UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Engenharia Civil, Arquitetura e Urbanismo

ROGERIO DE ALMEIDA

SIMULAÇÃO PROBABILÍSTICA DE EVENTOS DE SECA COM APLICAÇÃO DE CÓPULAS GAUSSIANA E ARQUIMEDIANAS

CAMPINAS

2020

ROGERIO DE ALMEIDA

SIMULAÇÃO PROBABILÍSTICA DE EVENTOS DE SECA COM APLICAÇÃO DE CÓPULAS GAUSSIANA E ARQUIMEDIANAS

Tese de Doutorado apresentada a Faculdade de Engenharia Civil, Arquitetura e Urbanismo da Unicamp, para obtenção do título de Doutor em Engenharia Civil, na área de Recursos Hídricos, Energéticos e Ambientais.

Orientador: Prof. Dr. Paulo Sergio Franco Barbosa

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO ROGERIO DE ALMEIDA E ORIENTADO PELO PROF. DR. PAULO SERGIO FRANCO BARBOSA.

ASSINATURA DO ORIENTADOR

____

CAMPINAS

2020

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Luciana Pietrosanto Milla - CRB 8/8129

AL64s

Almeida, Rogerio de 1973-
Simulação probabilística de eventos de secas com aplicação de cópulas gaussiana e arquimedianas / Rogério de Almeida. – Campinas, SP : [s.n.], 2020

Orientador: Paulo Sergio Franco Barbosa.
Tese (doutorado) – Universidade Estadual de Campinas, Faculdade de Engenharia Civil, Arquitetura e Urbanismo.

1. Engenharia hidrológica. 2. Previsão hidrológica. 3. Seca. 4. Cópulas. 5. Cópulas (Estatística matemática). I. Barbosa, Paulo Sérgio Franco, 1957-. II. Universidade Estadual de Campinas. Faculdade de Engenharia Civil, Arquitetura e Urbanismo. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Probabilistic simulation of drought events with application of gaussian and archimedian copulas
Palavras-chave em inglês:
Hydrological engineering
Hydrological forecast
Drought
Copulas
Copulas (Mathematical statistics)
Área de concentração: Recursos Hídricos, Energéticos e Ambientais
Titulação: Doutor em Engenharia Civil
Banca examinadora:
José Anderson do Nascimento Batista
Rosângela Ballini
Jurandir Zullo Junior
Jefferson Nascimento de Oliveira
Benedito Cláudio da Silva
Data de defesa: 05-02-2020
Programa de Pós-Graduação: Engenharia Civil

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: http://orcid.org/0000-0002-0469-3233
- Currículo Lattes do autor: http://lattes.cnpq.br/5641497444183363

UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA CIVIL, ARQUITETURA E URBANISMO

SIMULAÇÃO PROBABILÍSTICA DE EVENTOS DE SECA COM APLICAÇÃO DE CÓPULAS GAUSSIANA E ARQUIMEDIANAS

Rogerio de Almeida

Tese de Doutorado aprovada pela Banca Examinadora, constituída por:

Prof. Dr. José Anderson Batista Nascimento
Presidente /FEC/UNICAMP

Profa. Dra. Rosângela Ballini
IE/UNICAMP

Prof. Dr. Jurandir Zullo Junior
CEPAGRI/ÚNICAMP

Prof. Dr. Jefferson Nascimento de Oliveira
FEIS/UNESP

Prof. Dr. Benedito Cláudio da Silva
IRN/UNIFEI

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 05 de fevereiro de 2020.

DEDICATÓRIA

A minha esposa, Viviane Aguirre Amo, que esteve ao meu lado em todos os momentos dessa longa jornada. Obrigado por todo amor, carinho, companheirismo e apoio incondicional.

Aos meus pais Laercio Ribeiro Duarte de Almeida e Maria Luisa Soldan de Almeida, exemplos de muito trabalho e dedicação. Obrigado por me darem uma educação de qualidade e por propiciarem um ambiente seguro para meu desenvolvimento como pessoa e como profissional.

Ao meu irmão José Francisco de Almeida.

AGRADECIMENTOS

A Deus, por tudo que tenho e por todas as oportunidades que me foram dadas.

Ao Prof. Dr. Paulo Sergio Franco Barbosa, meu orientador e um grande amigo. Obrigado pelo apoio, pela paciência, pelos ensinamentos e também pela confiança em mim depositada.

A minha tia Maria Terezinha Soldan de Alcântara por todo apoio, carinho e preocupação na fase inicial de meu doutorado.

Ao meu amigo Henrique Leme Felizatti pelo apoio e pelas ótimas discussões onde abordando os mais variados temas sobre estatística, simulações e modelagem de séries temporais.

Ao colega Erick Andrade Busato pelo apoio e pela troca de informações sobre o uso de acoplamentos na análise multivariada de variáveis aleatórias.

Ao meu amigo Jairo Colombo pelos bons momentos e pela alegre conivência durante nossa permanência na Faculdade de Engenharia Civil, Arquitetura e Urbanismo (FEC) da Universidade Estadual de Campinas (UNICAMP).

À CPFL na pessoa do Vice Presidente de Operações Reguladas, Luís Henrique Ferreira Pinto, que sempre me permitiu dedicar tempo para os estudos e conclusão deste trabalho.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) na forma de bolsa de Doutorado (DR-II) no pais (processo nº 02/02365-1).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (01-P-27235/2002).

RESUMO

Neste trabalho é apresentado um método para simular a ocorrência de eventos de seca, baseado na aplicação de cópulas Arquimedianas e Gaussiana. O método foi aplicado ao Rio Camanaducaia localizado na região sudeste do Brasil. A teoria dos runs, proposta por Yevjevich (1967), foi utilizada para definição e caracterização dos eventos de secas a partir dos dados históricos de vazões médias mensais de 1944 a 2016. Quatro níveis de truncamento foram utilizados para definir os eventos de seca, correspondentes aos percentis 98, 95, 92 e 90 das vazões médias mensais. Seis distribuições de probabilidade uni-variadas foram avaliadas para representar o comportamento probabilístico das durações e severidades dos eventos de seca, sendo elas: Gama com dois parâmetros; Exponencial; Weibull; Log-Normal; Pareto Generalizada e Generalizada de Valores Extremos. O Critério de Informação de Akaike e o teste de Kolmogorov-Smirnov foram utilizados para auxiliar na seleção dos modelos probabilísticos uni-variados mais adequados, e avaliação do ajuste aos dados observados, respectivamente. Tendo como base as distribuições de probabilidade uni-variadas mais adequadas, cópulas Arquimedianas, do tipo, Clayton, Frank, Gumbel, Joe e Ali-Mikhali-Haq, e Elíptica do tipo Gaussiana foram utilizadas na análise conjunta das durações e severidades dos eventos de seca. Os resultados demonstram que o método proposto é uma ferramenta viável para simular a ocorrência de eventos de seca aplicado ao Rio Camanducaia, localizado na região sudeste do Brasil. O método preserva as características probabilísticas das durações e severidades dos eventos de seca e suas medidas de correlação. As cópulas assimétricas com dependência na cauda superior e simétricas sem dependência nas caudas, se mostram adequadas na modelagem conjunta das durações e severidades. A análise do período de retorno dos eventos de seca demonstrou que grande parte das secas obtidas com base na série histórica de vazões possuem recorrência inferior a 50 anos, enquanto que os eventos mais críticos possuem recorrência superior a 300 anos.

ABSTRACT

This study presents a method to simulate the occurrence of drought events, based on the application of Archimedian and Gaussian copulas. The method was applied to the Camanaducaia River located in the southeastern region of Brazil. Yevjevich's (1967) theory of runs was used to define and characterize drought events from historical monthly average flow data from 1944 to 2016. Four truncation levels were used to define drought events, corresponding to 98, 95, 92 and 90 percentiles of the monthly average streamflows. Six univariate probability distributions were evaluated to represent the probabilistic behavior of drought durations and severities: 2-parameter Gamma; Exponential; Weibull; Log-Normal; Generalized Pareto and Generalized Extreme Values. The Akaike Information Criterion and the Kolmogorov-Smirnov test were used to assist in selecting the most appropriate univariate probabilistic models, and to evaluate those that better fit the observed data, respectively. Based on the most appropriate univariate probability distributions, Archimedean, Clayton, Frank, Gumbel, Joe and Ali-Mikhali-Haq, and Elliptic Gaussian copulas were used in the joint analysis of the durations and severities of the drought events. The results demonstrate that the proposed method is a useful tool to simulate the occurrence of drought events applied to the Camanducaia River, located in southeastern Brazil. The method preserves the probabilistic characteristics of drought durations and severities and the dependencies between both. Asymmetric copulas with dependence on the upper tail and symmetrical without dependence on the tails, are adequate in the joint modeling of durations and severities. The analysis of the return period of drought events also evidenced that most of the droughts obtained from the historical series of streamflow have a recurrence of less than 50 years, while the most critical events have a recurrence of more than 300 years.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AIC	Critério de Informação de Akaike.
AMH	Ali-Mikhail-Haq.
ANA	Agência Nacional de Águas.
AR(1)	Auto Regressivo de ordem 1ª ordem.
ARMA	Modelo Auto Regressivo de Média Móvel.
CDF	Função de Distribuição Acumulada
CLA	Clayton.
CR	Sala de Controle.
DAEE	Departamento de Águas e Energia Elétrica
DP	Desvio padrão.
DPI	Índice Potencial de Secas.
EXP	Exponencial.
FRA	Frank.
GAM2	Gama com dois parâmetros.
GAU	Gaussiana.
GEV	Generalizada de Valores Extremos.
GOF	Goodness-of-Fit.
GP	Pareto Generalizada.
GUM	Gumbel.
IBGE	Instituto Brasileiro de Geografia e Estatística.
JOE	Joe.
KS	Kolmogorov-Smirnov.
LOGN	Log-Normal.
Max	Máximo.
Med	Mediana.
Méd	Média.
Min	Mínimo.
MLE	Método da Máximo Verossimilhança.
$P98$	Percentil 98.
$P95$	Percentil 95.

$P92$	Percentil 92.
$P90$	Percentil 90.
Q_{7-10}	Vazão mínima com 10 anos de recorrência e duração de 7 dias.
r	Coeficiente de correlação de Pearson.
REC	Resolução Conjunta.
SDI	Índice de Vazão de Seca.
SHI	Índice Hidrológico Padronizado.
UGRHI	Unidade de Gerenciamento de Recursos Hídricos.
WBL	Weibull.

LISTA DE SÍMBOLOS

$\mathbf{1}$	Função indicador.
c	Função densidade da cópula.
C	Cópula.
$\hat{C}$	Cópula empírica.
$\mathcal{C}_0$	Classe de cópula paramétrica.
D_n	Estatística do teste de Kolmogorov-Smirnov para a amostra observada.
D_{n_i}	Estatística do teste de Kolmogorov-Smirnov para a i-ézima pseudo observação.
E	Valor esperado.
f	Função densidade de probabilidade.
F	Distribuição de probabilidade acumulada.
$\hat{F}$	Distribuição de probabilidade acumulada empírica.
F_k	Distribuição de probabilidade da k-ézima variável aleatória.
$\hat{F}_k$	Distribuição de probabilidade acumulada empírica da k-ézima variável aleatória.
$F_{X_1 X_2}$	Distribuição de probabilidade conjunta das variáveis aleatórias X_1 e X_2.
$F^{(-1)}$	Inversa da distribuição de probabilidade.
${F_k}^{(-1)}$	Inversa da distribuição de probabilidade da k-ézima variável aleatória.
$\mathcal{F}_0$	Classe de distribuição de probabilidade paramétrica.
h	Graus de liberdade.
H	Distribuição de probabilidade conjunta de um vetor de variáveis aleatórias.
H_0	Hipótese nula ou verdadeira.
H_1	Hipótese alternativa.
K	Defasagem ou lag.
L	Intervalo de tempo entre inicios de eventos consecutivos.
$\mathcal{L}$	Função de verossimilhança.
$\hat{l}$	Logaritmo da função de verossimilhança
m	Número de pseudo observações.

n	Número de observações.
N	Total de eventos ou total de classes de distribuições de probabilidade ou cópulas paramétricas.
$\mathbb{N}$	Conjunto dos números naturais.
p	Ordem da parte auto regressiva do modelo ARMA ou número de parâmetros da distribuição de probabilidade ou cópula.
P	Probabilidade.
q	Ordem da parte de média móvel do modelo ARMA.
q_0	Valor de arbitrário, valor referência, nível de truncamento ou vazão de referência.
Q	Variável aleatória (vazão) ou estatística dos testes de Box-Pierce e Ljung.
$\hat{r}_j$	Auto correlação estimada no lag j
$\mathbb{R}$	Conjunto dos números reias.
S_n	Estatística de Cramer-von Mises.
S_{ni}	Estatística de Cramer-von Mises para a i-ézima pseudo observação.
$t_{inicial}$	Tempo de início do evento.
t_{final}	Tempo de término do evento.
T	Tempo máximo.
T_{X_1}	Período de retorno da variável aleatória X_1.
T_{X_2}	Período de retorno da variável aleatória X_2.
$T_{X_1X_2}$	Período de retorno conjunto das variáveis aleatórias X_1 e X_2 para $P(X_1 \geq x_1, X_2 \geq x_2)$.
$\widehat{U}$	Pseudo observação da variável aleatória U.
$\boldsymbol{U}$	Vetor de variáveis aleatórias.
$\widehat{\boldsymbol{U}}$	Pseudo observação do vetor de variáveis aleatória $\boldsymbol{U}$.
X	Variável aleatória.
X_1	Variável aleatória (duração).
X_2	Variável aleatória (severidade).
$\boldsymbol{X}$	Vetor de variáveis aleatórias.
$\hat{X}_1$	Variável aleatória ajustada (duração).

