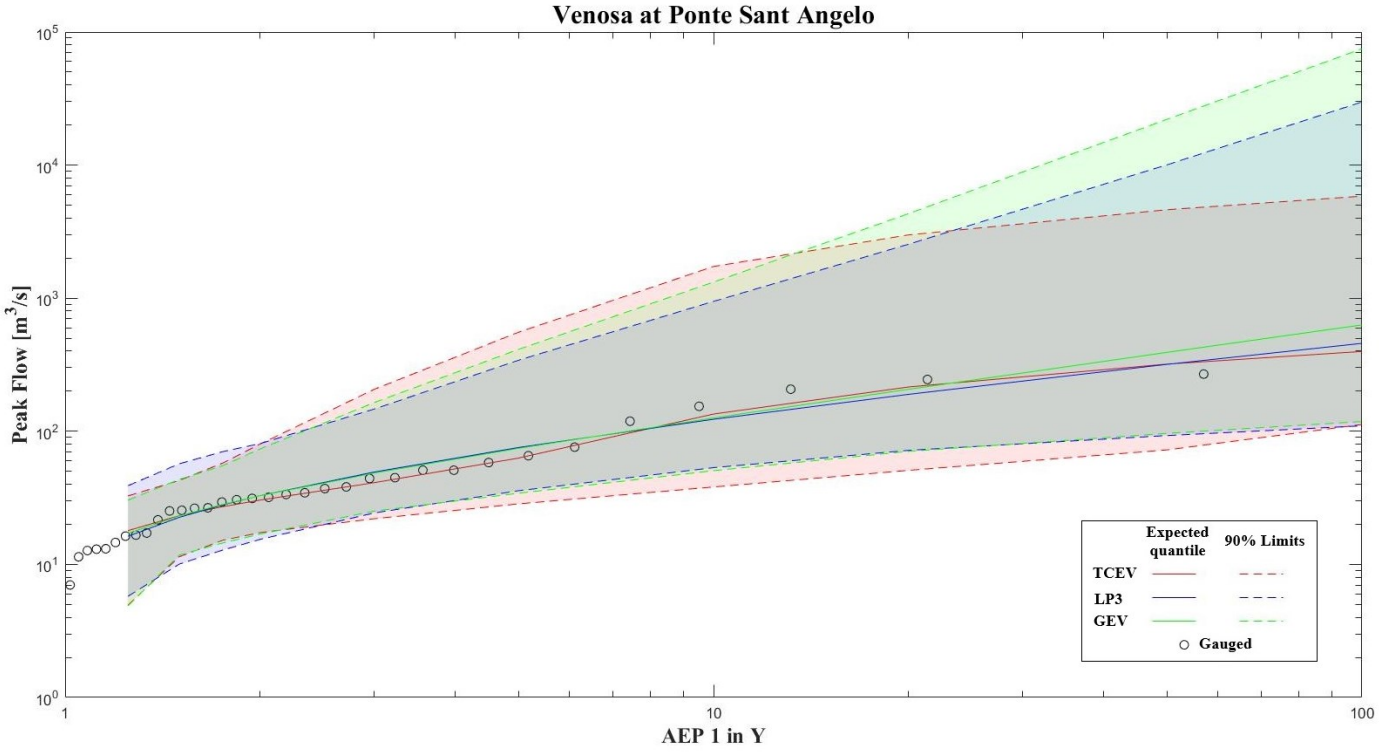

figure 6.52 – TCEV, LP3 and GEV probability plot for Venosa at Ponte Sant'Angelo

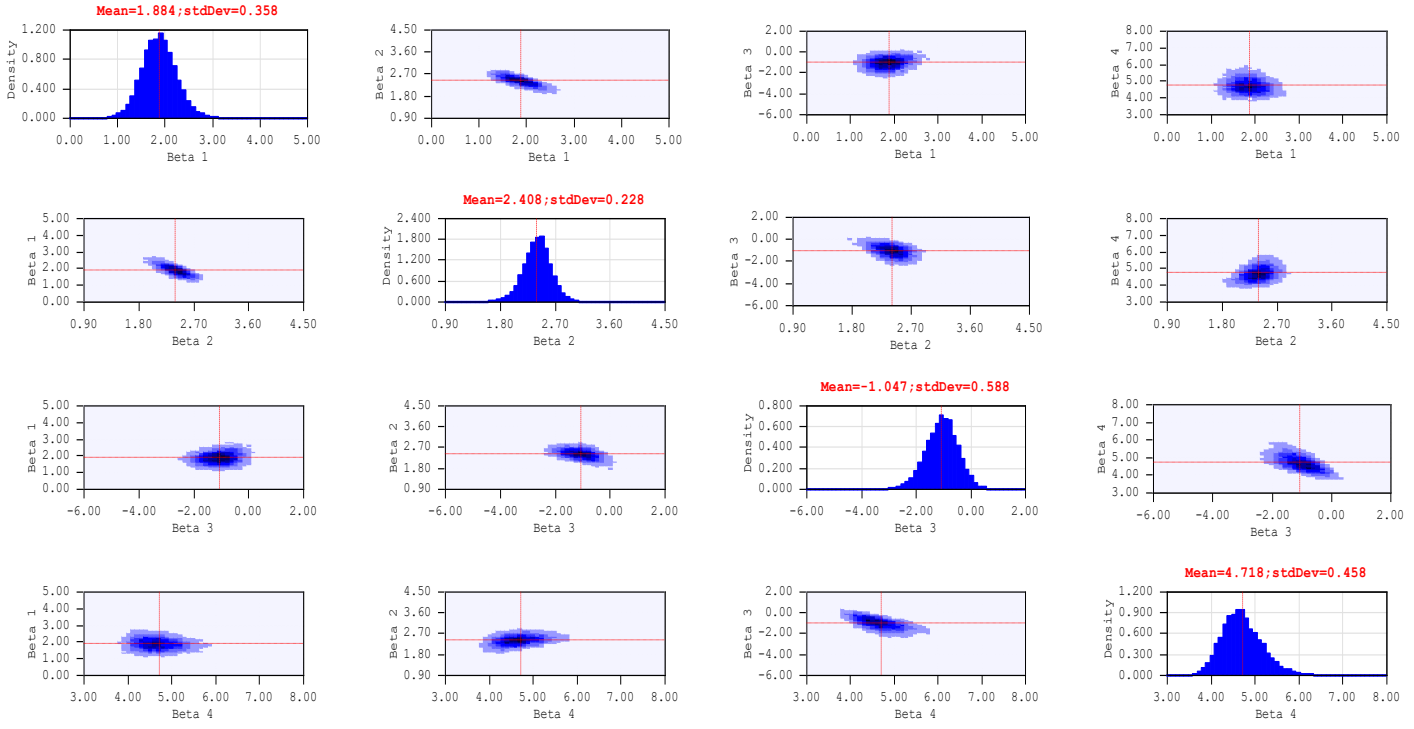

figure 6.53 – TCEV parameters probability plot for Venosa at Ponte Sant'Angelo

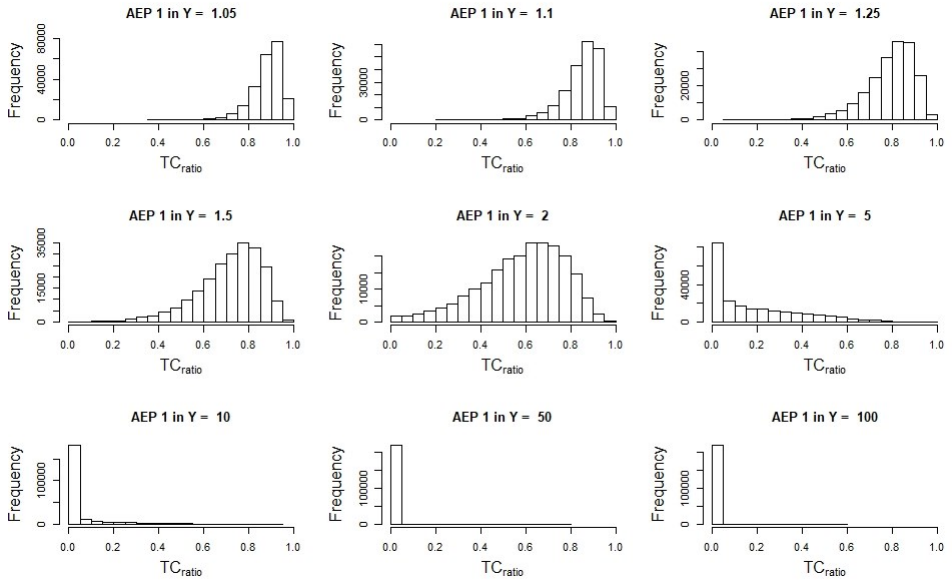

figure 6.54 – TC_{ratio} diagram for Venosa at Ponte Sant'Angelo

It is interesting to note how, according to the theoretical behavior of this basin, two components are effectively detected by the TCEV distribution, showing another quality of this model.

Tables 6.19-20 show, finally, how for AEP 1 in 50 TCEV provides lowest uncertainty with respect to the other distributions.

AEP 1 in Y	Distribution	Expcted Parameters Quantile	Quantile Confidence Limits 5% Limit	Quantile Confidence Limits 95% Limit
2	TCEV	30.41	17.34	80.3
	LP3	32.85	15.42	80.9
	GEV	32.79	16.87	73.8
10	TCEV	134.75	38.14	1731.3
	LP3	123.04	53.3	946.3
	GEV	125.85	50.57	1321.8
50	TCEV	319.61	72.39	4619.7
	LP3	318.1	92.83	10035.9
	GEV	391.46	96.38	22043.4
100	TCEV	397.78	112.47	5840.8
	LP3	457.84	109.88	29795.6
	GEV	629.76	118.37	74788

Table 6.19- Comparison of Selected Quantiles with 90% CL for Venosa at Ponte Sant'Angelo

	TCEV			
	$\ln(\Lambda_1)$	$\ln(\theta_1)$	$\ln(\Lambda_2)$	$\ln(\theta_2)$
Mean	1.88393	2.40764	-1.04715	4.71812
Std Dev	0.35783	0.22835	0.58784	0.45799

	LP3		
	Mean (ln flow)	loge [Std dev (ln flow)]	Skew (ln flow)
Mean	3.58222	-0.07628	0.58673
Std Dev	0.16314	0.14128	0.36238

	GEV		
	Location ζ	ln(Scale σ)	Shape ε
Mean	24.857	2.948	-0.676
Std Dev	3.803	0.220	0.192

Tables 6.20 – Comparison of estimates parameters for Venosa at Ponte Sant'Angelo

6.5 – Discussion

Goal of this chapter was to evaluate the suitability of the use of TCEV and Kappa four-parameter distributions for flood frequency analysis in three regions of Australia, i.e. New South Wales, Queensland and Northern Territory. Results shown a good fit of TCEV for most of sites located in New South Wales and in Queensland, except for the northern part of this latter state, where, together with Northern Territory, LP3 shows the better fit.

The TCEV ability in modelling "S-shape" behavior (that cannot be handled by a three-parameter distribution) in observed data is relevant, Usually, a practical approach for dealing with this issue is to censor some values (using the multiple Grubbs Beck). However, it is not surprising that GEV distribution do not provide good results for most sites: L-Moments ratio diagram showed that sample points on this diagram lies far from related line.

Some notes are required for describing and justifying analytical problems for Kappa distribution. In most cases, the implemented algorithm did not provide solutions. More in detail, was often impossible to compute the covariance matrix needed for implementing the Metropolis-Hastings algorithm. Only for some sites it was possible to obtain a result. An example is illustrated in fig. 6.55-6.56. In these figures, the problems in fitting the observed series can be noted. This is a common behavior for all sites for which Kappa was evaluated.

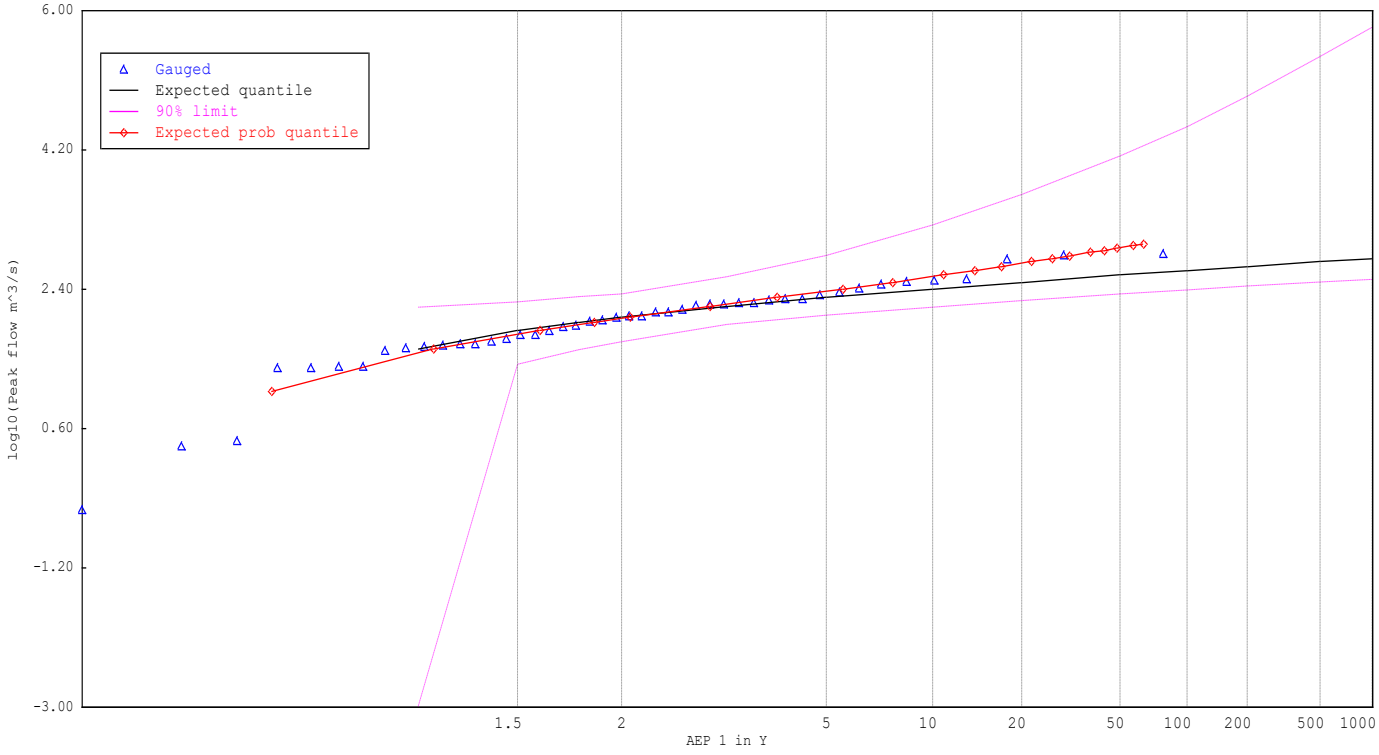

figure 6.55 – Kappa probability plot for Sandy Creek at Clermont (Qld)