$\hat{X}_2$	Variável aleatória ajustada (severidade).
y_1	Soma da derivação positiva entre inícios sucessivos de derivações positiva e negativa.
y_2	Soma da derivação negativa entre inícios sucessivos de derivações negativa e positiva.
Z	Intervalo de tempo entre eventos com durações maiores ou iguais a um valor arbitrário.
$\mathbb{Z}^+$	Conjunto dos número reais positivos no intervalo [0,1].
W	Intervalo de tempo entre eventos com severidades maiores ou iguais a um valor arbitrário.
Δx_1	Intervalo de tempo entre eventos.
Δx_2	Volume entre eventos.
α	Parâmetro de forma das distribuições de probabilidade Gama, Weibull, Pareto Generalizada e Generalizada de Valores Extremos.
$\hat{\alpha}$	Estimativa de α.
β	Parâmetro de escala das distribuições de probabilidade Gama, Weibull, Pareto Generalizada e Generalizada de Valores Extremos.
$\hat{\beta}$	Estimativa de β.
γ	Constante de Euler.
θ	Parâmetro (distribuição de probabilidade ou cópula).
$\hat{\theta}$	Estimador de máximo verossimilhança ou estimativa de θ.
λ	Taxa da distribuição de probabilidade Exponencial.
$\hat{\lambda}$	Estimativa de λ.
μ	Média.
$\hat{\mu}$	Estimativa de μ.
ξ	Parâmetro de locação das distribuições de probabilidade Pareto Generalizada ou Generalizada de Valores Extremos.
$\hat{\xi}$	Estimativa de ξ.
φ	Função geradora da cópula.
φ^{-1}	Inversa de φ.
$\varphi^{[-1]}$	Pseudo inversa de φ.

σ	Desvio padrão.
$\hat{\sigma}$	Estimativa de σ.
τ_1	Distância entre os inícios de derivações positivas.
τ_2	Distância entre os inícios de derivações negativas.
τ_3	Distância entre picos sucessivos de derivações positivas.
τ_4	Distância entre picos sucessivos de derivações negativas.
τ_5	Distância entre inícios sucessivos de derivações positiva e negativa.
τ_6	Distância entre inícios sucessivos de derivações negativa e positiva.
Θ	Conjunto dos parâmetros.
Φ^{-1}	Inversa da distribuição de probabilidade Normal Padronizada.
Γ	Função gama.
$\chi^2_{1-\alpha,h}$	Distribuição de probabilidade qui-quadrado com nível de significância $1-\alpha$ e h graus de liberdade.

SUMÁRIO

1. INTRODUÇÃO

Segundo Cunha et al. (1983), as secas, juntamente com as enchentes, ciclones tropicais e terremotos, são consideradas responsáveis por mais de 90% de todas as perdas causadas ao meio ambiente vital ao Homem. Outros desastres, tais como erupções vulcânicas, deslizamentos de encostas, avalanches, tornados, tsunamis e incêndios, embora de importância local, em geral, são de menor significância global, exceto se tomados em termos agregados. Eventos como enchentes, terremotos e ciclones iniciam de maneira repentina, possuindo curta duração e sendo restritos a um local de influência, já as secas iniciam lentamente, podendo atingir longas durações e cobrir vastas áreas. Diferentemente das enchentes, dos ciclones e dos terremotos, que são desastres naturais associados a um evento extremo de grande impacto, as secas são resultado da associação de eventos de baixo impacto acumulados ao longo do tempo.

O termo "seca" possui diferentes conotações em várias partes do mundo, segundo Hudson Jr. e Hazen (1964), como por exemplo, em Bali, na Indonésia, qualquer período de seis dias ou mais sem precipitações é considerado como seca, enquanto que, na Líbia, secas são somente reconhecidas após dois anos sem precipitação. No Egito, qualquer ano no Rio Nilo sem a ocorrência de cheias representa uma seca, independentemente das precipitações. Ressaltando-se que essa percepção de secas no Egito é anterior a construção do reservatório de Assuã finalizado em 1970.

Segundo Santos (1998), as secas também são percebidas de maneira distinta entre regiões, devidos às suas diferentes características climáticas, diferentes níveis de utilização da água, e também por serem resultado da inter-relação entre os sistemas naturais, sujeitos às flutuações climáticas, e os sistemas concebidos pelo Homem, com suas especificidades e vulnerabilidades. Este fato contribui para a dificuldade de uma definição rigorosa e universal para as secas e, consequentemente, de um modelo único de abordagem para o seu estudo, conduzindo, assim, a várias definições e diferentes metodologias de análise. Desta forma, Dracup et al. (1980a) ressalta a importância de se definir exatamente qual é o significado do termo "seca" antes que qualquer análise proposta seja desenvolvida e utilizada.

Segundo Mirsha e Singh (2010), na definição da seca e de seus impactos, podemos considerar os aspectos conceituais e os operacionais. Do ponto de vista conceitual, a definição de uma seca é realizada em termos relativos, como por exemplo, “um período seco”, “uma seca severa”, provendo pouca orientação sobre o fenômeno. Enquanto que, do ponto de vista operacional, as definições utilizam a análise de frequências das durações e severidades (comportamentos probabilísticos) tendo como base um período de tempo pré-determinado.

Muitas definições para o termo “seca” podem ser encontradas na literatura. A Agência Nacional de Águas (ANA) define as secas como a ausência prolongada, deficiência acentuada ou fraca distribuição de precipitações em uma determinada região, ou ainda, um período de tempo excepcionalmente seco, suficientemente prolongado para que a falta de precipitação provoque um grave desequilíbrio hidrológico.

Na definição apresentada pela ANA, pode-se notar que, a precipitação representa a causa primária das secas, enquanto o efeito decorrente das baixas taxas de precipitação sobre os sistemas de recursos hídricos de uma determinada região, em um determinado período de tempo, representa a causa secundária ou o impacto da seca. Sendo assim, quando as taxas de precipitação permanecem abaixo de um valor considerado como normal, durante um determinado período de tempo, em uma determinada área, as secas são classificadas como climáticas ou meteorológicas. Para o caso onde os recursos hídricos de uma determinada região, rios, lagos ou reservatórios, não são mais capazes de suprir adequadamente seus usuários, as secas são classificadas como hidrológicas.

Assim, as secas são classificadas de acordo com a natureza de seus déficits como apresentado em Dracup et al. (1980a), Beran e Rodier (1985), Wilhite e Glantz (1985), Mirsha e Singh (2010), podendo ser dos seguintes tipos:

1. Secas meteorológicas, definidas como falta de precipitação sobre uma região por um período de tempo;
2. Secas agrícolas, definidas como período de tempo onde a umidade do solo é insuficiente para suportar as culturas agrícolas;

3. Secas hidrológicas, definidas como falta de suprimento de água, onde os recursos hídricos superficiais ou subterrâneos são insuficientes para atender os seus usos; e
4. Secas socioeconômicas, quando os baixos níveis de recursos hídricos afetam a produtividade da sociedade.

Diante da diversidade de formas para se caracterizar uma seca, Yevejevich (1967) apresentou um método baseado no conceito de nível limiar (ou nível de truncamento), denominado de teoria dos runs. A teoria dos runs permite identificar a ocorrência de eventos de seca e quantificar suas principais características. A aplicação da teoria dos runs em uma série temporal de uma variável de interesse qualquer, como por exemplo, as vazões de um rio, permite obter-se uma amostra de eventos de seca, onde cada um é caracterizado por sua duração e severidade (déficit). Sendo a duração, o período de tempo no qual a variável de interesse permanece abaixo do nível limiar, e a severidade, o volume cumulativo da variável correspondente à duração.

Assim, as durações e severidades dos eventos de seca constituem variáveis aleatórias cujos comportamentos probabilísticos podem ser avaliados individualmente (uni-variado) ou conjuntamente (bivariado). Como as durações e severidades dos eventos de seca apresentam correlação significativa entre elas, o par duração-severidade passa a ter destaque nos estudos de seca.

Neste contexto, as funções de acoplamento, ou simplesmente cópulas, introduzidas por Sklar (1959), se apresentam como alternativa para suportar os estudos acerca do comportamento conjunto das durações e severidades dos eventos de seca (análise bivariada). As cópulas são funções que associam as distribuições de probabilidade marginais uni-variadas de duas ou mais variáveis aleatórias, sendo utilizadas como método geral para formulação de sua distribuição de probabilidade conjunta, "codificando" a estrutura de dependência entre elas. As informações pertinentes ao comportamento probabilístico de cada variável estão contidas em suas distribuições de probabilidade marginais uni-variadas, enquanto que a, informação de como essas variáveis se relacionam estão contidas na cópula.

1.1. OBJETIVO

O objetivo deste trabalho é apresentar um método de simulação probabilística de ocorrência de eventos de seca baseado na aplicação de cópulas Gaussiana e Arquimedianas. Como estudo de caso foram utilizadas as vazões médias mensais do Rio Camanducaia, localizado na Unidade de Gerenciamento de Recursos Hídricos Nº 5 (UGRHI 5) do Estado de São Paulo.

O trabalho está dividido em cinco capítulos. No Capítulo 2, é apresentada uma revisão bibliográfica em ordem cronológica trazendo os principais estudos publicados sobre secas hidrológicas ao redor do mundo. No Capítulo 3, são apresentados os materiais e métodos, contendo: a) as principais características da área em estudo; b) o método utilizado para identificação e caracterização dos eventos de secas; c) as distribuições de probabilidade e as funções de acoplamento utilizadas nas análises uni e bivariadas dos eventos, além dos métodos para estimativas de seus parâmetros e critério de seleção dos modelos probabilísticos mais adequados; e d) a formulação do período de retorno conjunto das durações e severidades dos eventos de seca. No Capítulo 4, são apresentados e discutidos os principais resultados no âmbito da análise uni e bivariada dos eventos de seca. O Capítulo 5 finaliza este trabalho, trazendo as principais conclusões e recomendações para estudos posteriores. No Apêndice A, é apresentada a série de vazões médias mensais do Rio Camanducaia de 1944 a 2016, utilizada como estudo de caso, e no Apêndice B, são apresentadas as principais funções do ambiente computacional e linguagem de programação R utilizadas nas análises dos eventos de seca.

2. REVISÃO BIBLIOGRÁFICA

Como será apresentado ao longo da revisão bibliográfica, a partir da publicação do trabalho de Yevejvich (1967), a teoria dos runs permitiu que que vários pesquisadores ao redor do mundo conduzissem estudos uni-variados das durações e severidades dos eventos de seca. Estes estudos foram de grande relevância para consolidar o entendimento do comportamento probabilístico de tais variáveis.

Entretanto, a análise isolada das durações e severidades não revela a significante relação de dependência que existe entre essas variáveis, fornecendo apenas uma avaliação limitada do comportamento dos eventos.

Devido à dependência observada entre as durações e severidades, a partir dos anos 2000, vários pesquisadores passaram a realizar analises conjuntas das principais variáveis dos eventos de secas (duração e severidade). Nos primeiros trabalhos, o comportamento probabilístico conjunto das durações e severidades foi representado por uma distribuição de probabilidade conjunta, que era obtida pelo produto da distribuição de probabilidade condicional das severidades, para uma dada duração, pela distribuição de probabilidade das durações (Shiau e Shen, 2001; Salas et al., 2005).

O inconveniente deste tipo de análise é que, ele necessita de uma amostra significativa de dados observados que nem sempre estão disponíveis no histórico. Assim, para ampliar o tamanho de suas amostras, os pesquisadores utilizaram técnicas de geração de séries temporais.

Neste contexto, a análise de anéis de arvores também foi utilizada como alternativa para ampliar as séries de dados observados. Pesquisadores como González e Valdés (2003) utilizaram a análise dos anéis de árvore para ampliar as séries mensais de Índice Palmer de Severidade de Secas, da Divisão Climática 5 do Texas, Estados Unidos, de 1895 a 2001 para 1569 a 2001.

De Michele e Salvadori (2003), introduzem a utilização das cópulas, apresentadas por Sklar (1959), como ferramenta alternativa para análise conjunta das durações e intensidades dos eventos de seca meteorológica da bacia de drenagem de La Presa, Thyrrhenian Liguria, na região noroeste da Itália. Os autores utilizaram dados horários de precipitação de duas estações meteorológicas de 1990 a 1996.

Recentemente, o uso das cópulas na análise bivariada das durações e severidades dos eventos de seca tem sido proposta em vários estudos. Importantes contribuições sobre a aplicação das cópulas em hidrologia também podem ser encontradas em Salvadori et al. (2007) e Genest e Favre (2007).

Além do trabalho de De Michele e Salvadori (2003), não se pode deixar de citar uma série de trabalhos sobre a análise conjunta das durações e severidades de eventos de seca meteorológicas, que por sua vez também serviram de base conceitual para o estudo de outros tipos de seca, são eles: Shiau (2006); Shiau e Modarres (2009); Mirabbasi et al. (2011); Ganguli e Redy (2012); Ekanayake e Perera (2014); Wang et al. (2017); Fan et al. (2017); Tosunoglua e Onof (2017); She e Xia (2018); e Azan et al. (2018).

Na Figura 2, é apresentada uma linha do tempo contendo os principais trabalhos sobre secas hidrológicas no âmbito das análises uni e bivariada das durações e severidades.