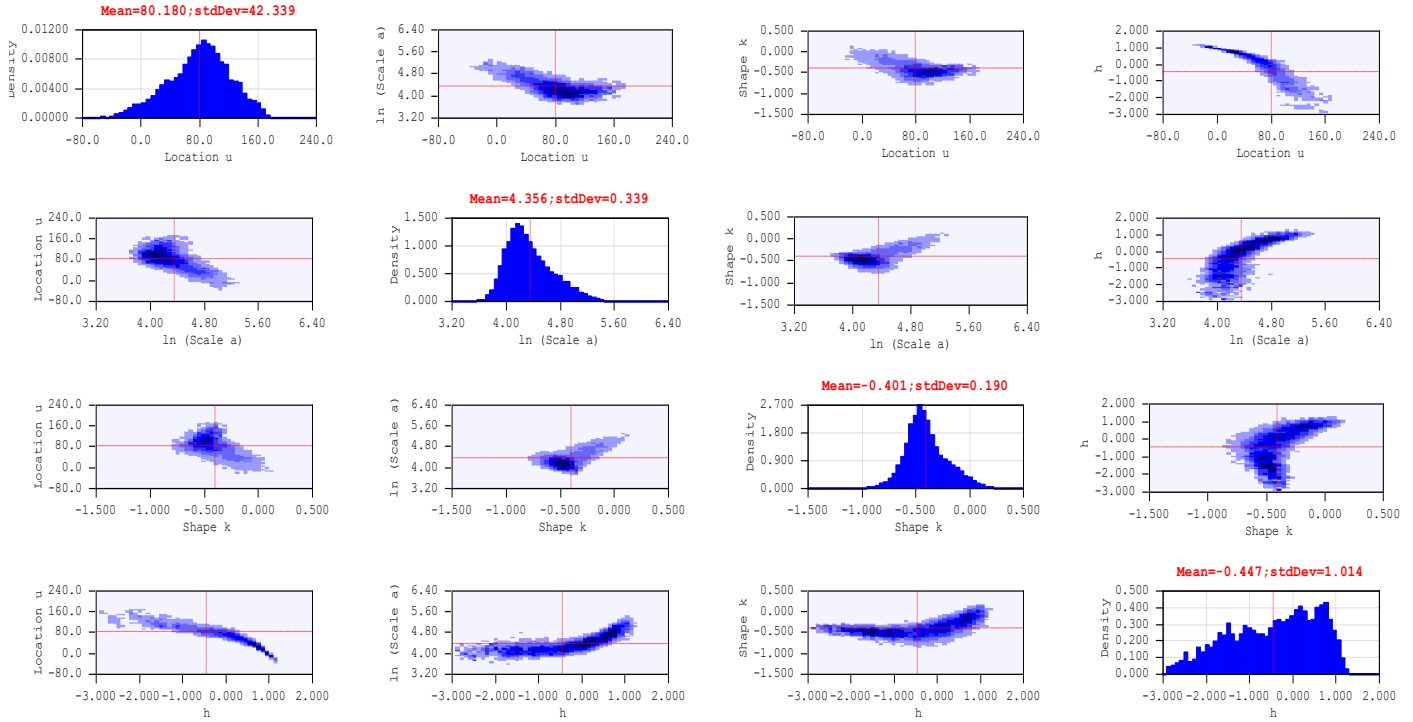

figure 6.56 – Kappa parameters probability plot for Sandy Creek at Clermont (Qld)

As noted in sub-paragraph 6.3.1.2, sites that show a good fit to TCEV distribution and that provide successful outcome to TC_{ratio} test, are clustered in a wide area of eastern Australia, which comprehend the whole NSW and a part of Queensland below the latitude of 23° south. This macro-region was recognized also in the work of Micevski et al. (2006) on multidecadal variability of floods in eastern Australia.

Moving into the framework that can be traced back to the works of Erskine and Warner (1988), Franks and Kuczera (2002) and Micevski et al. (2006) is fundamental for understanding the underlying mechanism to flood data. In fact, Franks and Kuczera (2002) showed how in these cases the classical hypothesis of independent and identical distributed annual maximum floods is questionable. Although floods are independent, a modulation of the Pacific Interdecadal Oscillation (IPO) on flood risk in New South Wales was detected. *Quae cum ita sint*, annual maximum floods cannot be considered anymore to be identically distributed. This can lead to relevant issues when dealing with short-term flood risk (Franks and Kuczera, 2002).

The characteristic of TCEV distribution of contemplating two different population can make this distribution eligible for interpreting the marginal distribution of the two populations. Results seems confirm this statement.

Furthermore, the Italian case study demonstrated the ability of TCEV in fitting samples characterized by the presence of different populations, each one linked to a precise condition of runoff generation. These conclusions can confirm an important ability of TCEV distribution in fitting particular series of flood data.

6.6 – Final remarks

Goal of this chapter was the evaluation of the applicability of TCEV and Kappa four-parameter distributions at-site analysis of annual maxima of peak flows for New South Wales, Queensland and Northern Territory (Australia). There is much

literature on flood frequency analysis in Australia, but no previous studies were found on the explicit use of TCEV and Kappa in this country for FFA.
The first step was to define an L-Moments Ratio Diagram, on which plotting also the area corresponding to TCEV distribution. This visual assessment of finding the candidate distributions was confirmed to be reliable, because of detected the presence of numerous sites in New South Wales and Queensland suitable for being in the TCEV "domain of attraction".
The second step was to apply the TC_{ratio} measure for discerning the presence of two different populations in a sample. There were two main results in this analysis:

1. TC_{ratio} revealed to be an affordable tool for detecting the presence of two components of different nature into a single population;
2. a cluster similar to that found by Micevski et al. (2006) was found in Queensland and New South Wales.

Visual inspection of probability plots corroborated the achievements of the first two phases of this preliminary analysis for this cluster. In fact, TCEV showed a great ability in fitting the "S-shape" of observed data, plotted with Cunnane's plotting position formula. Usually, such made data, when fitted with LP3 or GEV, require the censoring of part of the sample, that for TCEV seems to be not necessary. However, in some cases PILF produced a decrease in the skills of all distributions, TCEV included. While LP3 confirmed its goodness-of-fit (very similar to TCEV), GEV was considered inadequate for fitting data.
In Northern Queensland and Northern Territory, LP3 and GEV are the candidate distributions, where TCEV loses its abilities. TC_{ratio} test confirmed these conclusions.
In order to adopt also some parametric criteria widely used in literature, and increase the content of information on this analysis, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were computed for all sites. These

criteria showed a marked tendency in preferring LP3 distribution, also in TCEV was chosen for a not negligible number of sites. This result is in accordance with Laio et al. (2009), which denoted a particular behavior of these tests in preferring log-transformed distributions (LP3 in this case).

Focus was, then, moved only to the macro-region defined in 6.3.1.2. Comparative evaluation of uncertainty between TCEV, LP3 and GEV showed that TCEV has expected quantiles comparable with LP3, but lower uncertainty for 1 in Y AEP greater than 10. This result is really remarkable, because can led to an improvement of flood frequency analysis, were at-site approach to TCEV distribution was neglected since its introduction.

Floods in the eastern coast of Australia were investigated in different studies during last decades. For integrating this analysis, the conclusions of Franks and Kuczera (2002) and Micevski et al. (2006) are fundamental. These authors found that the assumptions of being independent and identically distributed of annual maximum flood for New South Wales (and, more in general, eastern Australia) is inconsistent, because of the modulation of Interdecadal Pacific Oscillation on flood risk. In those cases, the use a non-homogeneous probability models can be investigated. However, TCEV seems to represent the marginal distribution quite well, because of its ability in contemplating the presence of two different populations.
Finally, TCEV distribution showed a good fitting in basins characterized by two different mechanisms for runoff generation. This is an expected quality, because arises in the spite of its original formulation.

7- CONCLUSIONS

7.1 – Introduction

One of the main goals of this thesis was to implement a Bayesian estimation method that allows to perform an at-site analysis of TCEV and Kappa distributions with an appropriate quantification of uncertainty in parameters and quantiles estimates. It was a preparatory phase for testing the applicability of these distribution in a flood frequency analysis framework. In this conclusive chapters, the main findings of this thesis will be highlighted, with a final proposal for future research developments.

7.2 – The role of Kappa and TCEV distributions in the Extreme Value Theory

In chapter 2 the theoretical framework of extreme value theory was illustrated, with respect to two different theoretical derivations. The theory formulated by Todorovic and Zelenhasic, in particular, represents the basis for understanding the physical context in which these distributions were derived. TCEV distribution was indeed introduced considering the possibility that floods can be generated by two different mechanisms, one responsible for frequent events and the other of the rare. There two components were considered be composed by a Poissonian distribution of the number of occurrences in an annual interval, whilst their magnitude is defined by an exponential distribution.
In the case of Kappa, this distribution can be derived combining the hypothesis that the number of events in a year follows a binomial distribution, while the number of occurrences can be described by a Generalized Pareto.
It is thus clear the phenomenologically different underlying nature between these two distributions.

7.3 – A review on the frequentist use of TCEV and Kappa

All the most diffused findings about TCEV and Kappa distributions were derived in the framework of the frequentist inference. While for Kappa can be found in literature some reviews and investigations of its at-site estimations in terms of maximum likelihood, probability weighted moments and L-moments, for TCEV distribution the state of the art in more fragmented. In fact, although at-site estimation in very rare, it is difficult to find a complete and exhaustive review of its characteristics.

In Chapter 3 the theoretical background of classical estimation methods is provided, in order to be propaedeutic to the discussion about their implementation for the analysed distributions, illustrated in Chapter 4. In this latter section, after the descriptions of the theoretical aspects of these distributions, a focus on L-moments is given, particularly for TCEV. A description of is implementation is illustrated, and finally this distribution was introduced in the L-Moment Ratio Diagram. This allowed a clearer visual comparison with traditional candidate distributions in flood frequency analysis.

7.4 – The use of Bayesian inference for estimating and quantifying uncertainty

In chapter 5 the main concepts for describing Bayesian inference were discussed. Moving from the assumption that this way of making inference allows to an appropriate quantification to the uncertainty in parameter and quantile estimation, a Bayesian approach for their evaluation for at-site TCEV and Kappa was implemented. This goal was obtained modifying the source code of the software FLIKE (a software for Bayesian inference realized by prof. George Kuczera), and introducing all constraints and theoretical aspects of these distributions. Furthermore, this software was integrated with a MCMC Metropolis-Hastings algorithm for sampling the posterior distribution. Comparison with Importance Sampling (default method in FLIKE) allowed a

better discrimination between some particular degenerate cases. For example, when analysing a Gumbel distributed sample with TCEV distribution, a reliable interpretation for diagnose the parent distribution was recognized. In this way, was set an efficient analytical tool with a user-friendly interface able to give a useful interpretation of our results.
Sampled posterior distribution was exploited in order to define a visual tool able to discerning the presence of two different population in the same time series, TC_{ratio}. Based on the ratio between the first and the sum of the two components of TCEV distribution, TC_{ratio} revealed to be an efficient tool for concluding its proposal objective.

7.4 – Investigation of at-site flood frequency analysis for eastern and northern Australia

In chapter 6 an at-site frequency analysis for annual maximum of peaks flows was conducted for sites located in Australia, in particular in the regions of New South Wales, Queensland and Northern Territory. Four distributions were compared: GEV, LP3, TCEV and Kappa.
A first investigation for searching the candidate distribution was conducted with a visual analysis of L-Moments ratio diagram, that showed how TCEV seems a candidate distribution for fitting number of sites in NSW and in Queensland. In Northern Territory, instead, only a couple of records looked suitable for a good TCEV-fit. For the same reasons, Kappa denoted some problems in recognizig the underlying distribution.
The second step consisted in performing the visual test using TC_{ratio} measure for discriminating the presence of two populations into a single sample. This measure showed that for sites below the latitude of 23° south in the eastern Australia the presence of two different populations seems to be justifiable.
Then, a visual inspection of probability and posterior distribution plots showed how below the above-mentioned latitude TCEV fits in a comparable way with

LP3, which was considered advisable for that zone. This increasing in performances can be easily showed by the ability of TCEV in describing the "S-shape" of plotting position data. It was relevant to note that Kappa distribution denoted computational problems for most sites, making not possible a wide comparison with others distribution. Further investigations are required to fix this issue.

According with some authors that employed Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for model selection, these likelihood-based criteria were applied. From this analysis emerged that LP3 should be considered the preferable distribution for most of the analyzed sites, while TCEV has only a low percentage of acceptance as the best fit distribution. GEV was chosen only in some cases. Those results are not in conflict with previous findings, because these criteria were recognized choice more frequently log-transformed distributions.

Finally, because of differences in quantile estimates for all sites between TCEV and LP3 were not so evident, a comparison with connected uncertainty was conducted. As a result, it emerges how, for an assigned 1 in Y AEP, TCEV distribution presents systematically a considerable low uncertainty for the above defined macro-region.

In conclusion, TCEV seems to show the best fit performances for sites of NSW and Queensland below the latitude of 23° south in the eastern Australia. This region is surprisingly similar to that found by Micevski et al. (2006). This latter study was a consequence of a previous work of Franks and Kuczera (2002), that demonstrated how floods for NSW are sensible to long-term climate variations, modulated by El Nino Southern Oscillation, and explainable by values assumed by the Interdecadal Pacific Oscillation. As a consequence, authors stated that flood series cannot be considered homogeneous and composed by identically distributed floods.

In the spite of these works, the good fit of TCEV could be explained with its ability in interpreting the marginal distribution of this composite populations.

Finally, a single Italian case study was proposed. Also in this case, TCEV gave the best performances. This can be explained by the presence of floods that are generated by different runoff mechanisms. An interpretation of these phenomena in the framework of Theoretically derived distributions of floods was provided by Gioia et al. (2008)

7.5 – Future works

The development of a Bayesian approach for computation of parameters and quantile estimates, in the light of the findings about TCEV gave lots of suggestions for future works. The simplest idea that can be implemented is to study how at-site uncertainty can be reduced introducing into the analysis both regional and historical information. Furthermore, an application of TCEV in the regional frequency analysis framework can lead to relevant progresses in the management of flood risk to wide areas, with benefits detectable especially in ungauged basins.

Another purpose that should be explored is to extend analysis to New Zealand (were a good fit of TCEV was documented) and to Pacific areas for which LP3 distribution was considered giving an adequate fit to observed annual maxima.

The role of PILFs if TCEV distribution is also susceptible of investigations, searching for an analytical comparison with results derived by applying statistical tests as the multiple. Grubbs-Beck test.

Finally, insights can be found in problems arising with the estimation of Kappa distribution. A reliable solution could be the reduction of the feasible space in its parameter domain, in order to give to this distribution a more physical interpretation.

REFERENCES

Adams, C. A. and McMahon, T. A.: ESTIMATION OF FLOOD DISCHARGE FOR UNGAUGED RURAL CATCHMENTS IN VICTORIA., in National Conference Publication - Institution of Engineers, Australia., 1985.

Akaike, H.: A New Look at the Statistical Model Identification, IEEE Trans. Automat. Contr., doi:10.1109/TAC.1974.1100705, 1974.

Aldrich, J.: R. A. Fisher and the making of maximum likelihood 1912 - 1922, Stat. Sci., doi:10.1214/ss/1030037906, 1997.

Ames, W. F. and Brezinski, C.: Numerical recipes in Fortran 77, Volume 1., 1993.

Arnell, N. W. and Gabriele, S.: The performance of the two-component extreme value distribution in regional flood frequency analysis, Water Resour. Res., doi:10.1029/WR024i006p00879, 1988.

Arnell, N., & Beran, M. (1988). Probability-weighted moments estimators for TCEV parameters.

Asquith, W.H., 2018, lmomco---L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.3.2, Texas Tech University, Lubbock, Texas.

Bačová-Mitková, V. and Onderka, M.: Analysis of extreme hydrological Events on the Danube using the Peak Over Threshold method, J. Hydrol. Hydromechanics, doi:10.2478/v10098-010-0009-x, 2010.

Balkema, A. A. and de Haan, L.: Residual Life Time at Great Age, Ann. Probab., doi:10.1214/aop/1176996548, 1974.

Ball J, Babister M, Nathan R, Weeks W, Weinmann E, Retallick M, Testoni I, (Editors) (2019). Australian Rainfall and Runoff: A Guide to Flood Estimation, © Commonwealth of Australia (Geoscience Australia).

Beran, M., Hosking, J. R. M. and Arnell, N.: Comment on "Two-Component Extreme Value Distribution for Flood Frequency Analysis" by Fabio Rossi, Mauro Florentino, and Pasquale Versace, Water Resour. Res., doi:10.1029/WR022i002p00263, 1986.

Blum, A. G., Archfield, S. A. and Vogel, R. M.: On the probability distribution of daily streamflow in the United States, Hydrol. Earth Syst. Sci., doi:10.5194/hess-21-3093-2017, 2017.

Bobée, B. and Rasmussen, P. F.: Recent advances in flood frequency analysis, Rev. Geophys., doi:10.1029/95RG00287, 1995.

Bobée, B.: The Log Pearson type 3 distribution and its application in hydrology, Water Resour. Res., doi:10.1029/WR011i005p00681, 1975.

Bortkiewicz, L., von (1922). Variationsbreite und mittlerer Fehler, Sitzungsber. Berli. Math. Ges. 21, 3-11.

Buishand, T. A.: Statistics of extremes in climatology, Stat. Neerl., doi:10.1111/j.1467-9574.1989.tb01244.x, 1989.

Burnham, K. P. and Anderson, D. R.: Multimodel inference: Understanding AIC and BIC in model selection, Sociol. Methods Res., doi:10.1177/0049124104268644, 2004.

Castellarin, A, Kohnová, S, Gaal, L, Fleig, A, Salinas, JL, Toumazis, A, Kjeldsen, TR & MacDonald, N (2012), Review of applied statistical methods for flood frequency analysis in Europe: WG2 of COST Action ES0901. (NERC) Centre for Ecology & Hydrology.

Castellarin, A., Camorani, G. and Brath, A.: Predicting annual and long-term flow-duration curves in ungauged basins, Adv. Water Resour., doi:10.1016/j.advwatres.2006.08.006, 2007.

Castillo, E.: Asymptotic Distributions of Maxima and Minima (I.I.D. Case), in Extreme Value Theory in Engineering., 1988.

Chauvenet, W.: A manual of spherical and practical astronomy., 1863.

Cohn, T. A., England, J. F., Berenbrock, C. E., Mason, R. R., Stedinger, J. R. and Lamontagne, J. R.: A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series, Water Resour. Res., doi:10.1002/wrcr.20392, 2013.

Coles, S. (2001). *An introduction to statistical modeling of extreme values* (Vol. 208, p. 208). London: Springer.

Connell, R. J. and Pearson, C. P.: Two-component extreme value distribution applied to Canterbury annual maximum flood peaks, J. Hydrol. New Zeal., 2001.

Conway, K.M. (1970), *Flood frequency analysis of some NSW coastal rivers*, Thesis (M. Eng. Sc.), University of New South Wales, Australia.

Cunnane, C. (1989). Statistical distributions for flood frequency analysis. Operational hydrology report (WMO).

Cunnane, C.: Factors affecting choice of distribution for flood series, Hydrol. Sci. J., doi:10.1080/02626668509490969, 1985.

Cunnane, C.: Review of statistical models for flood frequency estimation., 1987.

Duan, Q., Sorooshian, S. and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., doi:10.1016/0022-1694(94)90057-4, 1994.

Dupuis, D. J. and Winchester, C.: More on the four-parameter kappa distribution, J. Stat. Comput. Simul., doi:10.1080/00949650108812137, 2001.

Erskine, W. D. and Warner, R. F.: Geomorphic effects of alternating flood and drought-dominated regimes on NSW coastal rivers, Fluv. Geomorphol. Aust., 1988.

Fisher, R. A. and Tippett, L. H. C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample, Math. Proc. Cambridge Philos. Soc., doi:10.1017/S0305004100015681, 1928.

Franks, S. W. and Kuczera, G.: Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales, Water Resour. Res., doi:10.1029/2001wr000232, 2002.

Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum, in Ann. Soc. Polon. Math., vol. 6, pp. 93-116.

Gabriele, S. and Arnell, N.: A hierarchical approach to regional flood frequency analysis, Water Resour. Res., doi:10.1029/91WR00238, 1991.

Gabriele, S., & Iiritano, G. (1994). Alcuni aspetti teorici ed applicativi nella regionalizzazione delle piogge con il modello TCEV. GNDCI–Linea, 1 (in Italian)

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B.: Bayesian Data Analysis Second Edition.PDF, Wiley Interdiscip. Rev. Cogn. Sci., doi:10.1002/wcs.72, 2004.

Gilleland, E. and Katz, R. W.: ExtRemes 2.0: An extreme value analysis package in R, J. Stat. Softw., doi:10.18637/jss.v072.i08, 2016.

Gioia, A., Iacobellis, V., Manfreda, S. and Fiorentino, M.: Runoff thresholds in derived flood frequency distributions, Hydrol. Earth Syst. Sci., doi:10.5194/hess-12-1295-2008, 2008.

Gnedenko, B.: Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire, Ann. Math., doi:10.2307/1968974, 1943.

Gradshteyn, L., Y., and Ryzhik, I. M. (2007). Table of Integrals, Series and Products., Math. Comput., doi:10.2307/2007757.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R.: Probability weighted moments: Definition and relation to parameters of several

distributions expressable in inverse form, Water Resour. Res., doi:10.1029/WR015i005p01049, 1979.

Griffis, V. W. and Stedinger, J. R.: Log-Pearson type 3 distribution and Its application in flood frequency analysis. I: Distribution characteristics, J. Hydrol. Eng., doi:10.1061/(ASCE)1084-0699(2007)12:5(482), 2007.

Griffis, V. W. and Stedinger, J. R.: Log-pearson type 3 distribution and its application in flood frequency analysis. II: Parameter estimation methods, J. Hydrol. Eng., doi:10.1061/(ASCE)1084-0699(2007)12:5(492), 2007.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press.

Gumbel, E. J.: The Return Period of Flood Flows, Ann. Math. Stat., doi:10.1214/aoms/1177731747, 1941.

Haddad K, Rahman A (2008) Investigation on at-site flood frequency analysis in south-east Australia. IEM J Inst Eng Malays 69(3):59–64

Haddad, K. and Rahman, A.: Selection of the best fit flood frequency distribution and parameter estimation procedure: A case study for Tasmania in Australia, Stoch. Environ. Res. Risk Assess., doi:10.1007/s00477-010-0412-1, 2011.

Hastings, W. K.: Monte carlo sampling methods using Markov chains and their applications, Biometrika, doi:10.1093/biomet/57.1.97, 1970.

Hosking, J. R. M. (2019). Regional Frequency Analysis using L-Moments. R package, version 3.2. URL: https://CRAN.R-project.org/package=lmomRFA.

Hosking, J. R. M. and Wallis, J. R.: Parameter and quantile estimation for the generalized pareto distribution, Technometrics, doi:10.1080/00401706.1987.10488243, 1987.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resour. Res., doi:10.1029/92WR01980, 1993.

Hosking, J. R. M., Wallis, J. R. and Wood, E. F.: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, Technometrics, doi:10.1080/00401706.1985.10488049, 1985.

Hosking, J. R. M., Wallis, J. R., Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis, in Regional Frequency Analysis., 1997.

Hosking, J. R. M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, J. R. Stat. Soc. Ser. B, doi:10.1111/j.2517-6161.1990.tb01775.x, 1990.

Hosking, J. R. M.: Research Report Fortran routines for use with the method of, Mathematics, 2005.

Hosking, J. R. M.: The theory of probability weighted moments, Res. Rep. RC12210, New York., 1986.

Houghton, J. C.: Birth of a parent: The Wakeby Distribution for modeling flood flows, Water Resour. Res., doi:10.1029/WR014i006p01105, 1978.

Institution of Engineers Australia (I.E. Aust.) (1987) Australian rainfall and runoff: a guide to flood estimation ARR (1987) In: Pilgrim HD (ed) The Institute of Engineers Australia, Canberra (Version 3)

Interagency Advisory Committee on Water Data (IACWD), 1982. Guidelines for determining flood flow Frequency. Bulletin 17B of the Hydrology Subcommittee, OWDC, US Geological Survey, Reston, VA.

Jenkinson, A. F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. R. Meteorol. Soc., doi:10.1002/qj.49708134804, 1955.

Katz, R. W., Brush, G. S. and Parlange, M. B.: Statistics of extremes: Modeling ecological disturbances, Ecology, doi:10.1890/04-0606, 2005.

Kendall, M. G.: Studies in the History of Probability and Statistics: XI. Daniel Bernoulli on Maximum Likelihood, Biometrika, doi:10.2307/2333125, 1961.

Kendall, M., & Stuart, A. (1977). The advanced theory of statistics. Vol. 1: Distribution theory. *London: Griffin, 1977, 4th ed.*

Kotz, S. and Nadarajah, S.: Extreme Value Distributions: Theory and Applications, J. Zhejiang Univ. Sci., doi:10.1142/p191, 2000.

Koutsoyiannis, D. (2003, October). On the appropriateness of the Gumbel distribution for modelling extreme rainfall. In Proceedings of the ESF LESC Exploratory Workshop held at Bologna (pp. 24-25).

Koutsoyiannis, D., Mamassis, N., Efstratiadis, A., Zarkadoulas, N., & Markonis, I. (2012). floods in Greece. Changes of flood risk in Europe, 238-256.

Koutsoyiannis, D., Zarkadoulas, N., Angelakis, A. N. and Tchobanoglous, G.: Urban water management in ancient Greece: Legacies and lessons, J. Water Resour. Plan. Manag., doi:10.1061/(ASCE)0733-9496(2008)134:1(45), 2008.

Kuczera, G.: Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference, Water Resour. Res., doi:10.1029/1999WR900012, 1999.

Laio, F., Di Baldassarre, G. and Montanari, A.: Model selection techniques for the frequency analysis of hydrological extremes, Water Resour. Res., doi:10.1029/2007WR006666, 2009.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R.: Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles, Water Resour. Res., doi:10.1029/WR015i005p01055, 1979.

Lang, M., Ouarda, T. B. M. J. and Bobée, B.: Towards operational guidelines for over-threshold modeling, J. Hydrol., doi:10.1016/S0022-1694(99)00167-5, 1999.

Loeve, M. (1977). Notions of Measure Theory. In *Probability Theory I* (pp. 53-147). Springer, New York, NY

Madsen, H., Rasmussen, P. and Rosbjerg, D.: Comparison of annual maximum series and partial duration series for modelling exteme hydrological events: 1. At sit modelling, Water Res. Res., 1997.

Matalas, N. C. and Wallis, J. R.: Eureka! It fits a Pearson type: 3 distribution, Water Resour. Res., doi:10.1029/WR009i002p00281, 1973.

Matalas, N. C., Slack, J. R. and Wallis, J. R.: Regional skew in search of a parent, Water Resour. Res., doi:10.1029/WR011i006p00815, 1975.

Mays, L. W., Koutsoyiannis, D. and Angelakis, A. N.: A brief history of urban water supply in antiquity, Water Sci. Technol. Water Supply, doi:10.2166/ws.2007.001, 2007.

McCabe, G. J. and Wolock, D. M.: Climate change and the detection of trends in annual runoff, Clim. Res., doi:10.3354/cr008129, 1997.

McMahon, T. A. and Srikanthan, R.: Log Pearson III distribution - Is it applicable to flood frequency analysis of Australian streams?, J. Hydrol., doi:10.1016/0022-1694(81)90100-1, 1981.

Merz, B. and Thieken, A. H.: Separating natural and epistemic uncertainty in flood frequency analysis, J. Hydrol., doi:10.1016/j.jhydrol.2004.11.015, 2005.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E.: Equation of state calculations by fast computing machines, J. Chem. Phys., doi:10.1063/1.1699114, 1953.