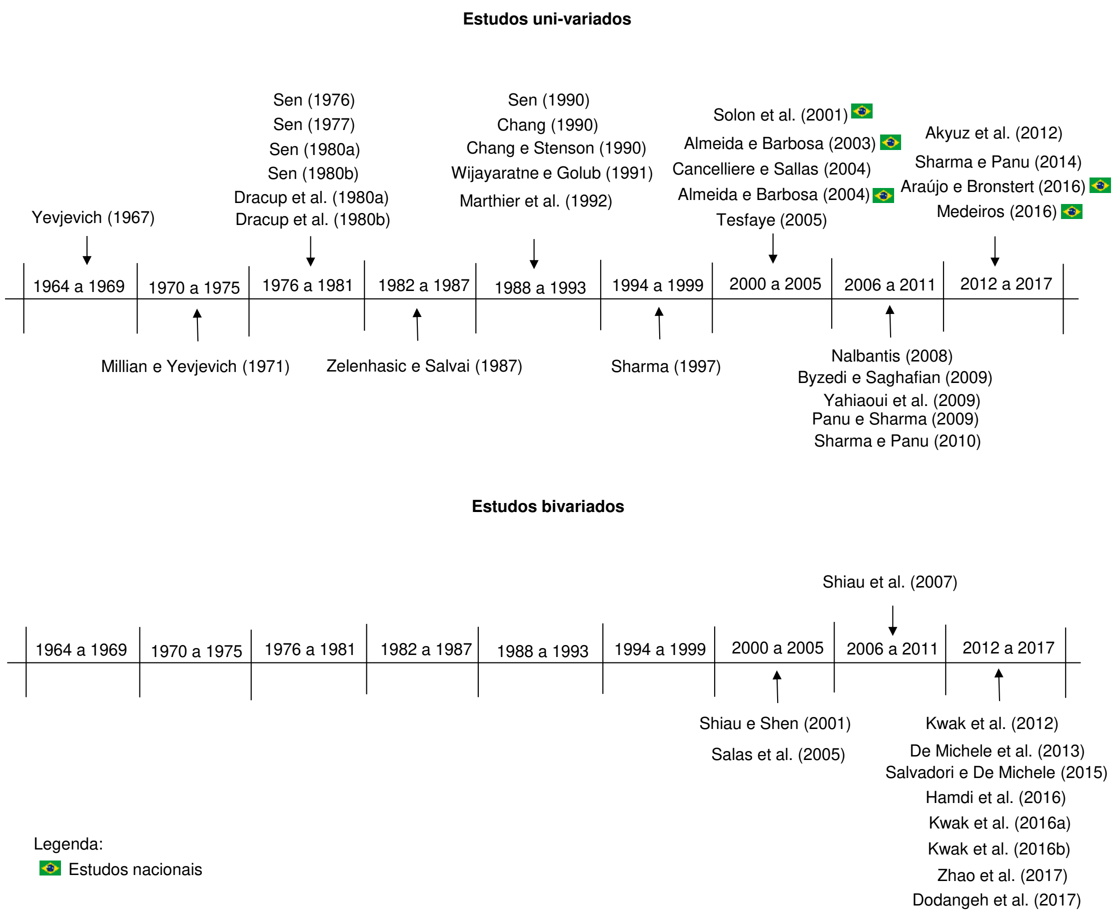


Figura 2. Linha do tempo dos principais eventos de secas hidrológicas

Como pode-se observar na Figura 2, a analise uni-variada dos eventos de seca ainda tem recebido nos últimos anos atenção de alguns pesquisadores ao redor do mundo, com um número significativo de publicações desde o trabalho de Yevjevich (1967).

A Figura 2 mostra que os períodos de cinco anos com maior número de publicações sobre a análise uni-variada dos eventos de seca foram: 1976 a 1981, 1988 a 1993, 2006 a 2011 e 2012 a 2017. A análise bivariada das durações e severidades é relativamente recente e com publicações concentradas entre 2012 e 2017.

Na sequência deste capítulo, os trabalhos citados na Figura 2 serão descritos em ordem cronológica, apresentando suas principais contribuições para os estudos de seca, com o merecido destaque para o trabalho de Yevjevich (1967).

2.1. PRINCIPAL REFERÊNCIA DOS ESTUDOS DE SECA

O estudo publicado por Yevjevich (1967) se tornou a principal referência para os estudos de secas em todo o mundo. Segundo o autor, o primeiro passo para a investigação de qualquer problema é a definição do problema em si. Esse é um dos principais obstáculos na investigação dos eventos de seca, pois os diferentes campos do conhecimento apresentam uma diversidade de formas para caracterizá-los. Por exemplo, os engenheiros vêem as secas do ponto de vista da precipitação, do escoamento superficial e do armazenamento de água. Os economistas veem as secas do ponto de vista das áreas das atividades humanas que são afetadas. Os agricultores possuem outro ponto de vista, para eles a seca está relacionada com a quantidade de água que uma determinada cultura necessita. Em seu trabalho, o autor busca uma “definição objetiva” para as secas. Segundo ele, o termo “definição objetiva” implica que os critérios, as metodologias e as técnicas utilizadas para a definição das secas, mesmo que em diferentes lugares, permitam a mesma interpretação do fenômeno, conduzindo aos mesmos resultados, tendo como referência a mesma base de dados. Com relação às secas hidrológicas, o autor as define como a medida da deficiência do suprimento de água no tempo, no espaço ou em ambos, onde os seguintes parâmetros estão envolvidos:

1. Duração;
2. Área de extensão;
3. Severidade ou déficit; e
4. Probabilidade de recorrência, tempo de recorrência ou período de retorno.

O autor também destaca como sendo o primeiro passo para uma “definição objetiva” a seleção da variável, ou variáveis, que irão representar a seca. A definição de uma sequência como uma sucessão de eventos similares precedidos e sucedidos por eventos diferentes representa o melhor conceito de “definição objetiva” para as secas. As sequências derivadas de uma série temporal de uma variável aleatória $Q = \{q_t, t \in [0, T]\}$ podem ser definidas de várias formas como apresentado na Figura 2.2.1.

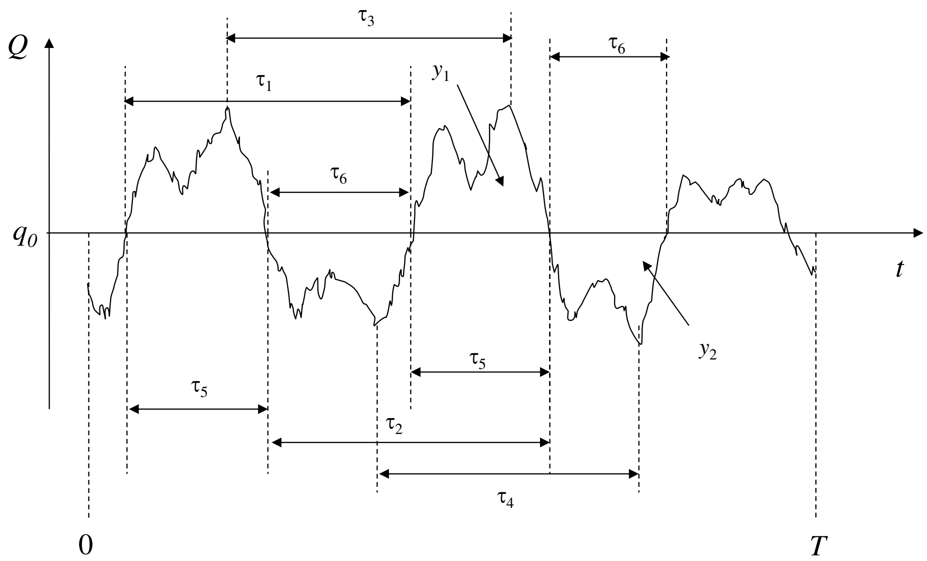


Figura 2.1. Conceito de sequências. Adaptado de Yevjevich (1967).

Selecionando um valor arbitrário q_0, denominado de nivle limiar, valor de referência ou nível de truncamento, a série da variável aleatória será interceptada em várias partes, e a relação entre o valor de referência e todos os outros valores da variável aleatória serve como base para várias definições de sequências. Entretanto, o valor de referência não necessita ser uma variável constante, podendo ser uma função determinística, uma variável aleatória ou combinação de ambas.

Vários conceitos de sequências são utilizados na literatura, sendo alguns deles apresentados a seguir (Yevjevich, 1967):

1. τ_1: distância entre os inícios de derivações positivas;
2. τ_2: distância entre os inícios de derivações negativas;
3. τ_3: distância entre picos sucessivos de derivações positivas;
4. τ_4: distância entre picos sucessivos de derivações negativas;
5. τ_5: distância entre inícios sucessivos de derivações positiva e negativa;
6. τ_6: distância entre inícios sucessivos de derivações negativa e positiva;
7. y_1: soma da derivação positiva entre inícios sucessivos de derivações positiva e negativa; e
8. y_2: soma da derivação negativa entre inícios sucessivos de derivações negativa e positiva.

Segundo Yevjevich (1967), a distância entre os inícios sucessivos de derivações negativa e positiva (τ_6), que representa o comprimento de uma seqüência negativa, pode ser utilizada para representar a duração de uma seca. A soma de uma derivação negativa (y_2) pode ser utilizada para indicar a deficiência do suprimento de água (déficit) ou a severidade da seca. A principal vantagem da utilização das sequências como elemento para a definição de secas é a possibilidade de determinar suas propriedades tendo como base a série histórica de dados ou uma série sintética obtida por um método de geração qualquer. A definição do valor de referência é de grande importância, pois à medida que seu valor é alterado, as variáveis τ_6 e y_2 sofrem mudanças substanciais em seus valores. Segundo o autor, o valor de referência pode ser definido como a média ou a mediana dos valores da variável aleatória, ou até mesmo como um valor que expressa o quanto a média da variável aleatória varia em relação ao seu desvio padrão.

Yevjevich (1967) considerou o comprimento das sequências (duração da seca) como uma variável aleatória independente, sendo seu comportamento probabilístico representado por uma distribuição de probabilidade Geométrica, enquanto que a determinação da soma das sequências é mais complexa e depende da distribuição de probabilidade da variável aleatória analisada.

Em seu trabalho, o autor utilizou o método de Monte Carlo para a geração de três tipos de variáveis aleatórias independentes: uma com distribuição Normal Padronizada e duas com distribuição Log-Normal Padronizada com coeficientes de assimetria 1 e 2, respectivamente. As variáveis foram submetidas a vários valores de referência para a determinação das somas das sequências. Desta forma, o autor pode avaliar os comportamentos referentes à média, variância e o coeficiente de assimetria, os quais foram plotados em função das probabilidades dos valores da variável aleatória serem menores que o valor de referência. O autor também avaliou a distribuição de probabilidade das somas das sequências.

O conceito de sequencias apresentado por Yevjevich (1967) para a identificação e caracterização de eventos de seca ficou conhecido na literatura como teoria dos runs. Além da análise quantitativa, Yevjevich (1967) apresenta em seu trabalho capítulos específicos abordando temas como, explicações gerais sobre as secas, previsão de secas continentais e medidas para combate de secas severas. Com relação às explicações sobre as secas, o autor traz os principais atores, denominados por ele como casuais, que afetam a ocorrência de secas. Dentre esses fatores, o autor destaca aqueles referentes à circulação atmosférica, circulação oceânica e áreas continentais:

1. Circulação atmosférica: as secas dependem das características hidrometeorológicas da região. Como a precipitação e o escoamento superficial resultante dependem da formação de áreas de instabilidade (chuvas), a falta ou o número reduzido dessas formações pode ser o principal fator casual de ocorrência de secas;
2. Circulação oceânica: a distribuição de áreas mais quentes e mais frias no oceano e suas evoluções ao longo do tempo afetam a formação de chuvas. Por exemplo, há uma correlação positiva entre a temperatura média da superfície do oceano no oeste da Califórnia e a precipitação média nas montanhas de Sierra Nevada. As águas frias da plataforma do Atlântico, persistentemente localizadas na costa

de Nova York e Nova Inglaterra nos últimos seis anos, podem explicar as secas da costa nordeste dos Estados Unidos. No caso do Brasil, podem ser citados os efeitos dos fenômenos El Ninõ e La Ninã que afetam o regime de chuvas e, consequentemente, a ocorrência dos eventos de secas. No fenômeno El Ninõ, há um aquecimento das águas do oceano Pacífico provocando, de forma geral, aumento das precipitações na região sul e em partes nas regiões sudeste e centro-oeste, com episódios de secas nas regiões nordeste e norte. No caso da La Ninã, ocorre um resfriamento das águas do oceano Pacífico provocando efeitos contrários ao do El Ninõ; e

3. Áreas continentais: similarmente ao que ocorre na circulação oceânica, áreas continentais muito frias ou muito quentes podem afetar a ocorrência de secas em áreas continentais adjacentes. Uma correlação negativa é frequentemente observada entre precipitação e temperatura do verão de áreas frias, pode ser responsável pela perda ou pelo um avanço substancial das massas de neve e de gelo nas regiões polares.

No contexto da previsão de secas continentais, Yevjevich (1967) enfatiza a complexidade do fenômeno. Segundo o autor, a característica aleatória predominante das séries anuais de precipitação e vazão não fornece uma perspectiva promissora para a previsão do fenômeno. A previsão que o autor se refere em seu trabalho é a utilização de padrões de observações passadas das séries temporais para definir unicamente (e não em termos probabilísticos) o futuro. De acordo com o autor, uma boa aproximação para a previsão de secas continentais deve incorporar informações das condições oceânicas e da circulação atmosférica.

Para as medidas de combate a secas severas, Yevjevich (1967) aborda em seu trabalho alternativas como a redução do consumo de água, a implementação de planos de contingência e aumento das linhas de suprimento por meio de estruturas temporárias ou transferência de água entre regiões. A redução do consumo é uma alternativa de fácil implementação, porém menos atrativa devido às suas implicações (repercussões) em termos políticos no que diz respeito à pressão social. Os planos de contingência não são relativamente caros, pois pequenos investimentos realizados aos longos dos anos nos sistemas de abastecimento de água podem reduzir de forma substancial os efeitos de uma seca. A constituição de fundos para financiamento de estudos de cheias e secas, para o planejamento e até para o combate de seus efeitos é muito importante, pois na maioria das vezes as soluções para os problemas estarão disponíveis na ocorrência de eventos mais críticos. Geralmente, cortes nestes fundos, quando eles existem, são proporcionais ao tempo decorrido a partir da ocorrência de eventos extremos. Outro ponto de atenção levantado pelo autor é que os planos de contingência são raramente finalizados e, se completados, a recomendação é que sejam implementados de forma lenta e gradual. Infelizmente, a realidade é que, na maioria dos casos, regiões, cidades, grupo de cidades ou até mesmo estados não possuem planos de contingência para combate às secas. Soluções improvisadas para o abastecimento de água a partir das fontes disponíveis na região atingidas por uma seca são mais uma regra do que uma exceção. É fato que qualquer solução improvisada terá um custo mais alto ou menor rendimento do que qualquer medida estruturada em planos de contingência para combate às ecas.