Micevski, T., Franks, S. W. and Kuczera, G.: Multidecadal variability in coastal eastern Australian flood data, J. Hydrol., doi:10.1016/j.jhydrol.2005.11.017, 2006.

Mielke, P. W.: Another Family of Distributions for Describing and Analyzing Precipitation Data, J. Appl. Meteorol., doi:10.1175/1520-0450(1973)012<0275:afodfd>2.0.co;2, 1973.

Montanari, A., Shoemaker, C. A. and Van De Giesen, N.: Introduction to special section on uncertainty assessment in surface and subsurface hydrology: An overview of issues and challenges, Water Resour. Res., doi:10.1029/2009WR008471, 2009.

Morgan, E. C., Lackner, M., Vogel, R. M. and Baise, L. G.: Probability distributions for offshore wind speeds, in Energy Conversion and Management., 2011.

Murshed, M. S., Seo, Y. A. and Park, J. S.: LH-moment estimation of a four parameter kappa distribution with hydrologic applications, Stoch. Environ. Res. Risk Assess., doi:10.1007/s00477-013-0746-6, 2014.

Nathan RJ and Weinmann PE (1991) Application of at-site and regional flood frequency analyses. In: Proceedings International Hydrology Water Resources Symposium, Perth, 2–4 October, 769:774

National Research Council (2000), Risk Analysis and Uncertainty in Flood Damage Reduction Studies, Natl. Acad., Washington, D. C.

Papalexiou, S. M. and Koutsoyiannis, D.: Battle of extreme value distributions : A global survey on extreme daily rainfall, Water Resour. Res., doi:10.1029/2012WR012557, 2013.

Pappenberger, F. and Beven, K. J.: Ignorance is bliss: Or seven reasons not to use uncertainty analysis, Water Resour. Res., doi:10.1029/2005WR004820, 2006.

Parida, B. P.: Modelling of Indian summer monsoon rainfall using a four-parameter Kappa distribution, Int. J. Climatol., doi:10.1002/(SICI)1097-0088(199910)19:12<1389::AID-JOC435>3.0.CO;2-T, 1999.

Park, J. S. and Park, B. J.: Maximum likelihood estimation of the four-parameter Kappa distribution using the penalty method, Comput. Geosci., doi:10.1016/S0098-3004(01)00069-3, 2002.

Park, J. S. and Yoon Kim, T.: Fisher information matrix for a four-parameter kappa distribution, Stat. Probab. Lett., doi:10.1016/j.spl.2007.03.002, 2007.

Parkes, B. and Demeritt, D.: Defining the hundred year flood: A Bayesian approach for using historic data to reduce uncertainty in flood frequency estimates, J. Hydrol., doi:10.1016/j.jhydrol.2016.07.025, 2016.

Paul, H., Rahman, A. and Haque, M.: Application of arr flike for at-site flood frequency analysis: A case study in New South Wales, Australia, in 37th Hydrology and Water Resources Symposium 2016: Water, Infrastructure and the Environment, HWRS 2016., 2016.

Peirce, B.: Criterion for the rejection of doubtful observations, Astron. J., doi:10.1086/100273, 1852.

Penta, A., Rossi, F., Silvagni, G., Veltri, M., & Versace, P. (1980). Un modello stocastico per l'analisi delle massime piogge giornaliere in presenza di grandi nubifragi (A stochastic model for analysis of daily rainfall maxima in presence of disastrous storms). Atti del XVII Convegno di Idraulica e Costruzioni Idrauliche (in italian)

Pickands, J. I.: Statistical Inference Using Extreme Order Statistics, Ann. Stat., doi:10.1214/aos/1176343003, 1975.

Rahman, A. S., Haddad, K., & Rahman, A. (2014). Impacts of outliers in flood frequency analysis: A case study for Eastern Australia. *J. Hydrol. Environ. Res*, *2*(1), 17-13.

Rahman, A. S., Rahman, A., Zaman, M. A., Haddad, K., Ahsan, A. and Imteaz, M.: A study on selection of probability distributions for at-site flood frequency analysis in Australia, Nat. Hazards, doi:10.1007/s11069-013-0775-y, 2013.

Ramachandra Rao, A., Hamed, K. H., Ramachandra Rao, A. and Hamed, K. H.: Parameter and Quantile Estimation, in Flood Frequency Analysis., 2000.

Rasmussen, P. F., & Rosbjerg, D. (1991). Prediction uncertainty in seasonal partial duration series. Water resources research, 27(11), 2875-2883.

Rasmussen, P. F.: Generalized probability weighted moments: Application to the generalized Pareto distribution, Water Resour. Res., doi:10.1029/2001WR900014, 2001.

Reis, D. S. and Stedinger, J. R.: Bayesian MCMC flood frequency analysis with historical information, in Journal of Hydrology., 2005.

Rider, P. R. (1933). Criteria for rejection of observations. *Washington University Studies (New Series), Science and Technology, 8*

Rossi, F., Fiorentino, M. and Versace, P.: Two-Component Extreme Value Distribution for Flood Frequency Analysis, Water Resour. Res., doi:10.1029/WR020i007p00847, 1984.

Rossi, F., Versace, P. (1981): Criteri e metodi per l'analisi statistica delle piene. Pubbl.del P.F.Conservazione del Suolo "Analisi delle piene" (in italian)

Schwarz, G.: Estimating the Dimension of a Model, Ann. Stat., doi:10.1214/aos/1176344136, 1978.

Shane, R. M., & Lynn, W. R. (1964). Mathematical model for flood risk evaluation. *Journal of the Hydraulics Division*, *90*(6), 1-20.

Singh, K.P. and Sinclair, R.A.: Two- distribution method for flood- frequency analysis, ASCE J Hydraul Div, 1972.

Singh, V. P. and Deng, Z. Q.: Entropy-based parameter estimation for Kappa distribution, J. Hydrol. Eng., doi:10.1061/(ASCE)1084-0699(2003)8:2(81), 2003.

Singh, V. P. and Strupczewski, W. G.: On the status of flood frequency analysis, Hydrol. Process., doi:10.1002/hyp.5083, 2002.

Snyder, D. L., & Miller, M. I. (2012). Random point processes in time and space. Springer Science & Business Media

Stedinger, J. R., Vogel, R. M. and Foufoula-Georgiou, E.: Frequency Analysis of Extreme Events, Handb. Hydrol., 1993.

Sugiura, N.: Further Analysis of the Data by Akaike' S Information Criterion and the Finite Corrections, Commun. Stat. - Theory Methods, doi:10.1080/03610927808827599, 1978.

Tingsanchali, T.: Urban flood disaster management, in Procedia Engineering., 2012.

Todhunter, I.: A history of the mathematical theory of probability from the time of Pascal to that of Laplace / by I. Todhunter., 1865.

Todorovic, P. and Zelenhasic, E.: A Stochastic Model for Flood Analysis, Water Resour. Res., doi:10.1029/WR006i006p01641, 1970.

Todorovic, P., & Yevjevic, V. (1969). Stochastic process of precipitation. *Hydrology papers (Colorado State University); no. 35*.

Todorovic, P.: Stochastic models of floods, Water Resour. Res., doi:10.1029/WR014i002p00345, 1978.

Totaro, V., Gioia, A. and Iacobellis, V.: Power of parametric and non-parametric tests for trend detection in annual maximum series, Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2019-363, 2019.

Van Gelder, P. H. A. J. M.: Statistical estimation methods in hydrological engineering, in Analysis and Stochastic Modeling of Extreme Runoff in Eurasian Rivers Under Conditions of Climate Change, edited by W. M. L. L.M. Korytny, pp. 11–57, Irkutsk., 2004.

Velickov, S.: Nonlinear Dynamics and Chaos with Applications to Hydrodynamics and Hydrological Modelling., 2014.

Vicens, G. J., Rodriguez-Iturbe, I. and Schaake, J. C.: A Bayesian framework for the use of regional information in hydrology, Water Resour. Res., doi:10.1029/WR011i003p00405, 1975.

Vogel, R. M., McMahon, T. A. & Chiew, F. H. S. (1993a): Floodflow frequency model selection in Australia, J. Hydrol., doi:10.1016/0022-1694(93)90288-K.

Vogel, R. M., Thomas Jr, W. O., & McMahon, T. A. (1993b). Flood-flow frequency model selection in southwestern United States. Journal of Water Resources Planning and Management, 119(3), 353-366.

von Mises, R.: La distribution de la plus grande de n valeurs, Am. Math. Soc., 1936.

Wallis, J. R., Matalas, N. C. and Slack, J. R.: Just a moment!, Water Resour. Res., doi:10.1029/WR010i002p00211, 1974.

Winchester, C. (2000). On estimation of the four-parameter kappa distribution. Dalhousie University.

Wood, E. F. and Rodríguez-Iturbe, I.: A Bayesian approach to analyzing uncertainty among flood frequency models, Water Resour. Res., doi:10.1029/WR011i006p00839, 1975.

Wood, E. F. and Rodríguez-Iturbe, I.: Bayesian inference and decision making for extreme hydrologic events, Water Resour. Res., doi:10.1029/WR011i004p00533, 1975.

Zelenhasic, E.: Theoretical probability distributions for flood peaks, Colo State Univ (Fort Collins), Hydrol Pap 42, 1970.

APPENDIX A: L-Moments for Kappa distribution

Case 1: $\varepsilon \neq 0$

Case 1.a: $h > 0, \varepsilon > -1$

$$\begin{cases} 1 \cdot \beta_0 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{\Gamma(1+\varepsilon)\Gamma\left(\frac{1}{h}\right)}{h^{(1+\varepsilon)}\Gamma\left(1+\varepsilon+\frac{1}{h}\right)}\right] \\ 2 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{2\Gamma(1+\varepsilon)\Gamma\left(\frac{2}{h}\right)}{h^{(1+\varepsilon)}\Gamma\left(1+\varepsilon+\frac{2}{h}\right)}\right] \\ 3 \cdot \beta_2 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{3\Gamma(1+\varepsilon)\Gamma\left(\frac{3}{h}\right)}{h^{(1+\varepsilon)}\Gamma\left(1+\varepsilon+\frac{3}{h}\right)}\right] \\ 4 \cdot \beta_3 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{4\Gamma(1+\varepsilon)\Gamma\left(\frac{4}{h}\right)}{h^{(1+\varepsilon)}\Gamma\left(1+\varepsilon+\frac{4}{h}\right)}\right] \end{cases}$$

System (4.10), because of (3.22-3.25), become:

$$
\begin{cases}
\zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{\Gamma(1+\varepsilon)\Gamma\left(\frac{1}{h}\right)}{h^{(1+\varepsilon)}\Gamma\left(1+\varepsilon+\frac{1}{h}\right)}\right] = \ell_1 \\
\dfrac{\sigma}{\varepsilon}\dfrac{\Gamma(1+\varepsilon)}{h^{(1+\varepsilon)}}\left[\dfrac{\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{1}{h}\right)} - 2\dfrac{\Gamma\left(\frac{2}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{2}{h}\right)}\right] = \ell_2 \\
\dfrac{\left[-6\dfrac{\Gamma\left(\frac{3}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{3}{h}\right)} + 6\dfrac{\Gamma\left(\frac{2}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{2}{h}\right)} - \dfrac{\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{1}{h}\right)}\right]}{\left[\dfrac{\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{1}{h}\right)} - \dfrac{2\Gamma\left(\frac{2}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{2}{h}\right)}\right]} = t_3 \\
\dfrac{\left[-20\dfrac{\Gamma\left(\frac{4}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{4}{h}\right)} + 30\dfrac{\Gamma\left(\frac{3}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{3}{h}\right)} - 12\dfrac{\Gamma\left(\frac{2}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{2}{h}\right)} + \dfrac{\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{1}{h}\right)}\right]}{\left[\dfrac{\Gamma\left(\frac{1}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{1}{h}\right)} - \dfrac{2\Gamma\left(\frac{2}{h}\right)}{\Gamma\left(1+\varepsilon+\frac{2}{h}\right)}\right]} = t_4
\end{cases}
$$

Case 1.b: $h = 0, \varepsilon > -1$

$$
\begin{cases}
1 \cdot \beta_0 = \zeta + \dfrac{\sigma}{\varepsilon}[1 - \Gamma(1+\varepsilon)] \\
2 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}[1 - 2^{-\varepsilon}\Gamma(1+\varepsilon)] \\
3 \cdot \beta_2 = \zeta + \dfrac{\sigma}{\varepsilon}[1 - 3^{-\varepsilon}\Gamma(1+\varepsilon)] \\
4 \cdot \beta_3 = \zeta + \dfrac{\sigma}{\varepsilon}[1 - 4^{-\varepsilon}\Gamma(1+\varepsilon)]
\end{cases}
$$

System (4.10), because of (3.22-3.25), become:

$$\begin{cases} \zeta + \dfrac{\sigma}{\varepsilon}[1 - \Gamma(1+\varepsilon)] = \ell_1 \\ \dfrac{\sigma}{\varepsilon}\Gamma(1+\varepsilon)\Gamma(1-2^{-\varepsilon}) = \ell_2 \\ \dfrac{-2\cdot 3^{-\varepsilon} + 3\cdot 2^{-\varepsilon} - 1}{1-2^{-\varepsilon}} = t_3 \\ \dfrac{-5\cdot 4^{-\varepsilon} + 10\cdot 3^{-\varepsilon} - 2^{-\varepsilon} + 1}{1-2^{-\varepsilon}} = t_4 \end{cases}$$

Case 1.c: $h < 0, -1 < \varepsilon > -1/\varepsilon$

$$\begin{cases} 1\cdot\beta_0 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{\Gamma(1+\varepsilon)\Gamma\left(-\varepsilon - \dfrac{1}{h}\right)}{(-h)^{(1+\varepsilon)}\Gamma\left(1 - \dfrac{1}{h}\right)}\right] \\ 2\cdot\beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{2\cdot\Gamma(1+\varepsilon)\Gamma\left(-\varepsilon - \dfrac{2}{h}\right)}{(-h)^{(1+\varepsilon)}\Gamma\left(1+\varepsilon + \dfrac{2}{h}\right)}\right] \\ 3\cdot\beta_2 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{3\Gamma(1+\varepsilon)\Gamma\left(-\varepsilon - \dfrac{3}{h}\right)}{(-h)^{(1+\varepsilon)}\Gamma\left(1 - \dfrac{3}{h}\right)}\right] \\ 4\cdot\beta_3 = \zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{4\Gamma(1+\varepsilon)\Gamma\left(-\varepsilon - \dfrac{4}{h}\right)}{(-h)^{(1+\varepsilon)}\Gamma\left(1 - \dfrac{4}{h}\right)}\right] \end{cases}$$

System (4.10), because of (3.22-3.25), become:

$$
\begin{cases}
\zeta + \dfrac{\sigma}{\varepsilon}\left[1 - \dfrac{\Gamma(1+\varepsilon)\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{(-h)^{(1+\varepsilon)}\Gamma\left(1-\frac{1}{h}\right)}\right] = \ell_1 \\
\dfrac{\sigma}{\varepsilon}\dfrac{\Gamma(1+\varepsilon)}{(-h)^{(1+\varepsilon)}}\left[2\dfrac{\Gamma\left(-\varepsilon-\frac{2}{h}\right)}{\Gamma\left(1-\frac{2}{h}\right)} - \dfrac{\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}\right] = \ell_2 \\
\dfrac{\left[-6\dfrac{\Gamma\left(-\varepsilon-\frac{3}{h}\right)}{\Gamma\left(1-\frac{3}{h}\right)} + 6\dfrac{\Gamma\left(-\varepsilon-\frac{2}{h}\right)}{\Gamma\left(1-\frac{2}{h}\right)} - \dfrac{\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}\right]}{\left[2\dfrac{\Gamma\left(-\varepsilon-\frac{2}{h}\right)}{\Gamma\left(1-\frac{2}{h}\right)} - \dfrac{\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}\right]} = t_3 \\
\dfrac{\left[-20\dfrac{\Gamma\left(-\varepsilon-\frac{4}{h}\right)}{\Gamma\left(1-\frac{4}{h}\right)} + 30\dfrac{\Gamma\left(-\varepsilon-\frac{3}{h}\right)}{\Gamma\left(1-\frac{3}{h}\right)} - 12\dfrac{\Gamma\left(-\varepsilon-\frac{2}{h}\right)}{\Gamma\left(1-\frac{2}{h}\right)} + \dfrac{\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}\right]}{\left[2\dfrac{\Gamma\left(-\varepsilon-\frac{2}{h}\right)}{\Gamma\left(1-\frac{2}{h}\right)} - \dfrac{\Gamma\left(-\varepsilon-\frac{1}{h}\right)}{\Gamma\left(1-\frac{1}{h}\right)}\right]} = t_4
\end{cases}
$$

Case 2: $\varepsilon = 0$

Case 2.a: $h > 0$

$$
\begin{cases}
1 \cdot \beta_0 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(h) + \psi\left(1+\frac{1}{h}\right)\right] \\
2 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(h) + \psi\left(1+\frac{2}{h}\right)\right] \\
3 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(h) + \psi\left(1+\frac{3}{h}\right)\right] \\
4 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(h) + \psi\left(1+\frac{4}{h}\right)\right]
\end{cases}
$$

System (4.10), because of (3.22-3.25), become:

$$\begin{cases} \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(h) + \psi\left(1+\dfrac{1}{h}\right)\right] = \ell_1 \\ \alpha\left[\psi\left(1+\dfrac{2}{h}\right) - \psi\left(1+\dfrac{1}{h}\right)\right] = \ell_2 \\ \dfrac{2\psi\left(1+\dfrac{3}{h}\right) - 3\psi\left(1+\dfrac{2}{h}\right) + \psi\left(1+\dfrac{1}{h}\right)}{\psi\left(1+\dfrac{2}{h}\right) - \psi\left(1+\dfrac{1}{h}\right)} = t_3 \\ \dfrac{5\psi\left(1+\dfrac{4}{h}\right) - 10\psi\left(1+\dfrac{3}{h}\right) + 6\psi\left(1+\dfrac{2}{h}\right) - \psi\left(1+\dfrac{1}{h}\right)}{\psi\left(1+\dfrac{2}{h}\right) - \psi\left(1+\dfrac{1}{h}\right)} = t_4 \end{cases}$$

Case 2.b: $h = 0$

$$\begin{cases} 1 \cdot \beta_0 = \zeta + \alpha\gamma \\ 2 \cdot \beta_1 = \zeta + \alpha[\gamma + \ln(2)] \\ 3 \cdot \beta_1 = \zeta + \alpha[\gamma + \ln(3)] \\ 4 \cdot \beta_1 = \zeta + \alpha[\gamma + \ln(4)] \end{cases}$$

System (4.10), because of (3.22-3.25), become:

$$\begin{cases} \zeta + \alpha\gamma = \ell_1 \\ \alpha \ln(3) = \ell_2 \\ \dfrac{\ln\left(\dfrac{3^2}{2^3}\right)}{\ln 2} = t_3 \\ \dfrac{\ln\left(\dfrac{2^{16}}{3^{10}}\right)}{\ln 2} = t_4 \end{cases}$$

Case 2.c: $h < 0$

$$\begin{cases} 1 \cdot \beta_0 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(-h) + \psi\left(-\dfrac{1}{h}\right)\right] \\ 2 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(-h) + \psi\left(-\dfrac{2}{h}\right)\right] \\ 3 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(-h) + \psi\left(-\dfrac{3}{h}\right)\right] \\ 4 \cdot \beta_1 = \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(-h) + \psi\left(-\dfrac{4}{h}\right)\right] \end{cases}$$

System (4.10), because of (3.22-3.25), become:

$$\begin{cases} \zeta + \dfrac{\sigma}{\varepsilon}\left[\gamma + \ln(-h) + \psi\left(-\dfrac{1}{h}\right)\right] = \ell_1 \\ \alpha\left[\psi\left(-\dfrac{2}{h}\right) - \psi\left(-\dfrac{1}{h}\right)\right] = \ell_2 \\ \dfrac{2\psi\left(-\dfrac{3}{h}\right) - 3\psi\left(-\dfrac{2}{h}\right) + \psi\left(-\dfrac{1}{h}\right)}{\psi\left(-\dfrac{2}{h}\right) - \psi\left(-\dfrac{1}{h}\right)} = t_3 \\ \dfrac{5\psi\left(-\dfrac{4}{h}\right) - 10\psi\left(-\dfrac{3}{h}\right) + 6\psi\left(-\dfrac{2}{h}\right) - \psi\left(-\dfrac{1}{h}\right)}{\psi\left(-\dfrac{2}{h}\right) - \psi\left(-\dfrac{1}{h}\right)} = t_4 \end{cases}$$

APPENDIX B: Dataset

B.1 – New South Wales

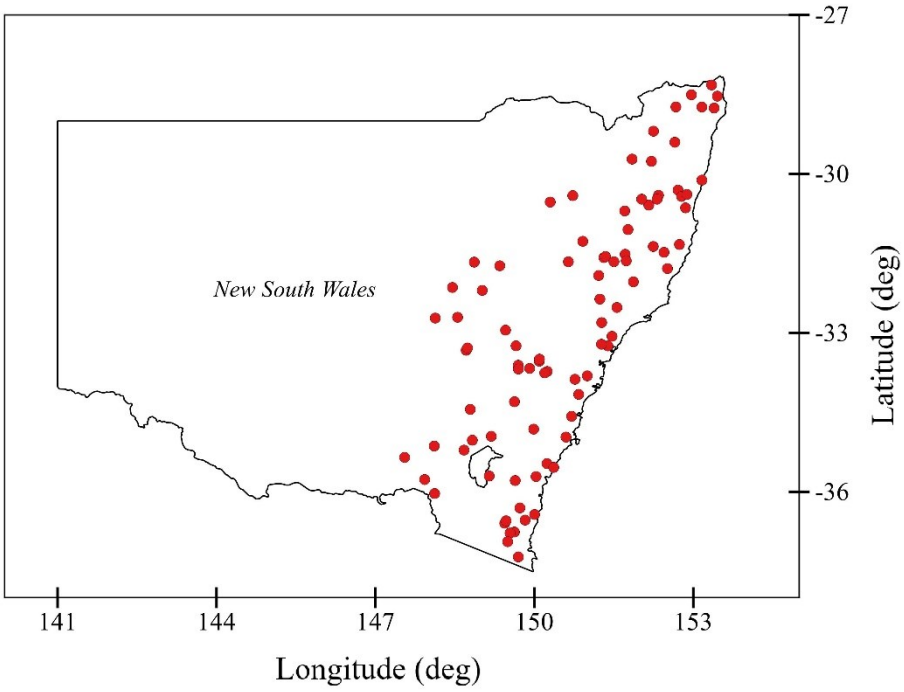

figure B.1 – Available sites for New South Wales

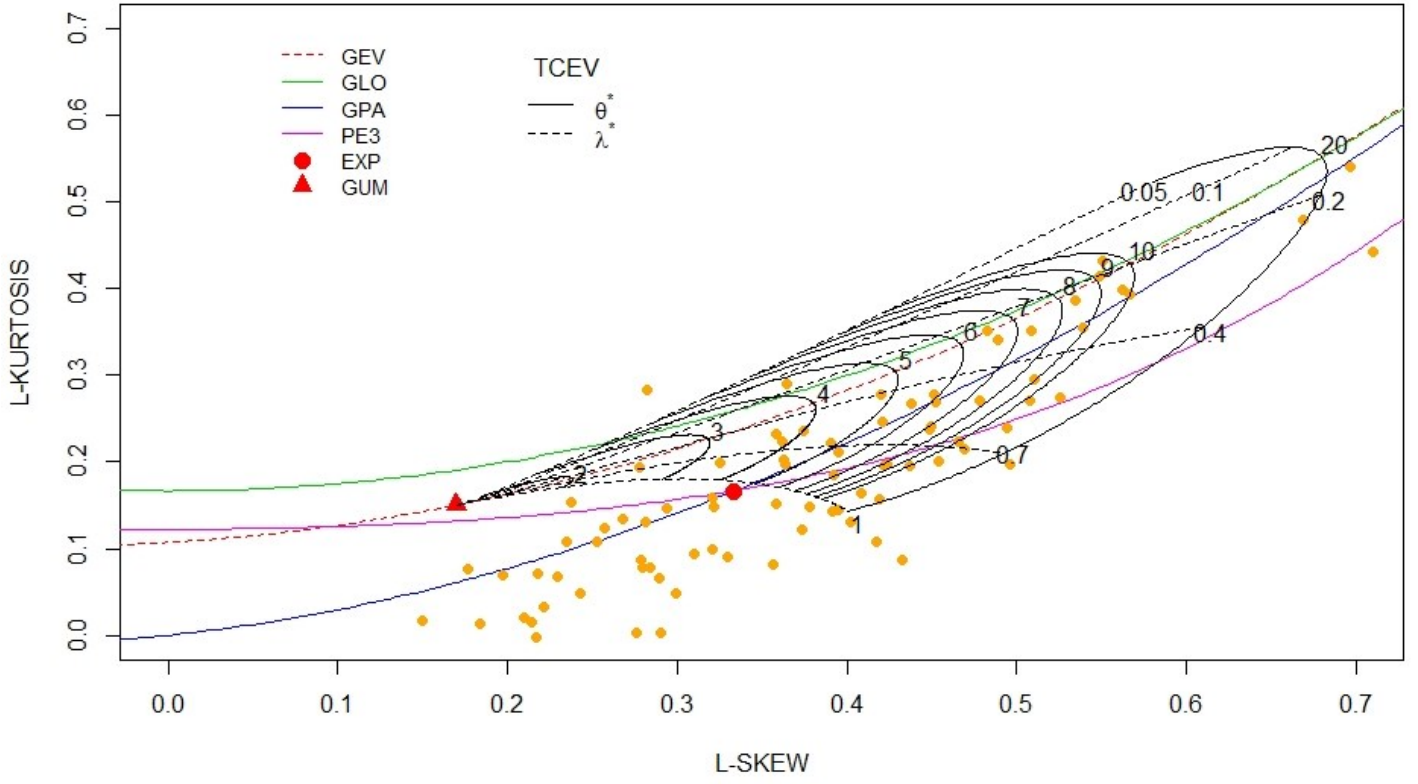

figure B. 2 – L-Moments Ratio Diagram for available catchments of NSW

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
201005	Boat Harbour No.20.55 cm	Rous	-28.3217	153.3467	1958-1985
202001	Durrumbul (Sherrys Crossing)	Brunswick	-28.5333	153.4567	1972-2011
203005	Wiangaree	Richmond	-28.5067	152.9667	1982-2011
203010	Rock Valley	Leycester	-28.7383	153.1633	1986-2011
203014	Eltham	Wilsons	-28.7583	153.3950	1987-2011
204008	Ebor	Guy Fawkes	-30.4050	152.3450	1983-2011
204017	Dorrigo No.2 & No.3	Bielsdown Ck	-30.3067	152.7133	1972-2011
204031	Shannon Vale	Mann	-29.7217	151.8450	1992-2011
204033	Billyrimba	Timbarra	-29.1950	152.2500	1979-2011
204034	Newton Boyd	Henry	-29.7633	152.2117	1972-2011
204043	Bonalbo	Peacock Ck	-28.7367	152.6733	1961-2011
204067	Fine Flower	Gordon Brook	-29.4033	152.6533	1983-2011
205002	Thora	Bellinger	-30.4267	152.7800	1983-2011
205006	Bowraville	Nambucca	-30.6417	152.8550	1972-2006
205007	Woolgoolga	Woolgoolga Ck	-30.1183	153.1633	1961-1982
205014	Gleniffer Br	Never Never	-30.3883	152.8800	1983-2006
206001	Jeogla	Styx	-30.5900	152.1617	1979-2011
206014	Coninside	Wollomombi	-30.4783	152.0267	1955-2011
206017	Causeway (Hatchery)	Serpentine Ck	-30.4783	152.3183	1962-1985
206018	Apsley Falls	Apsley	-31.0517	151.7683	1961-2011
206034	Abermala	Mihi Ck	-30.7000	151.7067	1985-2010
207013	D/S Bunnoo R Junction	Ellenborough	-31.4817	152.4483	1976-2011
207014	Avenel	Wilson	-31.3333	152.7400	1985-2011