Por fim, Yevjech (1967) aborda em seu trabalho transferência de água a longas distâncias entre regiões como forma de combate a secas severas, fazendo um paralelo aos sistemas de transmissão de energia elétrica.

2.2. ANALISES UNI-VARIADAS DOS EVENTOS DE SECA

Millian e Yevjevich (1971) estudaram as probabilidades de ocorrência dos eventos históricos de seca utilizando a teoria dos runs, tendo como base as definições do maior comprimento (duração) e da maior soma das sequências negativas (déficit). Foram estudadas 10 séries de vazões médias anuais, quatro para rios dos Estados Unidos e as demais para rios da Colômbia, Laos, Suíça, Romênia, Suécia e União Soviética. As séries de vazões média anuais tinham duração de 22 a 150 anos. Os autores também estudaram 20 séries de precipitações médias anuais dos Estados Unidos com durações de 30 a 79 anos. Os autores avaliaram quatro níveis de truncamento para identificação e caracterização dos eventos de seca. Também foram geradas séries anuais com cinco tamanhos diferentes (25, 50 100, 200 e 500) para variáveis com distribuição Normal e Log-Normal Independente, e um processo estocástico estacionário dependente, o qual seguia um modelo Auto Regressivo de 1ª ordem, ou AR(1). Os autores concluíram que, para as 10 séries de vazões médias anuais e 20 séries de precipitações médias anuais, os períodos de retorno dos maiores comprimentos e das maiores somas das sequências negativas são praticamente os mesmos para um determinado nível de truncamento.

Sen (1976) propôs uma solução analítica simples para problemas relativos às probabilidades e propriedades de períodos úmidos e secos de séries anuais de vazão. A metodologia proposta pelo autor tinha como objetivo principal encontrar a probabilidade e as propriedades estatísticas dos runs. O autor estudou o comportamento da média dos comprimentos (durações) das sequencias positivas e negativas, como função do nível de truncamento, bem como da estrutura de correlação da série de dados, utilizando um modelo de Markov de lag 1. Foram utilizados cinco níveis de truncamento e coeficientes de correlação serial de lag 1 variando de 0 a 0,9. Como estudo de caso, foram utilizados dados de vazões do Rio Cine, na Turquia, com uma série de 32 anos de observações O autor concluiu que, para um processo AR(1), as propriedades estatísticas dos runs são funções do coeficiente de correlação serial de primeira ordem da série de dados.

Sen (1977), em um trabalho semelhante ao realizado em 1976, avaliou o comportamento da soma das sequencias (déficits, para os casos das secas e excessos, para o caso das cheias) considerando um processo independente Log-Normal e um processo Normal de Markov de lag 1. O autor concluiu que a análise pode ser aplicada para processos dependentes e independentes com distribuições Normal ou não, e que a variância e a média da soma das sequencias aumenta com o aumento do coeficiente de correlação entre observações sucessivas.

Sen (1980a) estudou o comportamento das durações críticas de secas hidrológicas para um processo estocástico periódico. A metodologia foi aplica a quatro rios da Turquia, com registros de vazões variando de 31 a 34 anos, e bacias hidrográficas com áreas de drenagem de 928 a 63.874 km^2. Os rios utilizados pelo autor foram os rios B. Menderes, Manavgat, Kizilirmak e Firat. O autor comenta que a estrutura periódica definida pela transição de probabilidades entre estados (seco e não seco) reflete no valor esperado da maior duração das secas. O autor também comentou que em amostras menores a periodicidade das secas é mais acentuada que em amostras maiores. À medida que o tamanho da amostra aumenta, a periodicidade diminui e se aproxima de um valor assintótico. O autor demostra este fato apresentando um gráfico onde são mostrados os valores esperados da maior duração das secas e comprimento da amostra de dados.

Sen (1980b) estudou o comportamento das durações e severidades críticas de secas para quatro rios ao longo do mundo, dois nos Estados Unidos (Mississipi e Missouri), um na Suíça (Rhine) e outro na Romênia (Danube), tendo como base suas séries de vazões médias anuais com registros variando de 65 a 150 anos. A metodologia apesentada foi baseada em processos independente e de Markov de lag 1. O autor avaliou vários níveis de truncamento e observou que, para as durações críticas, a metodologia representou razoavelmente os dados históricos. Para os déficits críticos, os autores apresentaram os resultados apenas para dois rios (Suíça e Romênia), concluindo que a metodologia não reproduz adequadamente os dados observados.

Dracup et al. (1980a) apresentaram várias considerações a respeito dos eventos de secas. Tais considerações incluem:

1. A seleção da natureza dos déficits de água a serem estudados (hidrológica, meteorológica ou agrícola);
2. A seleção do período médio de tempo utilizado para discretizar a série temporal de dados (mensal, semanal ou anual);
3. A seleção do nível de truncamento a ser utilizado para obter os eventos de secas (média ou mediana); e
4. Um método de regionalização ou padronização das secas.

Com relação ao primeiro item acima, os autores mencionam que a decisão inicial a ser tomada é a definição da natureza do déficit a ser estudado. Tais déficits podem ser representados em termos da precipitação (secas meteorológicas), da vazão (secas hidrológicas), da umidade do solo (secas agrícolas) ou qualquer combinação das três variáveis. Claramente, a natureza do ciclo hidrológico indica que períodos onde ocorrem baixos valores de precipitação, vazão e umidade do solo estão intimamente relacionados. Sendo a precipitação uma das variáveis de entrada do ciclo hidrológico e a vazão uma das variáveis de saída, se o interesse é determinar a causa dos eventos de secas, a atenção deve ser focada na precipitação. Caso o interesse seja na determinação dos efeitos e dos impactos da seca, a atenção deve ser focada nas secas hidrológicas ou agrícolas. Depois de selecionada a natureza do déficit de água, deve-se decidir qual unidade de tempo será utilizada como período médio para discretizar as variáveis meteorológicas ou hidrológicas de interesse. A seleção de um período médio de tempo depende basicamente dos propósitos do estudo. Entretanto, deve-se lembrar que a escolha do período médio de tempo afeta o estudo das secas com relação a dois aspectos:

1. Definição do tamanho da amostra de eventos; e
2. Grau de correlação entre eles.

Dracup et al. (1980b) caracterizaram estatisticamente os eventos de secas hidrológicas e de altas vazões em vários postos fluviométricos da Califórnia. Os autores utilizaram como base para seu estudo as vazões médias anuais históricas. Os eventos de secas hidrológicas e de altas vazões foram obtidos utilizando-se um nível de truncamento correspondente à vazão média anual de longo termo de cada posto fluviométrico. Além da severidade e das durações, os autores também quantificaram a magnitude dos eventos como sendo a relação entre suas severidades e durações.

Segundo Dracup et al. (1980b), a severidade e a duração são consideradas como parâmetros primários, pois dependem diretamente das vazões, enquanto que a magnitude é considerada como um parâmetro secundário, pois depende da severidade e da duração. Em seu estudo, os autores realizaram as seguintes análises estatísticas para a severidade, duração e magnitude:

1. Estacionariedade;
2. Aleatoriedade, representada pela dependência serial de ordem 1;
3. Correlação entre os parâmetros; e
4. Correlação cruzada entre os parâmetros de eventos sucessivos de secas hidrológicas e de altas vazões.

Dentre os pontos mais relevantes, os autores observaram que os resultados dos diagramas de dispersão obtidos na análise de aleatoriedade forneceram uma indicação excelente da máxima resposta de cada bacia hidrográfica em termos de severidade e duração para os períodos analisados, sendo esta indicação muito significante para o estudo de operação de reservatórios quando comparada com a tradicional metodologia do período crítico. Com relação à correlação cruzada entre eventos sucessivos de secas e altas vazões, foi observada independência entre ambos. Os autores também citam que a avaliação do comportamento estatístico dos parâmetros dos eventos de secas hidrológicas plurianuais representa um passo essencial para a implementação de uma análise de frequências, e na concepção de modelos estocásticos de geração de séries de eventos.

Zelenhasic e Salvai (1987) estudaram o comportamento dos eventos de secas hidrológicas no Rio Sava em Sr. Mitrovica, Iuguslávia, cuja área de drenagem abrange 87.996 km^2. Os eventos de secas hidrológicas foram obtidos para um nível de truncamento representado pela vazão que possui 90% de probabilidade de ser superada, percentil 90, aplicado sob o período de registros de vazões semanais compreendido entre os anos de 1926 a 1983. A independência das durações e dos déficits dos eventos, observados em série, foi constatada para um nível de significância de 95% pelas funções de autocorrelação. A distribuição de probabilidade Exponencial foi utilizada para representar o comportamento das severidades e das durações dos eventos, sendo que tal hipótese não pode ser rejeitada para um nível de significância de 95% segundo o teste de Kolmogorov-Smirnov. Os autores também utilizaram um modelo de Poison para representar a ocorrência dos eventos de secas para cinco períodos de tempo distintos, 180, 210, 240, 310 e 365 dias. A aderência do modelo de Poison foi verificada pelo teste do qui-quadrado para um nível de significância de 95%. Com base nas distribuições de probabilidade das severidades e das durações, e no modelo de Poison, os autores obtiveram distribuições de probabilidade condicional teóricas para representar o máximo déficit e a máxima duração que podem ocorrer em um intervalo de tempo de um ano. Os modelos teóricos obtidos foram validados pelo teste de Kolmogorov-Smirnov para um nível de significância de 95%.

Sen (1990) utilizou um modelo de cadeia de Markov de 2ª ordem para representar as durações críticas dos eventos de secas hidrológicas de três rios de diferentes partes do mundo: Rio Misissippi, nos Estados Unidos; Rio Rhine, na Romênia; e Rio Danube, na Suiça. O estudo teve como base as séries de vazões médias anuais com registros variando de 96 a 150 anos. O nível de truncamento considerado foi a média das vazões anuais. Em suas conclusões, o autor comenta que a probabilidade da maior duração de uma seca depende do tamanho da amostra, nível de truncamento e da transição de probabilidade entre estados (seco e não seco), mas é independente da distribuição de probabilidade do fenômeno hidrológico.

Chang (1990) utilizou em seu estudo o conceito de intensidade dos eventos de secas, expresso pelo autor como o número esperado de eventos de secas que podem ocorrer em um determinado intervalo de tempo. Em seu estudo, foi utilizada a série de vazões médias diárias da estação fluviométrica Higby do Rio Scioto (tributário do Rio Ohaio) para o período compreendido entre os anos de 1979 a 1986, cuja área de drenagem é de 13.289 km^2. Os níveis de truncamento utilizados no estudo correspondem aos percentis 50, 70, 80 e 90. Como conclusão, o autor identificou uma elevação das intensidades dos eventos de seca hidrológica na região.

Chang e Stenson (1990) discutem as secas hidrológicas na forma de vazões mínimas para 18 estações fluviométricas na bacia hidrográfica do Rio Scioto em Ohio nos Estados Unidos. As estações fluviométricas analisadas possuem áreas de drenagem de 316 a 13.283 km^2. Os autores analisaram primeiramente as vazões mínimas anuais para períodos de retorno de 2, 5, 10, 20, 50 e 100 anos. Os resultados obtidos indicaram que, para os rios analisados, considerando um período de retorno de 100 anos, as secas, quando representadas pelas vazões mínimas anuais, são praticamente iguais a zero. Na sequência os autores avaliaram a vazão para definir um evento de seca, considerando como nível de truncamento percentis das séries de vazões médias diárias. Os autores utilizaram níveis de truncamento correspondentes a vários percentis, variando do 10 ao 95. Os autores observaram que médias diárias com percentil 95 são aproximadamente iguais às vazões mínimas anuais com período de retorno de dois anos. Para os níveis de truncamento equivalentes aos percentis 70, 80, 90 e 95, os autores avaliaram os coeficientes de correlação entre as durações e severidades dos eventos de secas, observando uma forte correlação entre eles, mas que o coeficiente de correlação diminui a medida que se eleva o nível de truncamento.

Wijayaratne e Golub (1991) realizaram um estudo sobre secas hidrológicas utilizando séries sintéticas de vazões médias anuais. Como estudo de caso, os autores utilizaram o Rio Pequest em New Jersey, cuja série de dados de vazões históricas abrangia os anos de 1922 a 1985. A metodologia utilizada para a geração das séries sintéticas de vazões médias anuais foi baseada em um modelo híbrido, que representava a parte determinística com base em uma análise de harmônicos e a parte estocástica com base em um modelo auto regressivo de média móvel. Os autores geraram vários comprimentos de séries e observaram que aquelas com 240 eventos de vazões médias anuais eram estatisticamente próximas da série histórica de dados. Foram geradas séries com comprimentos de 60, 120, 180 e 240 anos. Para a definição dos eventos de seca, seis níveis de truncamento foram aplicados à série histórica de vazões. Os níveis de truncamento variaram de 85% a 110% da vazão média anual, com intervalos de 5%. Para a seleção do nível de truncamento mais conveniente foram utilizados dois critérios:

1. Para um dado nível de truncamento, a média das durações deveria ser aproximadamente igual a média da duração dos excessos; e
2. A duração média não poderia ser significantemente diferente da semi-periodicidade do ciclo hidrológico (obtida com a análise dos harmônicos da série de dados históricos).