207015	**Mount Seaview**	**Hastings**	**-31.3717**	**152.2450**	**1985-2011**
208006	Forbesdale (Causeway)	Barrington	-32.0383	151.8700	1973-2011
208007	Nowendoc	Nowendoc	-31.5183	151.7150	1974-2011
208009	Barry	Barnard	-31.5817	151.3133	1986-2011
208015	Landsdowne	Landsdowne	-31.7883	152.5133	1986-2011
208024	D/S Back R Jctn	Barnard	-31.5600	151.3433	1983-2011
208026	Jacky Barkers	Myall	-31.6417	151.7350	1985-2011
208027	Measuring Weir	Barnard	-31.6583	151.5033	1988-2011
210018	Moonam Dam Site	Hunter	-31.9183	151.2150	1974-2011
210069	Pokolbin Site 4	Muggyrang Ck	-32.8083	151.2717	1965-1992
210084	The Rocks No.2	Glennies Ck	-32.3650	151.2383	1973-2010
210095	Vacy	Bucks Ck	-32.5233	151.5600	1976-1997
211008	Avondale	Jigadee Ck	-33.0667	151.4667	1975-2011
211010	U/S Wyong R (Durren La)	Jilliby Ck	-33.2483	151.3900	1985-2011
211014	Yarramalong	Wyong	-33.2167	151.2667	1977-2011
212011	Lithgow	Coxs	-33.5367	150.0933	1962-2011
212013	Narrow Neck	Megalong Ck	-33.7300	150.2433	1988-2010
212042	Mount Walker	Farmers Ck	-33.4983	150.0967	1985-2011
212045	Island Hill	Coxs	-33.7583	150.1967	1983-2011
212320	Mulgoa Rd	South Ck	-33.8783	150.7683	1972-2011
213004	Parramatta Hospital	Parramatta	-33.8133	151.0000	1984-2003
213200	Wedderburn	O'Hares Ck	-34.1633	150.8383	1979-2011
214003	Albion Park	Macquarie Rivule	-34.5767	150.7050	1979-2011
215008	Kadoona	Shoalhaven	-35.7900	149.6400	1972-2010
215014	Bungonia	Bungonia Ck	-34.8200	149.9883	1984-2011

216002	**Brooman**	**Clyde**	**-35.4700**	**150.2383**	**1961-2011**
216004	Falls Ck	Currambene Ck	-34.9700	150.5983	1971-2010
216008	Kioloa	Butlers Ck	-35.5417	150.3667	1986-2010
216009	Buckenbowra No.3	Buckenbowra	-35.7150	150.0333	1986-2011
218003	Yowrie	Yowrie	-36.3067	149.7283	1959-1984
219001	Brown Mountain	Rutherford Ck	-36.5967	149.4417	1949-2010
219004	Tantawangalo School	Tantawangalo Ck	-36.7617	149.6233	1944-1973
219006	Tantawangalo Mountain (Dam)	Tantawangalo Ck	-36.7817	149.5417	1952-2010
219010	Brown Mountain (U/S Divers	Bonar Ck	-36.5500	149.4667	1955-1974
219013	North Brogo	Brogo	-36.5367	149.8267	1962-1982
219015	near Bermagui	Nutleys Ck	-36.4317	150.0050	1966-1988
220002	Rocky Hall (Whitbys)	Stockyard Ck	-36.9450	149.4967	1961-1984
221010	Imlay Rd Br	Imlay Ck	-37.2317	149.6983	1982-2011
401016	The Square	Welumba Ck	-36.0350	148.1183	1984-2011
401017	Yarramundi	Mannus Ck	-35.7717	147.9300	1984-2011
410107	Mountain Ck	Mountain Ck	-35.0283	148.8300	1980-2011
410141	Michelago	Micaligo Ck	-35.7050	149.1483	1983-2011
410149	Nottingham Rd Br	Nottingham Ck	-35.2150	148.6733	1983-2011
410152	Edwardstown	Stony Ck	-35.1383	148.1100	1985-2009
410156	Book Book	Kyeamba Ck	-35.3533	147.5517	1986-2011
410160	White Hill	Williams Ck	-34.9550	149.1883	1990-2010
412076	Cudal	Bourimbla Ck	-33.3300	148.7133	1980-1999
412090	Cudal No.2	Boree Ck	-33.2867	148.7383	1970-1989
412096	Kennys Ck Rd	Pudmans Ck	-34.4467	148.7917	1976-2002
412110	U/S Giddigang Ck	Bolong	-34.3017	149.6250	1981-2001

419035	**Timbumburi**	**Goonoo Goonoo** Ck	**-31.2733**	**150.9150**	**1982-2011**
419044	Damsite	Maules Ck	-30.5333	150.3000	1969-1992
419047	Woodsreef	Ironbark Ck	-30.4100	150.7267	1989-2011
419076	Old Warrah	Warrah Ck	-31.6600	150.6433	1983-2011
420010	Bearbung	Wallumburrawang Ck	-31.6667	148.8667	1980-2001
420012	Neilrex	Butheroo Ck	-31.7350	149.3483	1980-2001
421034	Dam Site	Slippery Ck	-33.6733	149.9117	1980-2000
421048	Obley No.2	Little	-32.7083	148.5517	1987-2011
421055	Rawsonville	Coolbaggie Ck	-32.1450	148.4550	1981-2011
421066	Hill end	Green Valley Ck	-32.9500	149.4567	1977-1998
421068	Saxa Crossing	Spicers Ck	-32.2000	149.0167	1978-2002
421076	Peak Hill No.2	Bogan	-32.7233	148.1300	1981-2011
421101	U/S Ben Chifley Dam	Campbells	-33.6133	149.6967	1979-2002
421104	Stromlo	Brisbane Valley	-33.6850	149.7000	1980-2000
421106	Wiagdon	Cheshire Ck	-33.2467	149.6550	1981-2001

B.2 – Queensland

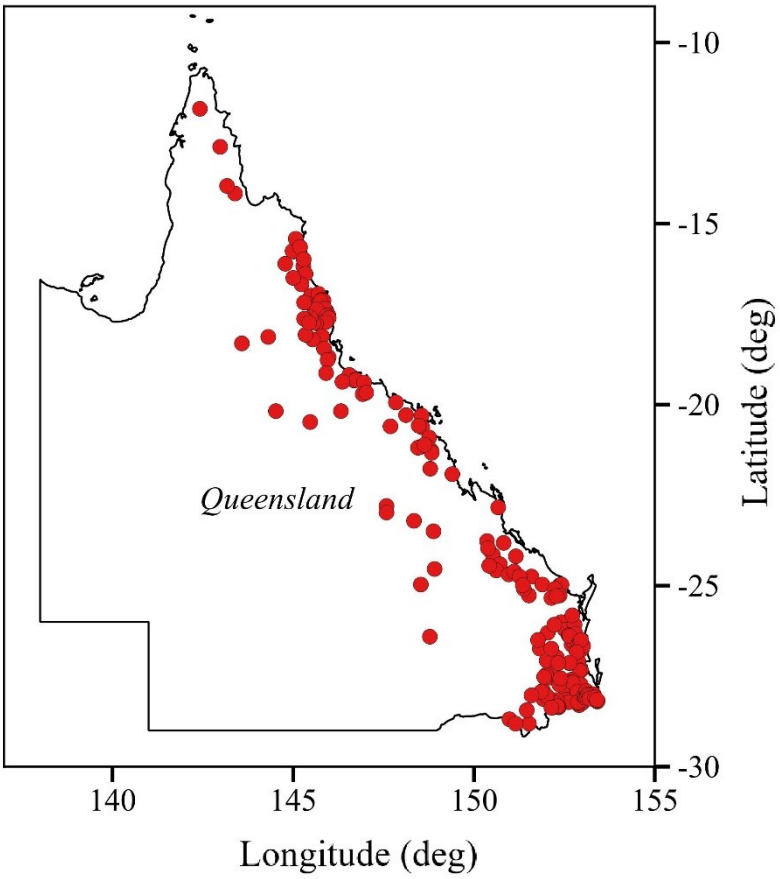

figure B.3 – Available sites for Queensland

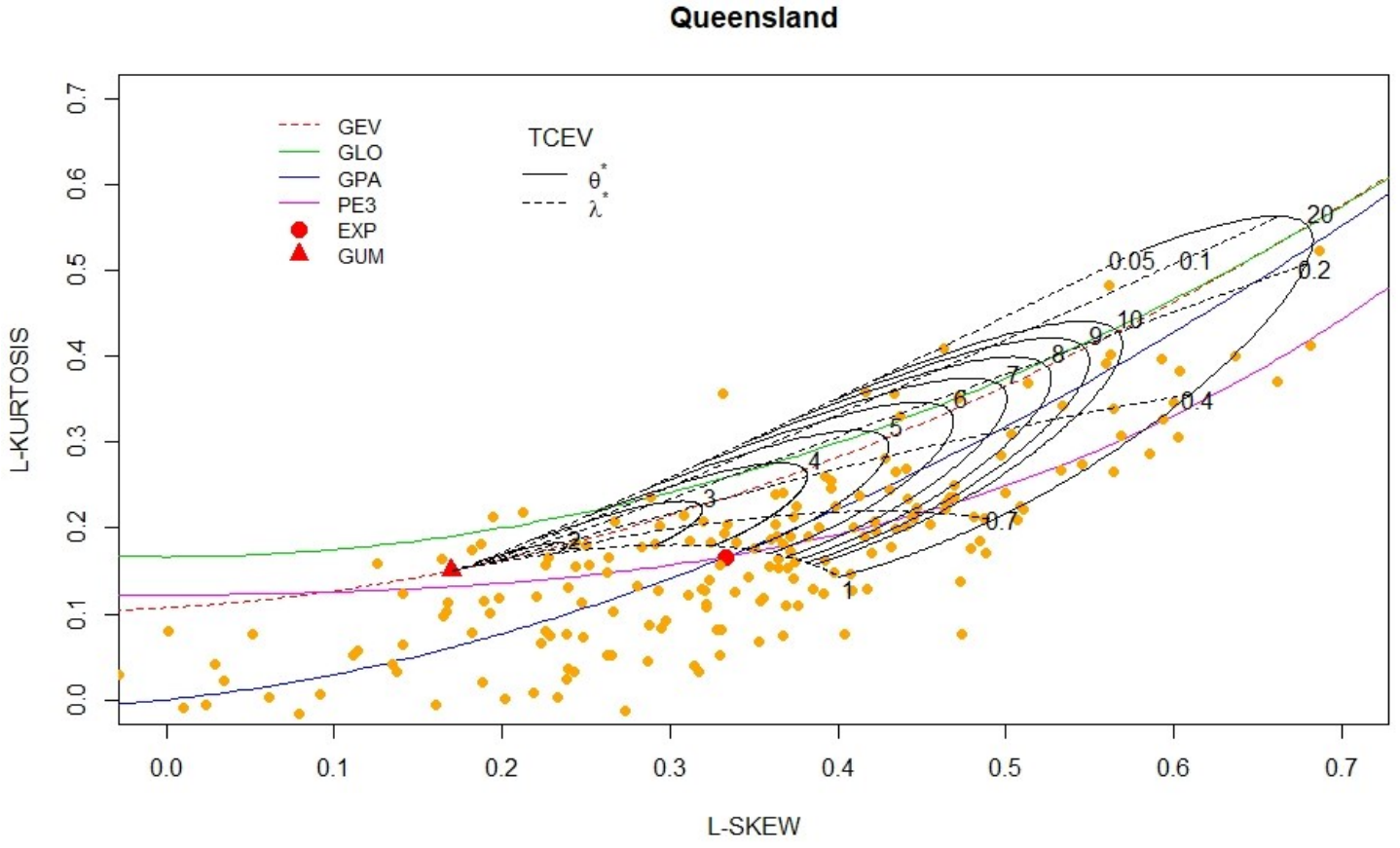

figure B.4 – L-Moments Ratio Diagram for available catchments of Quuensland

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
102101	Fall Ck	Pascoe	-12.88	142.98	1968-2011
104001	Telegraph Rd	Stewart	-14.17	143.39	1970-2011
105105	Developmental Rd	East Normanby	-15.77	145.01	1970-2011
105106	Mount Sellheim	West Normanby	-15.76	144.98	1971-2005
107001	Flaggy	Endeavour	-15.42	145.07	1959-2011
107002	Mount Simon	Annan	-15.65	145.19	1970-1989
108002	Bairds	Daintree	-16.18	145.28	1969-2011
108003	China Camp	Bloomfield	-15.99	145.29	1971-2011
108008	U/S Little Falls Ck	Whyanbeel Ck	-16.39	145.34	1991-2012
110003	Picnic Crossing	Barron	-17.26	145.54	1926-2011
110004	Malones	Emerald Ck	-16.99	145.49	1942-1962
110018	Railway Br	Mazlin Ck	-17.23	145.55	1992-2012
110101	Freshwater	Freshwater Ck	-16.94	145.70	1922-1958
111001	Gordonvale	Mulgrave	-17.10	145.79	1917-1972
111003	Aloomba	Behana Ck	-17.13	145.84	1943-1970
111005	The Fisheries	Mulgrave	-17.19	145.72	1967-2011
111007	Peets Br	Mulgrave	-17.14	145.76	1973-2011
111104	Powerline	Russell	-17.42	145.92	1967-1987
111105	The Boulders	Babinda Ck	-17.35	145.87	1967-2011
112001	Goondi	North Johnstone	-17.53	145.97	1929-1967
112002	Nerada	Fisher Ck	-17.57	145.91	1929-2011
112003	Glen Allyn	North Johnstone	-17.38	145.65	1959-2011

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
112004	Tung Oil	North Johnstone	-17.55	145.93	1967-2011
112101	U/S Central Mill	South Johnstone	-17.61	145.98	1917-2011
112102	Upper Japoonvale	Liverpool Ck	-17.72	145.90	1971-2012
113004	Powerline	Cochable Ck	-17.75	145.63	1967-2011
113007	Ebony Rd	Koolmoon Ck	-17.74	145.56	1986-2012
114001	Upper Murray	Murray	-18.11	145.80	1971-2011
116005	Peacocks Siding	Stone	-18.69	145.98	1936-1971
116008	Abergowrie	Gowrie Ck	-18.45	145.85	1954-2004
116010	Blencoe Falls	Blencoe Ck	-18.20	145.54	1961-2011
116011	Ravenshoe	Millstream	-17.60	145.48	1963-2011
116012	8.7KM	Cameron Ck	-18.07	145.34	1962-2011
116013	Archer Ck	Millstream	-17.65	145.34	1962-2011
116014	Silver Valley	Wild	-17.63	145.30	1962-2011
116015	Wooroora	Blunder Ck	-17.74	145.44	1967-2011
116017	Running Ck	Stone	-18.77	145.95	1971-2011
117002	Bruce HWY	Black	-19.24	146.63	1974-2011
117003	Bluewater	Bluewater Ck	-19.18	146.55	1974-2011
118003	Hervey Range Rd	Bohle	-19.32	146.70	1986-2012
118004	Middle Bohle R Junctio	Little Bohle	-19.33	146.68	1986-2005
118101	Gleesons Weir	Ross	-19.32	146.74	1916-1960
118106	Allendale	Alligator Ck	-19.39	146.96	1975-2011
119004	Bomb Range	Bullock Ck	-19.71	146.92	1972-1991
119006	Damsite	Major Ck	-19.67	147.02	1979-2011
120014	Oak Meadows	Broughton	-20.18	146.32	1971-1998