Como resultado, os autores observaram que um nível de truncamento de 95% da vazão média anual atendia os critérios estabelecidos pelos autores. Com este nível de truncamento, foram obtidas as principais estatísticas para as durações e as severidades dos eventos de seca hidrológica. Os autores observaram que as estatísticas associadas à duração e à severidade, obtidas com base nas séries sintéticas, apresentaram um comportamento próximo ao histórico, e que as discrepâncias observadas estavam associadas aos tamanhos das séries geradas. Quanto ao comportamento probabilístico, foram ajustadas distribuições de probabilidades do tipo Gama e Log-Normal, ambas com três parâmetros, para representar as durações e as severidades dos eventos de secas hidrológicas. Tais distribuições foram ajustadas tendo como referência os eventos gerados a partir da série histórica e da série sintética com duração de 240 anos. Os autores observaram que as curvas de probabilidade derivadas dos eventos gerados a partir da série sintética abrangiam uma maior variação do que aquela derivada dos dados históricos. Desta forma, as curvas de probabilidades derivadas das séries sintéticas são mais confiáveis para a previsão de eventos extremos, tais como, secas de maior duração e maior severidade para um dado período de retorno. Em suas conclusões, os autores citam que a metodologia apresentada pode ser estendida para diferentes bacias hidrográficas, possibilitando também a análise espacial das variações das características das secas hidrológicas, podendo ser de grande valia no planejamento futuro dos recursos hídricos.

Marthier et al. (1992) apresentaram uma análise estatística das durações e severidade dos eventos de seca hidrológica utilizando as distribuições de probabilidade Geométrica e Exponencial. Como estudo de caso, os autores utilizaram uma base de dados composta por vazões médias mensais compreendidas entre os anos de 1943 e 1989, correspondentes à área de drenagem de uma bacia hidrográfica da companhia hidroelétrica Alcan em Saguenay-Lac-Saint-Jean, em Québec no Canadá. Para caracterizar os eventos de seca hidrológica, os autores utilizaram um nível de truncamento igual ao percentil 35. O nível de truncamento foi calculado para os meses de janeiro a dezembro tendo como referência a série histórica de dados. A aplicação do nível de truncamento na série histórica de vazões médias mensais gerou dois conjuntos de dados, o das durações e o das severidades, cuja hipótese de independência das observações em série não pode ser rejeitada segundo o teste de Wald-Wolfwitz para um nível de significância de 95%. O teste de Grubbs e Beck também foi aplicado para verificar a presença de outliers. Para um nível de significância de 95% a hipótese da não presença de outliers não pode ser rejeitada. A análise visual do histograma de frequências das durações indicou que uma distribuição de probabilidade Geométrica poderia ser utilizada para representar tais dados. Tal hipótese de aderência não pôde ser rejeitada pelo teste do qui-quadrado para um nível de significância de 95%.

Para as severidades, os autores ajustaram uma distribuição de probabilidade Exponencial, cuja hipótese de aderência não pôde ser rejeitada pelo teste de Komogorov-Smirnov para um nível de significância de 95%. Os parâmetros das distribuições de probabilidade foram obtidos pelo Método da Máxima Verossimilhança. Os autores também citam que se houver a necessidade de uma melhor precisão para representar as severidades, distribuições de probabilidade com 2 ou 3 parâmetros podem ser apropriadas, como por exemplo, a Gama com 2 parâmetros ou 3 parâmetros. Entretanto, os autores alertam para o fato de que a estimativa de mais parâmetros pode implicar no aumento da variabilidade da variável estimada. Os autores concluíram que a metodologia apresentada pode ser utilizada para se ter um melhor conhecimento entre os eventos de seca hidrológica e as falhas energéticas em um contexto econômico, destacando também que são necessárias novas pesquisas para estudar a influência de diferentes níveis de truncamento nas durações e nas severidades das secas hidrológicas.

Sharma (1997) avaliou o comportamento das severidades e das durações dos eventos de seca hidrológica nos Rios Tana e Sondu, no Kenia, e no Rio Victoria Nile em Jinja, Uganda, com dados variando de 38 a 61 anos. Em sua análise, o autor trabalhou com as séries de vazões médias anuais normalizadas, onde as mesmas foram submetidas a seis níveis de truncamento. Tendo como referência os dados observados, um modelo teórico baseado em uma distribuição de probabilidade Normal pôde ser validado. Desta forma, o autor representou o comportamento dos valores esperados das durações e das severidades dos eventos de secas hidrológicas para seis níveis de truncamento e seis períodos de retorno: 10, 25, 50, 100, 200 e 500 anos. Em suas conclusões, os autores comentam que as severidades das secas diminuem a medida que diminui o nível de truncamento para um mesmo período de retorno. Esse mesmo comportamento foi observado para as durações.

Solon et al. (2001) estimaram o valor esperado da duração máxima de seca hidrológica no Açude Castanhão, localizado no Estado do Ceará, no Nordeste brasileiro. O Açude Castanhão é um reservatório formado pelo barramento do Rio Jaguaribe (considerado o maior rio intermitente do mundo), no chamado Boqueirão do Cunha, situado à cerca de 5,9 km ao sul do povoado Castanhão. As durações das secas foram determinadas pela teoria dos runs. No estudo, os autores assumiram um nível de truncamento constante e igual ao da vazão regularizada açude para um dado nível de garantia e para um dado horizonte de simulação, tanto anual como mensal. Para simular a operação do reservatório, foi gerada uma série sintética anual de 5.000 anos para o Rio Jaguaribe, desagregada em valores mensais. A série de 5.000 anos foi subdividida em 50 séries de 20 anos, 50 séries de 30 anos, 50 séries de 40 anos, 50 séries de 50 anos, 50 séries de 70 anos e 50 séries de 100 anos. O valor adotado para a vazão regularizada (nível de truncamento) foi a média dos 50 valores de vazões regularizadas obtidas pela operação simulada do reservatório para cada cenário de nível de garantia e horizonte de simulação. Os autores concluíram que, ao considerar-se níveis de garantia mensais, deve-se esperar valores maiores para as durações máximas das secas do que quando o reservatório é simulado com garantia anual. Os valores obtidos com garantias mensais no abastecimento oferecem uma maior confiabilidade que os valores com garantia anual, uma vez que a dispersão, traduzida pelo coeficiente de variação das séries de duração máxima das secas hidrológicas, são significativamente menores. Também observaram uma tendência de diminuição desta dispersão, para um mesmo nível de garantia, à medida em que se aumenta o horizonte de simulação. Este fato era esperado, uma vez que o processo vai cada vez menos sendo influenciado pelas condições iniciais assumidas para o reservatório, tendendo ao estado de equilíbrio.

Almeida e Barbosa (2003) apresentaram um estudo sobre o comportamento dos eventos de seca hidrológica para os Rios Piracicaba, Corumbataí, Camanducaia, Jaguari e Atibaia, localizados no Estado de São Paulo, Brasil. No estudo foram utilizadas as séries de vazões médias mensais compreendidas entre os anos de 1929 e 1999. Como nível de truncamento foram utilizadas projeções da demanda total de abastecimento urbano, industrial e rural para cada rio, para o horizonte dos anos 2000, 2005, 2010 e 2020. A análise baseada nos dados históricos indicou uma maior vulnerabilidade para a bacia hidrográfica do Rio Atibaia, na qual se observa um maior número de eventos de seca. Também foi observado pelos autores que, as durações e severidades do Rio Atibaia possuem elevado coeficiente de correlação de Pearson. Os autores também citam que, 90% dos eventos de seca observados a partir do histórico de vazões médias mensais possui durações inferiores a 20 dias. Fato que foi constatado pelas distribuições de probabilidade empírica dos déficits. Como conclusões, é ressaltada a importância de se entender a estrutura do comportamento dos eventos de seca, como informação adicional para o desenvolvimento de projetos e/ou programas de mitigação de impactos. Como complemento, os autores citam que não se deve avaliar o comportamento das secas somente com base no histórico de vazões observadas, pois este é somente uma das possíveis sequências sorteadas pela natureza. Assim, a utilização de técnicas como a geração de séries sintéticas pode fornecer um melhor entendimento do fenômeno.

Almeida e Barbosa (2004) realizaram um estudo sobre eventos de seca hidrológica no Rio Atibaia, localizado no Estado de São Paulo, Brasil, tendo como base os resultados apresentados em Almeida e Barbosa (2003). Os eventos foram avaliados para cinco níveis de truncamento, representados pelos percentis 25, 50 (mediana), 75, 90 e 99 da curva de vazões médias mensais. A bacia hidrográfica do Rio Atibaia possui área de aproximadamente 2.820 km^2. Os dados de vazões utilizados no estudo foram da estação fluviométrica 3D-006 localizada no município de Itatiba, Estado de São Paulo, compreendendo um período de 1929 a 2002. Os autores observaram um número crescente de eventos de seca a medida que se eleva o nível de truncamento, o que era esperado, sendo que, para o nível de truncamento correspondente o percentil 50, se atinge o maior número de eventos (ponto crítico de maior vulnerabilidade). Também foi constatada uma forte correlação de Pearson entre as durações e os déficits dos eventos de seca, como apresentada em Almeida e Barbosa (2003). Tendo como base tal comportamento, os autores propuseram a representação dos déficits por meio de uma equação linear dependente das durações, generalizada para cada nível de truncamento estudado. O comportamento probabilístico dos déficits, para cada nível de truncamento, foi representado por uma distribuição de probabilidade Gama com dois parâmetros, onde os mesmos foram estimados minimizando-se o somatório do erro quadrático entre as frequências observadas e teóricas. A hipótese de aderência da distribuição de probabilidade Gama foi verificada pelo teste de Kolmogorov-Smirnov para um nível de significância de 95%. Os autores propuseram uma generalização do comportamento dos parâmetros da distribuição de probabilidade Gama, permitindo a obtenção de uma distribuição de probabilidade genérica aplicável para os níveis de truncamento compreendidos entre os percentis 25 a 99 das vazões médias mensais. Os autores também avaliaram a representação da ocorrência anual de eventos de seca com a aplicação de uma distribuição de probabilidade de Poison, mas tal hipótese foi rejeitada.

Cancelliere e Sallas (2004) realizam um estudo das durações dos eventos de secas, tendo como base os dados de vazões médias mensais do Rio Niger em Koulikoro de 1907 a 1999. Os autores também ampliaram seu estudo com dados de precipitações médias mensais de Caltanissette, na Itália ,de 1879 a 1996, e Fort Colins, para a estação da Universidade de Colorado de 1901 a 1999. Também foram utilizadas as séries mensais do Índice de Precipitação Padronizada de 1896 a 2001, e de Índice Palmer de Seca de 1985 a 2001, ambas para o nordeste do Colorado. Os eventos de seca foram obtidos utilizando-se níveis de truncamento específicos para cada tipo de série de dados. A probabilidade das durações dos eventos foi obtida com base em um processo de Markov de 1ª ordem. Os autores utilizaram a metodologia apresentada em Shiau e Shen (2001) para obtenção do período de retorno das durações dos eventos. Os autores também avaliaram o efeito da suposição de independência e dependência entre os eventos na estimativa das probabilidades de ocorrência e dos períodos de retorno. Os autores verificaram que, para o conjunto de dados utilizados no estudo, os períodos de retorno obtidos com base na suposição de independência dos eventos conduzem a valores superiores àqueles baseados na suposição de dependência.

Tesfaye (2005) estudou o comportamento dos eventos de seca hidrológica no Rio Fraser próximo a Hope em British Columbia. O autor utilizou uma série histórica de dados de vazões médias semanais de outubro de 1912 a setembro de 1984. Para definir os eventos de seca hidrológica foi utilizada uma curva de demanda senoidal. Os autores avaliaram a duração máxima, a severidade máxima e o período de retorno das durações e severidades dos eventos. Uma análise confiável dos parâmetros que caracterizam os eventos de seca hidrológica, baseada na série de dados históricos, pode não ser alcançada devido ao número limitado de eventos. Desta forma o autor utilizou um modelo periódico auto regressivo de média móvel para gerar uma série sintética de vazões médias semanais com 1.000 anos de duração, obtendo, desta forma, um número maior de eventos. Com base nos eventos obtidos a partir da série sintética, foram determinados, por exemplo, os períodos de retorno para a máxima duração e a máxima severidade de 385 e 333 anos, respectivamente. Tais valores são bem diferentes daqueles obtidos com base na série histórica, seis e doze anos, respectivamente. Em suas conclusões, o autor cita que se deve realizar uma investigação detalhada das distribuições de probabilidade das durações e das severidades dos eventos de seca hidrológica, tanto para os dados históricos como para a série sintética, assim como uma análise da sensibilidade dos parâmetros que caracterizam os eventos em relação ao número de anos da série sintética.

Nalbantis (2008) apresentou uma metodologia baseada no Índice de Vazão de Secas (SDI) para caracterizar as secas hidrológicas em duas bacias hidrográficas da Grécia que abastecem a área metropolitana de Atenas, a bacia hidrográfica do Rio Evinos e a bacia hidrográfica do Rio Boeoticos Kephisos. O SDI é calculado para períodos de referência sobrepostos dentro de cada ano hidrológico, permitindo assim a categorização de secas num pequeno número de categorias, por exemplo: períodos sem secas, secas suaves, secas moderadas, secas severas e secas extremas, onde cada período possui uma probabilidade associada. Os autores concluíram que a metodologia pode ser facilmente implementada, mas requer dados de qualidade e com comprimento suficiente para estimar com precisão as frequências de fenômenos raros de seca.

Byzedi e Saghafian (2009) estudaram o comportamento das secas no sudoeste do Iran. Como estudo de caso, os autores utilizaram 54 estações fluviométricas da Bacia hidrográfica de Karoon, Dez e Karkhe. Os eventos de seca foram definidos pela teoria dos runs com um nível de truncamento referente à vazão do percentil 70. Para as severidades e durações dos eventos de seca, foram avaliadas distribuições de probabilidade Gama, Weibull, Log-Normal, Jonshon, Gumbel e Pareto Generalizada, cujos parâmetros foram estimados pelo Método da Máximo Verossimilhança. O teste do qui-quadrado foi utilizado para verificar a hipótese de adequação das distribuições aos dados observados. Das 54 estações fluviométricas, cinco foram descartadas, pois a hipótese de adequação dos modelos probabilísticos uni-variados foi rejeitada. A análise da frequência teve como base os valores máximos anuais de duração e severidade. Os autores realizaram uma análise regional das durações e severidades com dois anos de período de retorno. Os autores concluiram que a análise regional não indicou relações adequadas para as durações (regiões de comportamento homogêneo). Assim, somente para as severidades, os autores estabeleceram regressões regionais para as secas de dois anos. A análise regional do trabalho foi realizada com três regiões distintas, onde as severidades das secas com dois anos de período de retorno estão relacionadas com a área (km^2), com a densidade da rede de drenagem (km/km^2) e o total de precipitação (mm) de dezembro a fevereiro.

Yahiaoui et al. (2009) analisaram as frequências dos eventos de seca hidrológica na captação de Oued Mina no oeste da Argélia, com área de drenagem de 4.285 km^2. Para o estudo, foi utilizada uma série de vazões médias diárias de 19 anos. Para a identificação dos eventos de secas hidrológicas foi utilizado um nível de truncamento igual ao percentil 70. Para as durações e as severidades dos eventos de secas hidrológicas observados foram ajustadas as distribuições de probabilidades do tipo Pareto, Weibull e Log-Normal. Os autores observaram que a distribuição de probabilidade Log-Normal foi a melhor escolha para estimar as magnitudes das durações e dos eventos de secas hidrológicas. O ajuste das distribuições de probabilidade foi medido pelo coeficiente de correlação, pela raiz quadrada dos Erros Médios Quadráticos e pelo teste de Kolmogorov-Smirnov. Também foram calculadas, para vários períodos de retorno, as durações máximas e os volumes máximos como apresentado em Zelenhasic e Salvai (1987). Em função do número reduzido de eventos de secas hidrológicas os autores não validaram os valores teóricos das máximas durações e dos máximos volumes. Em suas conclusões, os autores destacam que as distribuições de probabilidade baseadas somente nas observações históricas podem conter incertezas.

Panu e Sharma (2009) estudaram o comportamento de dois parâmetros críticos das secas hidrológicas, a maior duração e a maior severidade para um determinado período de tempo. O estudo foi realizado para 25 rios do nordeste de Ontario, Canadá, utilizando as séries de vazões médias anuais, cujos valores podiam ser considerados independentes e obedecendo uma distribuição de probabilidade Normal com um nível de significância de 95%. No estudo, os autores compararam as durações máximas e as severidades máximas obtidas com base em equações do trabalho de Millan e Yevjevich (1971) com uma aproximação probabilística de tais parâmetros. A conclusão foi de que a aproximação probabilística foi marginalmente melhor. Os autores também apresentam uma avaliação espacial das secas hidrológicas na região nordeste de Ontario com base no Índice Potencial de Secas (DPI) de Horn (1989). Valores elevados de DPI indicam que a região é mais propensa a secas. Os autores observaram pelo DPI maior propensão a secas ao longo da fronteira de Ontario-Manitoba.

Sharma e Panu (2010) estudaram o comportamento dos eventos de seca hidrológica para 18 rios do Canadá, cujas bacias hidrográficas variam de 37 a 32.400 km^2, com períodos de dados variando de 1919 a 2005 os quais necessitaram de estimativas (preenchimento de falhas). As vazões médias diárias foram transformadas em vazões médias semanais. Os autores observaram que uma distribuição de probabilidade Gama com dois parâmetros se adequava bem aos dados de vazões médias semanais. As séries de vazões médias semanais são padronizadas, gerando uma série de vazões médias semanais estacionária com média zero e variância igual a um. Desta forma, os autores apresentam o conceito de Índice Hidrológico Padronizado (SHI), obtido com base na série histórica de vazões médias semanais padronizadas. O SHI é uma série de valores padronizados com média zero e variância igual a um, mas não normalmente distribuído. O nível de truncamento da série padronizada utilizado no estudo foi a mediana. Os autores estudaram então os comportamentos das durações médias dos eventos. Os valores esperados das durações foram obtidos com base no teorema dos números extremos (Sen, 1980b) e comparados com os valores observados. Segundo os autores, os resultados obtidos com o teorema dos números extremos não foram satisfatórios devido à forte estrutura de correlação da série de SHI. Desta forma, os autores propuseram a utilização de um modelo de cadeia de Markov para representar os valores esperados das durações que conduziu a resultados satisfatórios.

Akyuz et al. (2012) estudaram o comportamento das durações e severidades dos eventos de seca hidrológica de quatro rios, um em Nova Iorque, um na Suíça e dois na Turquia. O estudo teve como base as séries de vazões médias anuais, com períodos de dados variando entre 68 e 150 anos, com elevada auto correlação de ordem 1. O nível de truncamento das séries utilizado no estudo foi a mediana das vazões médias anuais. No estudo, os autores utilizaram um modelo de cadeia de Markov de 1ª e 2ª ordem, comparando os resultados obtidos com os de um modelo AR(1), que gerou uma série de vazões médias anuais de três milhões de ano.

Os autores concluíram que o modelo de cadeia de Markov de 1ª ordem não é adequado para séries com alta correlação, sendo melhor desempenho apresentado pelo modelo de cadeia de Markov de 2ª ordem.

Sharma e Panu (2014) realizaram um estudo similar ao apresentado em Sharma e Panu (2010) para os mesmos rios do Canadá. No estudo, os autores propuseram uma metodologia para prever as durações e magnitudes das secas. Os eventos de seca foram caracterizados para um nível de truncamento correspondente à mediana dos dados observados, aplicada às series de valores de SHI, discretizadas em intervalos semanais, mensais e anuais. O modelo baseado na teoria dos valores extremos conduz a resultados satisfatórios para os dados em escala anual e até certo ponto, para a escala mensal. A modelagem baseada em cadeia de Markov conduz a resultados semelhantes na mesma situação. Para os dados em escala semanal, a modelagem baseada em cadeia de Markov se mostra mais satisfatória. Os autores também avaliaram o desempenho da modelagem baseada em cadeia de Markov aplicada às series semanais submetidas à níveis de truncamento diferentes da mediana. Os níveis de truncamento utilizados no estudo foram os percentis 95 e 90. Nestes dois casos, a modelagem baseada em cadeia de Markov também conduziu a resultados satisfatórios.

Araújo e Bronstert (2016) apresentaram um método para avaliar secas hidrológicas no semiárido brasileiro. Os autores estudaram a bacia hidrográfica do Rio Jaguaribe. As secas hidrológicas foi avaliadas para:

1. Sistemas de grandes reservatórios, como é o caso dos Açudes Castanhão, Orós e Banabiú, para o período compreendido entre 1994 a 2012;
2. Reservatórios de todos os tamanhos e 17.986 cisternas; e
3. Reservatórios de tamanhos diferentes de forma plurianual.

Os autores observaram pelos resultados que a demanda de água é restrita na bacia hidrográfica, que as secas hidrológicas e meteorológicas estão frequentemente fora de fase, que há uma correlação negativa entre o nível de armazenamento e a severidade da seca e que pequenos sistemas de armazenamento não conseguem lidar com secas de longa duração.

Medeiros (2016) apresenta uma metodologia de avaliação de secas hidrológicas na qual se determina o volume de água abaixo de uma demanda requerida para diferentes intervalos de tempo, de onde se pode caracterizar a seca em função de sua duração, severidade e magnitude (teoria dos runs). Foram estudadas as secas ocorridas entre 1997 e 2015 em dois reservatórios de diferentes capacidades da bacia hidrográfica do Rio Piranhas-Açu no nordeste brasileiro. A bacia hidrográfica do Rio Piranhas-Açu, localizada nos Estados do Rio Grande do Norte e da Paraíba, está totalmente inserida na região semiárida do nordeste brasileiro, com área total de 43.580 km², abrange 147 municípios, dos quais 47 pertencem ao Rio Grande do Norte e 100 à Paraíba, com uma população total aproximada de 1,3 milhão de habitantes. Os reservatórios analisados foram o Açude de Cruzeta e a Barragem Armando Ribeiro Gonçalves. A autora também complementou sua análise utilizado o método de avaliação de secas hidrológicas apresentado por Araújo e Bronstert (2016) para comparar as características dos eventos de seca identificados por diferentes métodos para os mesmos reservatórios. Os resultados demonstraram que reservatório com maior capacidade de acumulação são mais eficientes e, assim, menos susceptível à seca. Foi verificado que a diferença básica entre as duas abordagens é o tempo de análise dos eventos de seca: enquanto pelo método baseado a teoria dos runs é possível estudar as secas ocorridas no passado para diagnosticar e fazer um planejamento dos usos dos reservatórios no futuro, o método apresentado por Araújo e Bronstert (2016) possibilita a avaliação das condições atuais para prever o início de uma seca hidrológica. A autora conclui também que as duas metodologias podem ser utilizadas simultaneamente pelos gestores de recursos hídricos, a fim de possibilitar uma análise mais abrangente dos eventos de seca ocorridos em bacias hidrográficas.

2.2. ANALISES BIVARIADAS DOS EVENTOS DE SECA

Shiau e Shen (2001) apresentaram um estudo sobre a análise de recorrência de secas hidrológicas de diferentes severidades, tendo como base as vazões médias mensais a montante do reservatório Shihmen, localizado no sudoeste de Taipei, na parte norte de Taiwan. O reservatório Shihmen é uma das maiores fontes de suprimento de água para uso urbano e agrícola da parte norte de Taiwan. Segundo os autores, a capacidade do reservatório é pequena quando comparada com as vazões de montante ou com as vazões demandadas por Taiwan. A série histórica de vazões médias mensais utilizada no estudo abrangia os anos de 1965 a 1992. O limitado número de eventos de seca hidrológica, obtido com base na teoria dos runs a partir da série histórica de vazões, se mostrou insuficiente para uma análise acurada da distribuição de probabilidade condicional das severidades das secas em função de suas durações. Desta forma, os autores utilizaram uma série sintética de 10.000 anos gerada a partir de um modelo de desagregação. Este modelo realizou primeiramente a geração dos valores anuais e posteriormente a desagregação destes em valores mensais. O nível de truncamento utilizado no estudo foi a mediana das vazões médias de cada mês. Para o cálculo da distribuição de probabilidade condicional das severidades para uma dada duração, os autores avaliaram o comportamento da distribuição de probabilidade das severidades para oito durações distintas de um a nove meses. A distribuição de probabilidade ajustada às severidades foi a Gama com dois parâmetros. Assim, para cada duração, foram calculados os parâmetros de forma e escala, os quais foram relacionados, por meio de regressão, às durações previamente selecionadas. Uma distribuição de probabilidade Geométrica foi utilizada para representar a duração dos eventos de secas hidrológicas. De posse da distribuição de probabilidade das severidades, das regressões dos parâmetros de forma e escala e da distribuição de probabilidade das durações, os autores obtiveram uma generalização da função de distribuição de probabilidade condicional das severidades para uma dada duração. Com relação ao comportamento dos parâmetros de forma e escala das distribuições de probabilidade das severidades, os autores observaram que o parâmetro de forma cresce linearmente com a duração, enquanto que o parâmetro de escala não apresenta nenhuma tendência de comportamento, sendo este considerado constante e igual à média dos valores obtidos para cada duração selecionada. Os autores utilizaram um modelo analítico para representar o intervalo de recorrência entre eventos de seca hidrológica. Os intervalos de recorrência adotados pelos autores foram de 1, 2, 5, 10, 20, 30 e 50 anos. As durações médias correspondentes a cada intervalo de recorrência foram obtidas pela generalização da distribuição de probabilidade condicional das severidades para uma dada duração. Como conclusão, os autores enfatizam que um passo importante na operação de reservatórios é o estudo de situações inesperadas de baixas vazões (secas hidrológicas), pois o entendimento do comportamento de tais situações é necessário para um planejamento eficiente dos recursos hídricos.

Salas et al. (2005) estudaram o comportamento dos eventos de secas hidrológicas no Rio Poudre, localizado no Colorado, cuja área de drenagem é de 1.890 mi^2. Os dados utilizados foram as vazões médias anuais compreendidas entre os anos de 1854 a 2002. Os eventos de secas foram obtidos para cinco níveis de truncamento, representados por percentuais da vazão média de longo termo. Distribuições de probabilidade do tipo Gama com dois parâmetros e Geométrica foram utilizadas para representar as severidades e as durações dos eventos de secas hidrológicas, respectivamente. Os autores também utilizaram o conceito de magnitude apresentado em Dracup et al. (1980b), o qual chamaram de intensidade, sendo representado por uma distribuição de probabilidade bivariada em função da duração. Devido à extensão da série de vazões, um modelo AR(1) foi utilizado para gerar uma série sintética de 50.000 anos. Com base nas distribuições de probabilidade das severidades e das durações, foram obtidas as distribuições de probabilidade bivariadas:

1. Das severidades serem maiores ou iguais a um determinado valor dada uma duração;
2. Das severidades e das durações serem maiores ou iguais a determinados valores;
3. Das intensidades serem maiores ou iguais a um determinado valor dada uma duração; e
4. Das intensidades e das durações serem maiores ou iguais a determinados valores.

Os autores avaliaram o comportamento dos períodos de retorno das durações dos eventos de secas tendo como referência as distribuições de probabilidade bivariadas. Uma análise de risco foi realizada gerando-se 2.000 séries sintéticas de vazões anuais com horizonte de 25 anos. A análise de risco avaliou a probabilidade de ocorrência, em 25 anos, de eventos com severidades maiores ou iguais a um determinado valor dada uma duração, e de eventos com intensidades maiores ou iguais a um determinado valor dada uma duração. Os autores concluíram que, de acordo com suas experiências, a metodologia apresentada é viável e pode dar boas informações a respeito das severidades das secas que podem ser esperadas na área de estudo.