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
120102	Keelbottom	Keelbottom Ck	-19.37	146.36	1968-2011
120120	Mt. Bradley	Running	-19.13	145.91	1976-2011
120204	Crediton Recorder	Broken	-21.17	148.51	1957-1987
120206	Mt Jimmy	Pelican Ck	-20.60	147.69	1961-1987
120216	Old Racecourse	Broken	-21.19	148.45	1970-2011
120307	Pentland	Cape	-20.48	145.47	1970-2011
121001	Ida Ck	Don	-20.29	148.12	1958-2011
121002	Guthalungra	Elliot	-19.94	147.84	1974-2011
122004	Lower Gregory	Gregory	-20.30	148.55	1973-2011
124001	Caping Siding	O'Connell	-20.63	148.57	1970-2011
124002	Calen	StHelens Ck	-20.91	148.76	1974-2011
124003	Jochheims	Andromache	-20.58	148.47	1977-2011
125002	Sarich's	Pioneer	-21.27	148.82	1961-2011
125004	Gargett	Cattle Ck	-21.18	148.74	1968-2011
125005	Whitefords	Blacks Ck	-21.33	148.83	1974-2011
125006	Dam Site	Finch Hatton Ck	-21.11	148.63	1977-2011
126003	Carmila	Carmila Ck	-21.92	149.40	1974-2011
129001	Byfield	Waterpark Ck	-22.84	150.67	1953-2011
130004	Old Stn	Raglan Ck	-23.82	150.82	1964-2011
130108	Curragh	Blackwater Ck	-23.50	148.88	1973-2005
130207	Clermont	Sandy Ck	-22.80	147.58	1966-2011
130208	Ellendale	Theresa Ck	-22.98	147.58	1965-2003
130215	Lilyvale Lagoon	Crinum Ck	-23.21	148.34	1977-2011
130319	Craiglands	Bell Ck	-24.15	150.52	1961-2011

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
130321	Mt. Kroombit	Kroombit Ck	-24.41	150.72	1964-2004
130335	Wura	Dee	-23.77	150.36	1972-2011
130336	Folding Hills	Grevillea Ck	-24.58	150.62	1973-2011
130348	Red Hill	Prospect Ck	-24.45	150.42	1976-2011
130349	Kingsborough	Don	-23.97	150.39	1977-2011
130413	Braeside	Denison Ck	-21.77	148.79	1972-2011
130503	Wyseby Stn	Carnarvon Ck	-24.97	148.53	1967-1987
130507	Planet Downs	Planet Ck	-24.54	148.91	1973-1992
133003	Marlua	Diglum Ck	-24.19	151.16	1969-2004
135002	Springfield	Kolan	-24.75	151.59	1966-2011
135004	Dam Site	Gin Gin Ck	-24.97	151.89	1966-2011
136006	Dam Site	Reid Ck	-25.27	151.52	1966-2011
136102	Meldale	Three Moon Ck	-24.69	150.96	1949-1980
136108	Upper Monal	Monal Ck	-24.61	151.11	1963-2011
136110	The Gorge	Baywulla Ck	-25.09	151.38	1965-1986
136111	Dakiel	Splinter Ck	-24.75	151.26	1966-2011
136112	Yarrol	Burnett	-24.99	151.35	1966-2011
136202	Litzows	Barambah Ck	-26.30	152.04	1921-2011
136203	Brooklands	Barker Ck	-26.74	151.82	1941-2011
136301	Weens Br	Stuart	-26.50	151.77	1936-2011
137001	Elliott	Elliott	-24.99	152.37	1949-2011
137003	Dr Mays Crossing	Elliott	-24.97	152.42	1975-2011
137101	Burrum HWY	Gregory	-25.09	152.24	1967-2011
137102	Eureka	Sandy Ck	-25.34	152.14	1967-1987

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
137201	Bruce HWY	Isis	-25.27	152.37	1967-2011
137202	Childers	Oaky Ck	-25.29	152.29	1967-1987
138002	Brooyar	Wide Bay Ck	-26.01	152.41	1910-2011
138003	Glastonbury	Glastonbury Ck	-26.22	152.52	1979-2011
138009	Tagigan Rd	Tinana Ck	-26.08	152.78	1975-2011
138010	Kilkivan	Wide Bay Ck	-26.08	152.22	1910-2011
138101	Kenilworth	Mary	-26.60	152.73	1921-1973
138102	Zachariah	Amamoor Ck	-26.37	152.62	1921-2011
138103	Knockdomny	Kandanga Ck	-26.40	152.64	1921-1954
138104	Kidaman	Obi Obi Ck	-26.63	152.77	1921-1963
138106	Baroon Pocket	Obi Obi Ck	-26.71	152.86	1941-1986
138107	Cooran	Six Mile Ck	-26.33	152.81	1948-2011
138110	Bellbird Ck	Mary	-26.63	152.70	1960-2011
138111	Moy Pocket	Mary	-26.53	152.74	1964-2011
138113	Hygait	Kandanga Ck	-26.39	152.64	1972-2011
138120	Gardners Falls	Obi Obi Ck	-26.76	152.87	1987-2012
138903	Bauple East	Tinana Ck	-25.82	152.72	1982-2012
141001	Kiamba	South Maroochy	-26.59	152.90	1938-2011
141003	Warana Br	Petrie Ck	-26.62	152.96	1959-2011
141004	Yandina	South Maroochy	-26.56	152.94	1959-2011
141006	Mooloolah	Mooloolah	-26.76	152.98	1972-2011
141008	Kiels Mountain	Eudlo Ck	-26.66	153.02	1983-2012
141009	Eumundi	North Maroochy	-26.50	152.96	1983-2012
142001	Upper Caboolture	Caboolture	-27.10	152.89	1966-2011

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
142201	Cashs Crossing	South Pine	-27.34	152.96	1918-1963
142202	Drapers Crossing	South Pine	-27.35	152.92	1966-2011
143010	Boat Mountain	Emu Ck	-26.98	152.29	1967-2011
143011	Raeburn	Emu Ck	-27.07	152.01	1966-1985
143015	Damsite	Cooyar Ck	-26.74	152.14	1969-2011
143033	New Beith	Oxley Ck	-27.73	152.95	1989-2012
143101	Mutdapily	Warrill Ck	-27.75	152.69	1915-1953
143102	Kalbar No.2	Warrill Ck	-27.92	152.60	1913-1970
143103	Moogerah	Reynolds Ck	-28.04	152.55	1918-1953
143107	Walloon	Bremer	-27.60	152.69	1962-2011
143108	Amberley	Warrill Ck	-27.67	152.70	1962-2011
143110	Adams Br	Bremer	-27.83	152.51	1972-2011
143113	Loamside	Purga Ck	-27.68	152.73	1974-2011
143203	Helidon Number 3	Lockyer Ck	-27.54	152.11	1927-2011
143208	Dam Site	Fifteen Mile Ck	-27.46	152.10	1957-1985
143209	Mulgowie2	Laidley Ck	-27.73	152.36	1958-2011
143212	Tenthill	Tenthill Ck	-27.56	152.39	1984-2012
143219	Spring Bluff	Murphys Ck	-27.47	151.99	1986-2012
143229	Warrego HWY	Laidley Ck	-27.56	152.39	1991-2012
143303	Peachester	Stanley	-26.84	152.84	1928-2011
143306	U/S Byron Ck Junct	Reedy Ck	-27.14	152.64	1976-2005
143307	Causeway	Byron Ck	-27.13	152.65	1976-2009
143921	Rosentretters Br	Cressbrook Ck	-27.14	152.33	1987-2012
145002	Lamington No.1	Christmas Ck	-28.24	152.99	1910-1954

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
145003	Forest Home	Logan	-28.20	152.77	1918-2011
145005	Avonmore	Running Ck	-28.30	152.91	1922-1952
145007	Hillview	Christmas Ck	-28.22	153.00	1955-1974
145010	5.8KM Deickmans Br	Running Ckreek	-28.25	152.89	1966-2011
145011	Croftby	Teviot Brook	-28.15	152.57	1967-2011
145012	The Overflow	Teviot Brook	-27.93	152.86	1967-2009
145013	Rudd's Lane	Christmas Ck	-28.17	152.98	1968-1987
145018	Up Stream Maroon Dam	Burnett Ck	-28.22	152.61	1971-2011
145020	Rathdowney	Logan	-28.22	152.87	1974-2011
145101	Lumeah Number 2	Albert	-28.06	153.04	1911-2011
145102	Bromfleet	Albert	-27.91	153.11	1919-2011
145103	Good Dam Site	Cainbable Ck	-28.09	153.08	1963-2011
145104	32.2KM	Canungra Ck	-28.06	153.12	1966-1987
145107	Main Rd Br	Canungra Ck	-28.00	153.16	1974-2011
146002	Glenhurst	Nerang	-28.00	153.31	1920-2011
146003	Camberra Number 2	Currumbin Ck	-28.20	153.41	1928-1982
146004	Neranwood	Little Nerang Ck	-28.13	153.29	1927-1961
146005	Chippendale	Tallebudgera Ck	-28.16	153.40	1927-1953
146007	Pump House	Tallebudgera Ck	-28.15	153.40	1936-1962
146010	Army Camp	Coomera	-28.03	153.19	1963-2011
146011		Nerangwhipbird	-28.09	153.26	1966-1985
146012	Nicolls Br	Currumbin Ck	-28.18	153.42	1971-2011
146014	Beechmont	Back Ck	-28.12	153.19	1972-2011
146020	Springbrook Rd	Mudgeeraba Ck	-28.09	153.35	1990-2012

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
146095	Tallebudgera Ck Rd	Tallebudgera Ck	-28.15	153.40	1971-2011
416303	Clearview	Pike Ck	-28.81	151.52	1935-1987
416305	Beebo	Brush Ck	-28.69	150.98	1969-2011
416312	Texas	Oaky Ck	-28.81	151.15	1970-2011
416410	Barongarook	Macintyre Brook	-28.44	151.46	1968-2011
422210	Tabers	Bungil Ck	-26.41	148.78	1967-2011
422302	Killarney	Spring Ck	-28.35	152.34	1910-1955
422303	Killarney	Spring Ck South	-28.36	152.34	1910-1955
422304	Elbow Valley	Condamine	-28.37	152.16	1916-1972
422305	Gillespies	Emu Ck	-28.22	152.28	1924-1945
422306	Swanfels	Swan Ck	-28.16	152.28	1920-2011
422307	Kings Ck	Kings Ck	-27.90	151.91	1921-1966
422313	Emu Vale	Emu Ck	-28.23	152.23	1948-2011
422317	Rocky Pond	Glengallan Ck	-28.13	151.92	1954-1991
422319	Allora	Dalrymple Ck	-28.04	152.01	1969-2011
422321	Killarney	Spring Ck	-28.35	152.33	1960-2011
422326	Cranley	Gowrie Ck	-27.52	151.94	1970-2011
422334	Aides Br	Kings Ck	-27.93	151.86	1970-2011
422338	Leyburn	Canal Ck	-28.03	151.59	1975-2011
422341	Brosnans Barn	Condamine	-28.33	152.31	1977-2011
422394	Elbow Valley	Condamine	-28.37	152.14	1973-2011
915011	Mt Emu Plains	Porcupine Ck	-20.18	144.52	1972-2011
917104	Roseglen	Etheridge	-18.31	143.58	1967-2011
917107	Mount Surprise	Elizabeth Ck	-18.13	144.31	1969-2011

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
919005	Fonthill	Rifle Ck	-16.68	145.23	1969-2011
919013	Mulligan HWY	McLeod	-16.50	145.00	1973-2011
919201	Goldfields	Palmer	-16.11	144.78	1968-2011
919305	Nullinga	Walsh	-17.18	145.30	1957-1992
922101	Racecourse	Coen	-13.96	143.17	1968-2011
926002	Dougs Pad	Dulhunty	-11.83	142.42	1971-2011