Shiau et al. (2007) estudaram o comportamento dos eventos de secas hidrológicas no Rio Amarelo na China. Foram utilizados os dados de vazões médias mensais do Rio Tangnaihe na entrada do reservatório Longyangxia, o qual está localizado a montante do Rio Amarelo, com registros de vazões de 1919 a 2002. Alguns períodos de dados tiveram que ser reconstituídos com base nas informações de vazões do Rio Lanzhou. Os eventos se seca foram definidos com base em um nível de truncamento correspondente à mediana das vazões médias mensais. As durações dos eventos de seca foram representadas por uma distribuição de probabilidade Exponencial Mista e as severidades por uma distribuição de probabilidade Gama com 2 parâmetros. Os parâmetros das distribuições de probabilidade foram estimados com base no Método da Máximo Verossimilhança. O comportamento probabilístico conjunto das durações e severidades, foi modelado por uma cópula Clayton. O parâmetro da cópula foi estimado pelo Método da Inferência das Funções Marginais. Com base na análise bivariada, os autores calcularam os períodos de retorno dos eventos de seca para 2, 5, 10, 20, 50 e 100 anos. Os autores em sua conclusão comentam que, a seca mais severa que ocorreu no Rio Amarelo durante o período de 1919 a 2002 é a seca de 1930-1933, com uma duração de 36 meses e severidade de 5.264,8 m^3/s por mês. O período de retorno desta seca, calculado com base nas equações apresentada no trabalho, é de 105 anos. Entretanto, se o menor evento de não seca neste período for ignorado, então a seca passa a ter uma duração de 66 meses com severidade de 10.985,8 m^3/s por mês e o período de retorno passa para 3.868 anos. Por outro lado, a seca de 1977-1988 tem um período de retorno de 4,4 anos. Isso pode sugerir que a redução dramática da vazão do Rio Amarelo em sua jusante está deteriorada por outros fatores, como por exemplo, atividade humana.

Kwak et al. (2012) estudaram o comportamento dos eventos de seca no Rio Namhan, localizado na Coreia. Os dados utilizados foram as vazões médias mensais de 1967 a 2007. O nível de truncamento utilizado no estudo foi o percentil 70 da vazão média mensal de cada mês. As durações dos eventos de seca foram representadas pela distribuição de probabilidade Weibull e as severidades pela distribuição de probabilidade Gama. Para a análise bivariada das durações e severidades, os autores utilizam uma cópula Clayton. Com base no modelo bivariado das durações e severidades, os autores também calcularam os períodos de retorno dos eventos de seca para 2, 3, 5, 10, 30, 50, 80, 100 200 e 500 anos. Os autores concluem que, as secas são elementos importantes no planejamento e gerenciamento de recursos hídricos. Tentar entender as secas é considerado como um importante passo na solução de problemas relacionados às mesmas na perspectiva dos recursos hídricos. De acordos com os autores as informações de duração, severidade e período de retorno podem ser utilizadas como um dado preliminar importante para tratar dos eventos futuros.

De Michele et al. (2013) estudaram o comportamento dos eventos de seca no Rio Po. A bacia hidrográfica do Rio Po é a maior da Itália, com 74.000 km^2, onde 70.000 km^2 estão localizados na Itália e 4.000 km^2 na Suiça e França. Os eventos de seca foram caracterizados para a série de vazões médias diárias da estação fluviométrica Pontelagoscuro, com registros de 1923 a 2007. O nível de truncamento utilizado pelos autores foi a vazão determinada pela Sala de Controle, ou Control Room (CR), do Conselho da Bacia Hidrografia como nível de limiar de “Alerta”. Na determinação dos eventos de seca, os autores utilizaram um intervalo de tempo mínimo de três dias entre eventos, e uma duração mínima de cinco dias para caracterizar eventos mutuamente dependentes. No estudo, os autores avaliaram o comportamento probabilístico das durações e as intensidades dos eventos de seca. As distribuições de probabilidade Exponencial, Gama, Weibull, Pareto Generalizada e Generalizada de Valores Extremos foram utilizadas na análise uni-variada das durações e intensidades. O Método da Máximo Verossimilhança foi utilizado para estimar os parâmetros das distribuições de probabilidade, o teste Kolmogorov-Smirnov para avaliar a hipótese de aderência das distribuições de probabilidade aos dados observados e o Critério de Informação de Akaike para selecionar aquela de melhor ajuste. Os p-valores do teste de Kolmogorov-Smirnov foram calculados por simulação de Monte Carlo. Os autores utilizaram uma técnica de randomização para tornar a amostra observada das durações (composta por valores discretos), em amostras contínuas.

Os autores justificaram a utilização da técnica de randomização com base na alegação de que, a análise das durações observadas, na sua forma discreta, com vários dados repetidos (denominados no jargão estatístico de ties), pode ser ambígua e questionável, uma vez que, o rank dos dados não pode ser definido. No processo de randomização, por uma questão de consistência estatística, várias amostras independentes foram geradas. Os autores trabalharam com a geração de 400 randomizações independentes de duração e intensidade. Tanto para as amostras randomizadas quanto para a amostra observada de intensidades, a distribuição de probabilidade Gama é selecionada como aquela com melhor ajuste. Para as durações, a distribuição de probabilidade Weibull é a de melhor ajuste para as amostras randomizadas, enquanto que, a distribuição de probabilidade Gama é a de melhor ajuste para a amostra observada. A hipótese de aderência das distribuições de probabilidade foi validada pelo teste de Kolmogorov-Smirnov considerando um nível de significância de 90%. Segundo os autores, a técnica de randomização afim de contornar a presença de ties é satisfatória. Na análise bivariada dos eventos de seca, os autores utilizaram as cópulas Clayton, Frank e Gumbel, os mixes ClaytonGumbel, ClaytonFrank e GumbelFrank e, a família extra parametrizada de Khoudraji-Liebsher, sendo elas, XClaytonGumbel, XClaytonFrank, XGumbelFrank, XClaytonClayton, XGumbelGumbel e XFrankFrank. As cópulas foram testadas considerando as amostras randomizadas. Seus parâmetros foram estimados pelo Método da Máximo Verossimilhança e a seleção do melhor modelo foi realizada com o Critério de Informação de Akaike. O teste de Goodness-of-Fit com um nível de significância de 90% foi utilizado para avaliar a hipótese de aderência dos modelos bivariados. O melhor modelo bivariado encontrado pelos autores foi o XClaytonFrank.

Os autores também trouxeram o conceito de período de retorno dinâmico, formulado com base na teoria das cópulas e calculado por meio da abordagem da sobrevivência de Kendall (survival Kendall). Em suas conclusões, os autores comentam que, os resultados apresentados no trabalho podem ser usados pelas autoridades gestoras das bacias hidrográficas para planejar suas intervenções em tempo real, a fim de mitigar os efeitos das baixas vazões.

Salvadori e De Michele (2015) estudaram o comportamento dos eventos de seca do Rio Po, na Itália. Foram utilizadas as vazões médias diárias de cinco estações fluviométricas, Piacenza, Cremona, Boretto, Borgoforte e Pontelagoscuro. Os dados observados de vazão são desde 1924 em Piacenza e Borgoforte, desde 1972 em Cremona, desde 1943 em Boretto e desde 1923 em Pontelagoscuro. Todas as séries de dados terminam em 2007. Os eventos de seca foram caracterizados pela teoria dos runs utilizando três niveis de truncamento referentes às vazões de “Alerta”, “Alarme” e “Emergência” determinadas pela CR do Conselho da Bacia Hidrografia. Os autores estudaram as mesmas variáveis utilizadas por De Michele et al. (2013), utilizando também, as mesmas distribuições de probabilidade uni-variadas e cópulas. Os autores apresentam os conceitos de Hazard Trajectoy e Hazard Fan. O conceito do Hazard Trajectory consiste basicamente em plotar um gráfico as durações e intensidades dos eventos de seca com as isolinhas de seus períodos de retorno e acompanhar a evolução dos pares duração e severidade ao longo do tempo em relação ao período de retorno. A ideia do Hazard Fan é combinar as informações sobre o estado atual de uma seca com possíveis cenários futuros. Em suas conclusões os autores comentam que, o trabalho apresentado foi motivado por três principais considerações: 1) as secas são fenômenos espacialmente estendidos; 2) várias são as variáveis que podem desempenhar um papel significativo na evolução de uma seca; e 3) as secas são caracterizadas por uma dinâmica temporal lenta variando de semanas a meses. As duas primeiras considerações sugerem fortemente o uso de ferramentas multidimensionais para a caracterização das secas, enquanto que a terceira indica a possibilidade de acompanhar a seca em tempo real, a sua a evolução e seu status de perigo. Assim como em De Michelle et al (2013), os autores escrevem que os resultados apresentados no trabalho podem ser usados pelas autoridades gestoras das bacias hidrográficas para planejar suas intervenções em tempo real, a fim de mitigar os efeitos das secas

Hamdi et al. (2016) estudaram o comportamento dos eventos de seca hidrológica no Rio Medjerda na Tunisia, considerado com um dos rios mais longos e mais importantes do pais. A bacia hidrográfica do Rio Merjerda tem 24.000 km^2, dos quais 16.300 km^2 estão localizados na Tunisia. No estudo, foram utilizados os dados de vazões médias diárias da estação fluviométrica Jendouba para o período compreendido entre 1966 a 2008. Para a caracterização dos eventos de seca, os autores avaliaram três níveis de truncamento, correspondentes aos percentis 90, 85 e 80. As distribuições de probabilidade utilizadas nas análises uni-variadas das durações e severidades foram a Exponencial, Pareto Generalizada, Gama, Log-Normal e Log-Normal com três parâmetros. O teste do qui-quadrado e o Critério de Informação de Akaike foram utilizados para verificar a hipótese de aderência das distribuições de probabilidade aos dados observados e selecionar o melhor modelo uni-variado, respectivamente. Para estimar os parâmetros das distribuições de probabilidade, os autores utilizaram os Métodos dos Momentos, da Máximo Verossimilhança e dos Momentos Ponderado por Probabilidade, mas nem todos os três métodos foram aplicados a todas distribuições. Segundo os autores, os resultados indicaram que, a hipótese de aderência das distribuições de probabilidade Exponencial, Pareto Generalizada, Gama e Log-Normal com três parâmetros, aos dados de duração, não pode ser rejeitada para um nível de significância de 95%, sendo a distribuição de probabilidade Log-Normal com três parâmetros aquela com menor Critério de Informação de Akaike. A mesma conclusão é obtida na análise uni-variada das severidades. É importante ressaltar que tais conclusões não são observadas claramente nos dados apresentados pelos autores. Os autores apresentam somente a analise uni-variada das durações e severidades para o nível de truncamento correspondente ao percentil 90. A análise conjunta das durações e severidades foi realizada para as cópulas Gumbel, Clayton, Frank, Galamos, Normal, Husler-Reiss e Plackett. Os autores concluíram que a cópula Frank é aquela com melhor aderência aos dados observados. A hipótese de aderência das cópulas foi verificada pelo teste de Goodness-of-Fit e o Critério de Informação de Akaike foi utilizado para selecionar a cópula com melhor ajuste. Os parâmetros das cópulas foram estimados pelo Método da Máximo Pseudo-Verossimilhança. É importante ressaltar que, no estudo os autores apresentam somente a análise conjunta das durações e severidades para o nível de truncamento correspondente ao percentil 90. Com base nos resultados, também foram calculados os períodos de retorno dos eventos de seca para 2, 5, 10, 20, 50 e 100 anos. Ao final do trabalho os autores comentam também que, as cópulas podem ser facilmente aplicadas para a representar a estrutura de dependência das variáveis das secas que são altamente correlacionadas.

Kwak et al. (2016a) estudaram os eventos de seca hidrológica em quatro rios da Califórnia, Estados Unidos, Rio Sacramento acima da ponde Bend, Rio Feather no Lago Oroville, Rio Yuba próximo a Smartville e Rio American em Folsom. Foram utilizados os registros de vazões médias anuais compreendidos no período de 1872 a 1977. Os autores utilizaram um modelo de Redes Neurais Artificiais para reconstituir a série de vazões de 1560 a 1871. Para isso eles utilizaram os dados dos anéis de arvores das proximidades da área em estudo. Na caracterização dos eventos de seca, foi utilizado um nível de truncamento correspondente à mediana das vazões médias anuais. Os autores utilizaram as distribuições de probabilidade Exponencial e Gama na análise uni-variada das durações e severidades, com 95% de significância de acordo com o Coeficiente de Correlação do Gráfico de Probabilidade. A análise conjunta das durações e severidades foi realizada com cópulas Frank, Clayton e Gumbel, cujos parâmetros foram estimados pelo Método dos Momentos. A Distância Quadrática Mínima entre os valores teóricos e empíricos foi utilizada para selecionar a melhor cópula para as durações e severidades. Os autores concluíram que a cópula Frank é aquela com melhor ajuste. Tendo como base a cópula selecionada, os autores calcularam os períodos de retorno dos eventos de seca para 5, 10, 20, 30 40, 50, 80, 100, 200 e 500 anos. Os autores observaram que maioria das secas que ocorreram nos últimos cinco séculos não tem mais que cinquenta anos de período de retorno.

Kwak et al. (2016b) estudaram o comportamento dos eventos de seca hidrológica no Rio Namhan, localizado na Coreia, com área de drenagem de 2.442 km^2. Os dados históricos utilizados foram as vazões médias mensais de 1967 a 2008. Os autores utilizaram um nível de truncamento correspondente à 0,75 das vazões médias mensais aplicado a cada mês. Os autores explicam que este conceito é adotado na Coreia no âmbito do planejamento de recursos hídricos. Os autores adotaram distribuições as distribuições de probabilidade Exponencial e Gama para representar o comportamento probabilístico uni-variado das durações e severidades dos eventos de seca, respectivamente, com um nível de significância de 95% de acordo com o Coeficiente de Correlação do Gráfico de Probabilidade. Para a análise conjunta das durações e severidades dos eventos de seca, os autores também utilizaram cópulas Arquimedianas do tipo Frank, Clayton e Gumbel. Os parâmetros das cópulas foram estimados pelo Método dos Momentos. A Distância Quadrática Mínima entre os valores teóricos e observados foi utilizada para avaliar a cópula mais apropriada. A cópula Gumbel foi selecionada como aquela com melhor ajuste. Os autores também calcularam com base na cópula selecionada os períodos de retorno dos eventos de seca para 2, 3, 5, 10, 30, 50, 80, 100, 200 e 500 anos. Modelos de clima foram utilizados para avaliar o comportamento futuro das secas. Com tais modelos foi gerada uma série de vazões futuras até 2100, de onde foram avaliados os eventos de secas para três cenários de vazões com 30 anos cada um (2011 a 2029, 2039 a 2069 e 2070 a 2100). Os autores concluíram que os conceitos estabelecidos em seu trabalho podem ser utilizados para o planejamento e gerenciamento dos recursos hídricos como um indicador quantitativo.