B.3 – Northern Territory

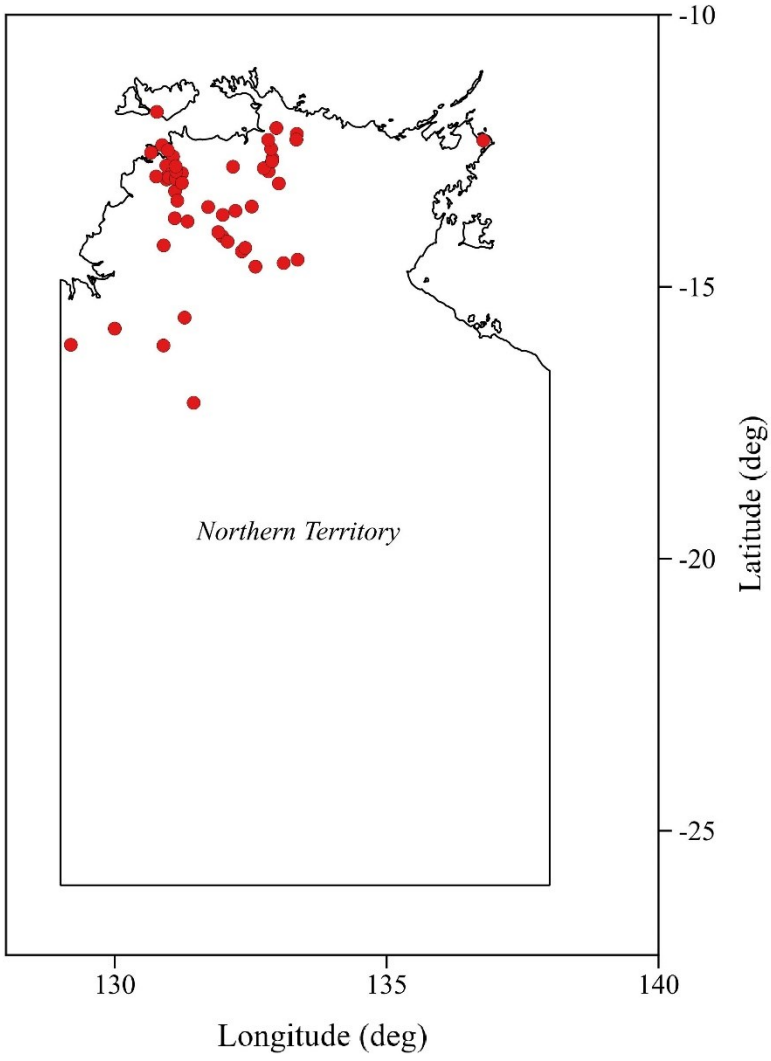

figure B.5 – Available sites for Northern Territory

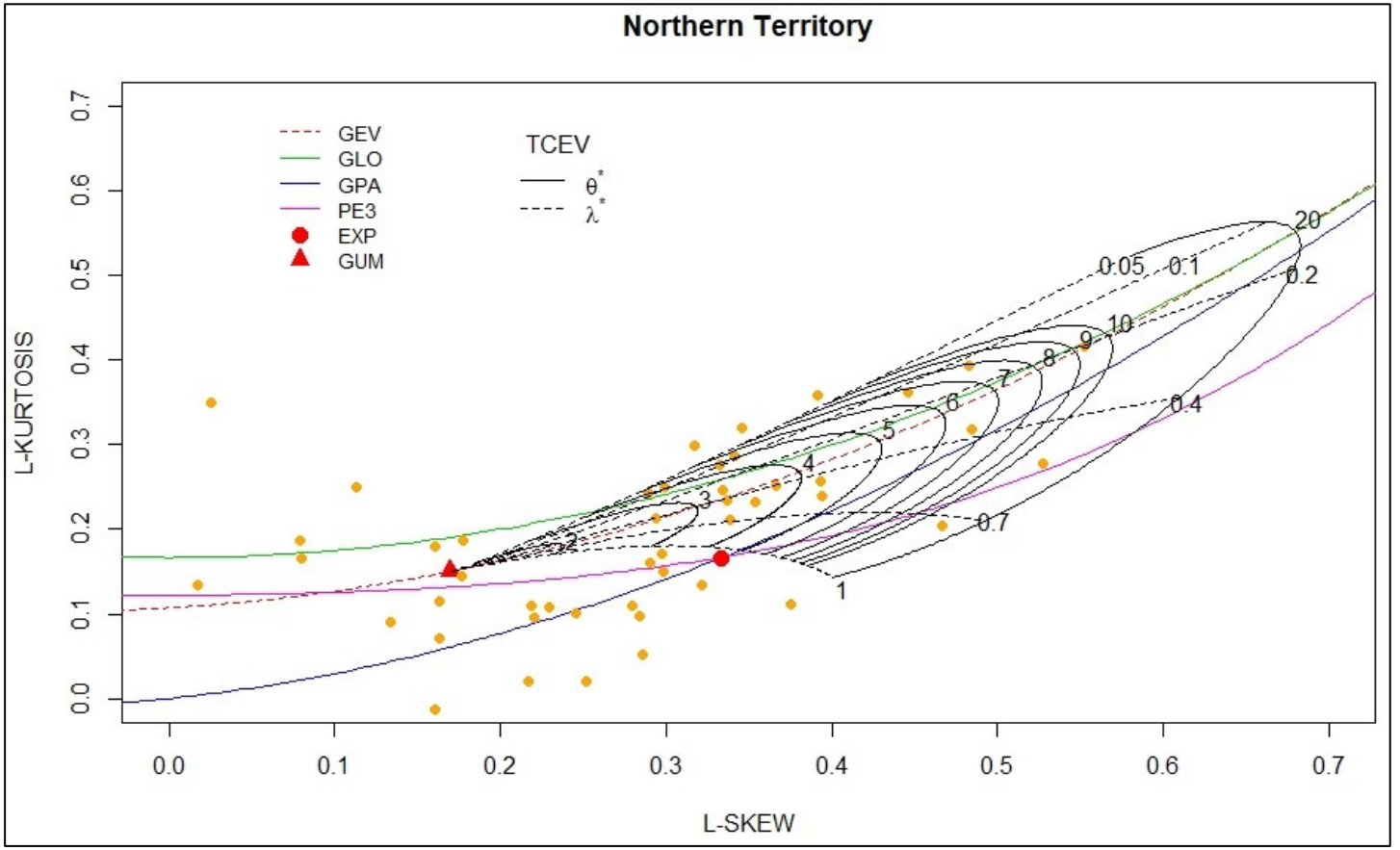

figure B. 6 – L-Moments Ratio Diagram for available catchments of Northern Territory

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
G8100189	Victoria HWY	Moriarty Ck	-16.065	129.1933	1967 - 1985
G8110004	Victoria HWY	East Baines	-15.7667	130	1963 - 2008
G8110014	U/S Fig Tree Yard	Sullivan's Ck	-15.565	131.285	1970 - 1992
G8110110	V.R.D. Rd Crossing	Surprise Ck	-16.0783	130.8967	1960 - 2003
G8110263	1.5 Miles D/S Bore	Bullock Ck	-17.1317	131.4517	1971 - 1992
G8140008	Old Railway Br	Fergusson	-14.07	131.9767	1958 - 2011
G8140061	Blue Hole	Copperfield Ck	-13.9933	131.9033	1958 - 1977
G8140063	D/S Old Douglas H/S	Douglas	-13.7967	131.3383	1958 - 2011
G8140086	D/S Stuart HWY	King	-14.6283	132.5883	1964 - 1986
G8140152	Dam Site	Edith	-14.1683	132.075	1962 - 2008
G8140158	Dam Site	McAdden Ck	-14.3483	132.3383	1964 - 2011
G8140159	Waterfall View	Seventeen Mile C	-14.2833	132.4	1963 - 2008
G8140161	Tipperary	Green Ant Ck	-13.7383	131.1033	1966 - 2011
G8140166	Gorge	Fish	-14.2367	130.9	1963 - 1985
G8150010	Batchelor Damsite	Finniss	-13.025	130.9533	1975 - 2011
G8150018	Stuart HWY	Elizabeth	-12.605	131.0733	1955 - 2011
G8150096	Cox Peninsula	Carawarra Ck	-12.5317	130.6683	1966 - 2011
G8150097	Rum Jungle +Ansto Eb4	East Finniss	-12.965	130.9683	1966 - 2009

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
G8150098	Tumbling Waters	Blackmore	-12.77	130.9483	1960 - 2010
G8150127	D/S McMillans Rd	Rapid Ck	-12.3933	130.8717	1964 - 2011
G8150151	U/S Darwin R Dam	Celia Ck	-12.91	131.0533	1972 - 2010
G8150180	Gitchams	Finniss	-12.97	130.7617	1961 - 2007
G8150200	Rum Jungle Rd Crossing	East Finniss	-12.99	131	1982 - 2007
G8150233	McArthur Park	Palmerston Catch	-12.4883	130.975	1984 - 2003
G8160235	Damsite	Takamprimili	-11.7817	130.775	1967 - 1986
G8170002	Railway Br	Adelaide	-13.2417	131.1083	1954 - 2007
G8170020	Dirty Lagoon	Adelaide	-12.91	131.235	1963 - 2011
G8170062	Eighty-Seven Mile Jump Up	Burrell Ck	-13.415	131.1517	1958 - 1985
G8170066	Stuart HWY	Coomalie Ck	-13.0133	131.1233	1958 - 2010
G8170075	U/S Manton Dam	Manton	-12.8783	131.13	1965 - 2010
G8170084	Tortilla Flats	Adelaide	-13.09	131.235	1960 - 2011
G8170085	Stuart HWY	Acacia Ck	-12.7833	131.12	1964 - 2011
G8180026	El Sherana Rd Crossing	Mary	-13.6017	132.22	1962 - 2011
G8180069	near Burrundie	McKinlay	-13.5317	131.7183	1959 - 2009
G8180252	D/S El Sherana Rd	Harriet Ck	-13.6767	131.9867	1965 - 2010
G8190001	U/S Arnhem HWY	West Alligator	-12.7917	132.175	1977 - 2010

Station ID	Station Name	River Name	Gauge Lat	Gauge Lon	Period of Record
G8200045	El Sherana (C)	South Alligator	-13.5233	132.52	1958 - 2009
G8200046	Coljon (C Part)	Deaf Adder Ck	-13.0983	133.0183	1972 - 1991
G8200049	near Nourlangie Rock	Koongarra Ck	-12.8767	132.83	1978 - 2005
G8200112	Kakadu HWY	Nourlangie Ck	-12.8183	132.7417	1962 - 2006
G8210001	Nimbuwah (C)	Cooper Ck	-12.1867	133.3483	1971 - 1992
G8210009	D/S Jabiru	Magela Ck	-12.6417	132.9	1972 - 2011
G8210012	George Town Crossing	Gulungul Ck (Bog	-12.69	132.8933	1972 - 1992
G8210016	Mt. Borradaile	Cooper Ck	-12.08	132.9733	1980 - 2006
G8210017	Jabiluka Billabong	Magela Ck Plains	-12.4617	132.875	1974 - 2006
G8210019	Outflow Main Channel	Magela Plains	-12.2967	132.8217	1976 - 2004
G8210024	D/S Nabarlek	Cooper Ck	-12.2933	133.34	1979 - 2006
G8260053	above Tidal Reach	Lower Latram	-12.3083	136.7783	1964 - 1984
G9030089	Rd Br	Waterhouse	-14.5617	133.1067	1973 - 2011
G9030090	Wattle Hill	Chambers Ck	-14.5	133.3633	1974 - 1992

CURRICULUM VITAE

Generalities

Name	Vincenzo Totaro
Place and date of birth	Foggia, Italy, 11 January 1987
Family	Unmarried
Structure	Polytechnical University of Bari - Department of Civil, Environmental, Land, Building Engineering and Chemistry
Address	Via E. Orabona 4, 70125, Bari (Italy)
Phone	+39 348 6716446 (Italy)
e-mail	vincenzo.totaro@poliba.it
web-page	https://www.researchgate.net/profile/Vincenzo_Totaro2

Education

- November 2011,16: three years degree in Civil Engineering with full marks and honors at Technical University of Bari, discussing a thesis on 'Hydraulic model of Carapelle rives'. Supervisor: prof. A. F. Piccinni
- February 2016, 18: two years Master's degree in Civil Engineering with full marks and honors at Technical University of Bari, discussing a thesis on 'Bayesian approach for evaluation of uncertainty in extreme value distributions'. Supervisor: prof. V. Iacobellis, cosupervisor: prof. A. Gioia

Positions

- June 28, 2019 – current: Research Fellow at Polytechnic University of Bari for the project “Prototyping and customization of a tertiary treatment aimed at the agricultural reuse of the effluents of the Taranto-Gennarini wastewater treatment plant”. Scientific responsible: prof. Alberto Ferruccio Piccinni
- November 1, 2016 – current: Ph. D. student in the XXXII cycle of the PhD course in Risk and environmental, territorial and building development at Polytechnic University of Bari. Supervisor: prof. Vito Iacobellis, Co-supervisor: prof. George A. Kuczera
- February, 2018 – July, 2018: visiting Ph. D. student at The University of Newcastle (Australia), under the supervision of prof. George Kuczera

Other activities

07/05/2019 – 14/07/2019, 15/06/2019 - 22/06/2019: visiting Ph.D. student at Department of Hydrology and Hydrodynamics - Institute of Geophysics, Polish Academy of Sciences, Warsaw (Poland) under the supervision of prof. Ewa Bogdanowicz and Krzysztof Kochanek.

Scientific interests

My research activities focus on Hydrology, with specific interest on: theoretically derived distributions of floods; Bayesian analysis; climate changes; non-stationarity in hydrological variables; DEM-based models for the evaluation of flood-prone areas.

- Fedele, V., Motta Zanin, G., Parisi, A., and Totaro, V. (2019): The complexity of risk in urban environment and the role of technological innovation. 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, pp. 1080-1084. doi: 10.1109/SMC.2019.8914640
- Totaro, V., Gioia, A., and Iacobellis, V.: Power of parametric and non-parametric tests for trend detection in annual maximum series, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-363, Accepted for publication, 2019.
- Totaro V., Peschechera G., Gioia A., Iacobellis V., Fratino U. (2019) Comparison of Satellite and Geomorphic Indices for Flooded Areas Detection in a Mediterranean River Basin. In: Misra S. et al. (eds) Computational Science and Its Applications – ICCSA 2019. ICCSA 2019. Lecture Notes in Computer Science, vol 11622. Springer, Cham
- Balacco G., Totaro V., Gioia A., Piccinni A.F. (2019) Evaluation of Geomorphic Descriptors Thresholds for Flood Prone Areas Detection on Ephemeral Streams in the Metropolitan Area of Bari (Italy). In: Misra S. et al. (eds) Computational Science and Its Applications – ICCSA 2019. ICCSA 2019. Lecture Notes in Computer Science, vol 11622. Springer, Cham
- Gioia, A., Totaro, V., Bonelli, R., Esposito, A. A., Balacco, G., & Iacobellis, V. (2018, May). Flood susceptibility evaluation on ephemeral streams of Southern Italy: a case study of Lama Balice. In International Conference on Computational Science and Its Applications (pp. 334-348). Springer, Cham.
- Iacobellis, V., Martellotta, A. M., Gioia, A., Prato, D., Totaro, V., Bonelli, R., Balacco G. & Esposito, A. A. (2018, May). Investigation of a Flood Event Occurred on Lama Balice, in the Context of Hazard Map Evaluation in Karstic-Ephemeral Streams. In International Conference on Computational Science and Its Applications (pp. 317-333). Springer, Cham.
- Totaro, V., Gioia, A., Novelli, A., & Caradonna, G. (2017, July). The use of geomorphological descriptors and landsat-8 spectral indices data for flood areas evaluation: a case study of Lato river basin. In International

Conference on Computational Science and Its Applications (pp. 30-44). Springer, Cham.

- Fiore G., Gioia A., Totaro V., Caradonna G., Novelli A., Tarantino E., Iacobellis V., Fratino U. (2016). L'utilizzo di immagini satellitari per la definizione di aree inondabili. In: Atti conferenza nazionale ASITA. vol. XX, p. 771-778, ISBN: 978-88-941232-6-5, Cagliari, 8- 10 Novembre 2016

Bari, 20/12/2019