Zhao et al. (2017) estudaram o comportamento das secas hidrológicas na bacia hidrográfica do Rio Weihe, na China, com 134.800 km^2, para o período compreendido entre 1960 a 2012. Foram estudadas três estações fluviométricas, Linjiacun, Xianyang e Huaxian, localizadas a montante, no meio e a jusante do Rio Weihe. O Rio Weihe é o maior tributário do Rio Amarelo no norte da China. Os dados de vazões médias mensais utilizados no estudo, foram obtidos no Instituto de Conservação do Solo e da Água. Foram estudados os eventos de seca somente para um nível de truncamento correspondente ao percentil 25 das vazões médias mensais. Os autores testaram seis distribuições de probabilidade para representar as durações e severidades dos eventos de seca, Exponencial, Gama com dois parâmetros, Log-Normal, Pareto Generalizada, Generalizada de Valores Extremos e Weibull. O teste de Anderson-Darling foi aplicado para verificar a hipótese de aderência dos modelos uni-variados e o Critério de Informação de Akaike utilizado para selecionar a distribuição de probabilidade com melhor ajuste. Para a análise conjunta das durações e severidades, os autores utilizaram três cópulas Arquimedianas, Clayton, Frank e Gumbel-Hougard. O Método dos Mínimos Quadrados Ordinários e o Critério de Informação de Akaike foram usados para testar a hipótese de aderência cópulas e selecionar aquela de melhor ajuste, respectivamente. Os parâmetros das cópulas foram estimados pelo Método da Inferência das Funções Marginais. Tanto para as durações quanto para as severidades, a hipótese de aderência da distribuição de probabilidade Pareto Generalizada foi rejeitada para as três estações fluviométricas. A hipótese de aderência de uma distribuição de probabilidade Exponencial aos dados observados das durações dos eventos de seca da estação fluviométrica Huaxian também foi rejeitada. O teste de Anderson-Darling foi aplicado com um nível de significância de 99% para a análise uni-variada, assim, para p-valores menores que 0,01, a hipótese de aderência das distribuições de probabilidade aos dados observados foi rejeitada. Para a análise conjunta das durações e severidades, as cópulas com melhor ajuste aos dados observados foram a Gumbel para a estação fluviométrica Linjiacun e Frank para as estações fluviométricas Xianyang e Huaxian. Com base nas cópulas selecionadas, os autores obtiveram os períodos de retorno dos eventos de seca para 2, 5, 10, 20, 50 e 100 anos. Também foi avaliada a sensibilidade do período de retorno quanto as distribuições de probabilidade uni-variadas e as cópulas. Em suas conclusões os autores comentam que, o período de retorno é sensível às distribuições uni-variadas de probabilidade e as cópulas. Portanto, é importante selecionar apropriadamente tais funções, e avaliar as sensibilidades e incertezas, com mais atenção na modelagem dos eventos levando em consideração a condição dos recursos hídricos e a exigência do nível de gestão.

Dodangeh et al. (2017) compararam os métodos paramétricos e não paramétricos para a estimativa dos parâmetros de cópulas, aplicadas no estudo das secas na bacia hidrográfica do Rio Karkhen, no sudoeste do Iran, com 43.000 km^2. No estudo foram utilizados os dados de vazões médias diárias de 3 estações fluviométricas, Tang Sazbon, localizada no principal rio, Rio Simereh, e Huleilan e Polchehr localizadas em dois tributários, nos Rios Chazman e Gamasiab, respectivamente. O período de dados corresponde de 1970 a 2012. Os autores utilizaram a teoria dos runs com um nível de truncamento referente ao percentil 75 da série de vazões para identificação e caracterização dos eventos de seca. Para as durações e severidades, treze distribuições de probabilidade foram avaliadas, cujos parâmetros foram estimados pelo Método da Máximo Verossimilhança. Os testes de Kolmogorov-Smirnov e do qui-quadrado foram utilizados para avaliar a hipótese de adequação dos modelos probabilísticos uni-variados e selecionar os mesmos. O nível de significância utilizado pelos autores foi de 95%. Os autores concluíram que as distribuições de probabilidade Gama e Log-Normal são as mais adequadas para representar as severidades dos eventos de secas nas estações fluviométricas Tang Sazbon, Huleilan e Polchehr. Para as durações, as distribuições de probabilidade com melhor adequação foram a Generalizada de Valores Extremos, Exponencial e Gama. Os autores utilizaram o coeficiente de correlação de Pearson para medir a dependência entre os valores de duração e severidade dos eventos de seca. O comportamento probabilístico conjunto das durações e severidades foi modelado pelas cópulas Clayton, Gumbel e Frank, cujos parâmetros foram estimados pelos Métodos da Máximo Verossimilhança e τ de Kendall. Os ajustes das cópulas foram avaliados pelos Critérios de Akaike, Baysiano e pelo Erro Quadrático Médio. Os autores concluíram que para as estações fluviométricas Tang Sazbon e Polchehr o método não paramétrico do τ de Kendall conduz a menores valores de Erro Quadrático Médio quando comparado ao Método da Máximo Verossimilhança (paramétrico). Para a estação fluviométrica Huleilan, o método paramétrico produz um Erro Quadrático Médio menor. Desta forma, os autores concluem que a utilização do método não paramétrico para estimar os parâmetros das cópulas pode ser empregado para construir a distribuição de probabilidade conjunta das durações e severidades dos eventos de seca das estações fluviométricas da bacia hidrográfica do Rio Karkheh. Assim, a cópula Frank foi a selecionada para a estação fluviométrica Tang Sazbon e a Gumbel para as estações fluviométricas Huleilan e Polchehr. Também foram avaliados os períodos de retorno conjunto das durações e severidades dos eventos de seca. Os resultados também mostraram que os períodos de retorno da estação fluviométrica localizada no principal rio são menores em comparação com aqueles localizadas nos tributários para o mesmo evento de seca. Também foi observado que o método não paramétrico de estimativa dos parâmetros das cópulas afeta a escolha da distribuição de probabilidade conjunta das durações e severidades e, consequentemente, o período de retorno dos eventos de seca. Maiores períodos de retorno são obtidos pelo método paramétrico em comparação com o método não paramétrico para a bacia hidrográfica do Rio Karkheh.

Ao longo da revisão bibliográfica, observa-se no âmbito da análise bivariada que os pesquisadores concentram quase que exclusivamente suas análises no cálculo dos períodos de retorno dos eventos de seca, tendo as cópulas como base para obtenção da distribuição de probabilidade conjunta das durações e severidades.

O período de retorno é uma importante ferramenta utilizada nos estudos dos sistemas de recursos hídricos. No caso dos eventos de seca, o cálculo do período de retorno permite a obtenção dos possíveis pares duração-severidade com a mesma recorrência. Devido ao tamanho limitado das séries de dados hidrológicos, é comum a obtenção de um número reduzidos de eventos de seca, que associados à análise do período de retorno, podem conduzir a uma falsa percepção de segurança hídrica. Fica evidente a necessidade da utilização de técnicas de geração de cenários aplicadas aos eventos de seca para que as autoridades gestoras das bacias hidrográficas possam mensurar o nível de risco de seus sistemas (Figura 2.2.1). Neste contexto, as cópulas se apresentam como alternativa l para geração de cenários de eventos de seca, cuja aplicação específica ainda é uma lacuna na bibliografia, necessitando de mais pesquisas e aprofundamentos.

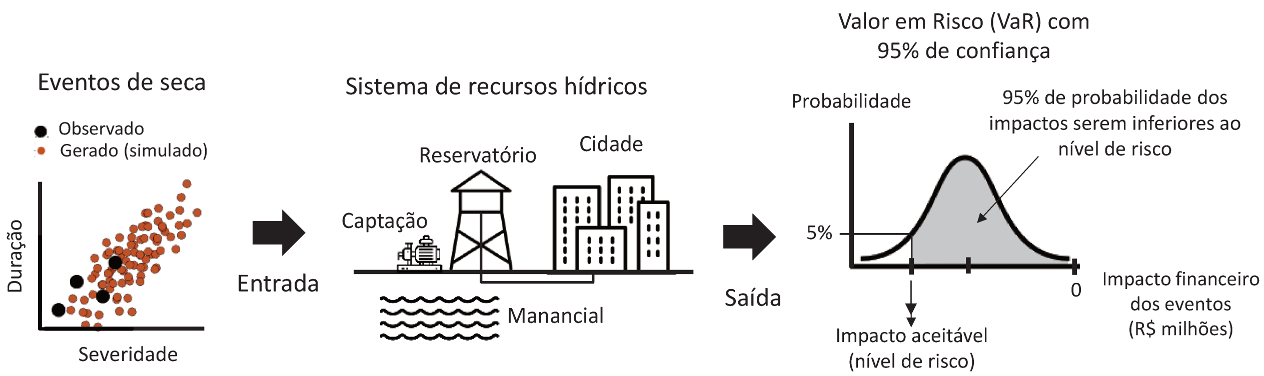


Figura 2.2.1. Geração de cenários para a quantificação do nível de risco.

3. MATERIAIS E MÉTODOS

3.1. ESTUDO DE CASO

No presente trabalho, é avaliado o comportamento dos eventos de seca do Rio Camanducaia localizado na UGRHI Nº 5 do Estado de São Paulo, pertencente à bacia hidrográfica do Rio Tiête (Figura 3.1.1), no sudeste do Brasil.

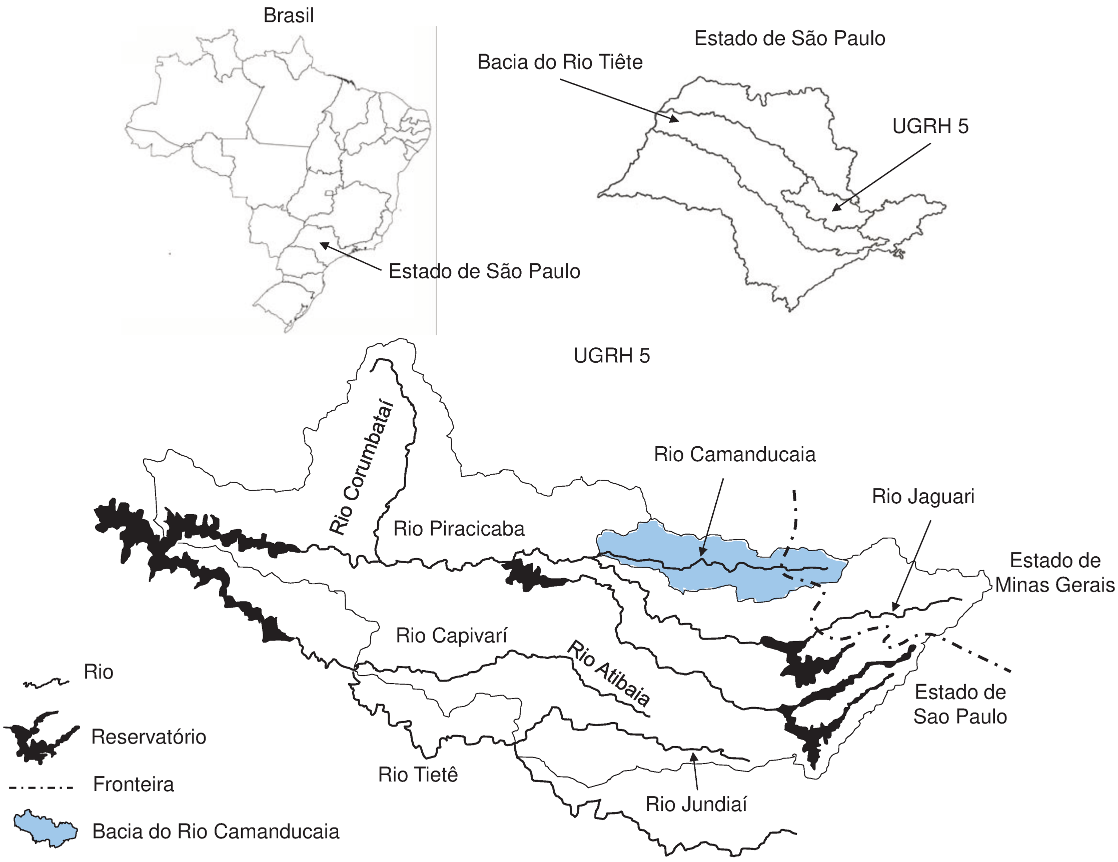


Figura 3.1.1. UGRHI Nº 5. Adaptado de Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2011).

A bacia hidrográfica do Rio Camanducaia abrange uma área de 1.032 km^2, dos quais 860 km^2 estão localizados no Estado de São Paulo e o restante no Estado de Minas Gerais. O Rio Camanducaia não possui reservatório e é responsável pelo abastecimento de água, total ou parcial, de 11 municípios, totalizando uma população em torno de 301 mil habitantes segundo os dados da Fundação Sistema Estadual de Análise de Dados do Estação de São Paulo e Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2017.

O Rio Camanducaia também é um dos principais afluentes do Rio Jaguari, outra importante fonte de abastecimento de água para a região. O Rio Jaguari faz parte do Sistema Cantareira que abastece a região metropolitana de São Paulo com mais de 10 milhões de pessoas.

Na Figura 3.1.2 é apresentada a bacia hidrográfica do Rio Camanducaia e os municípios abastecidos pelo mesmo.

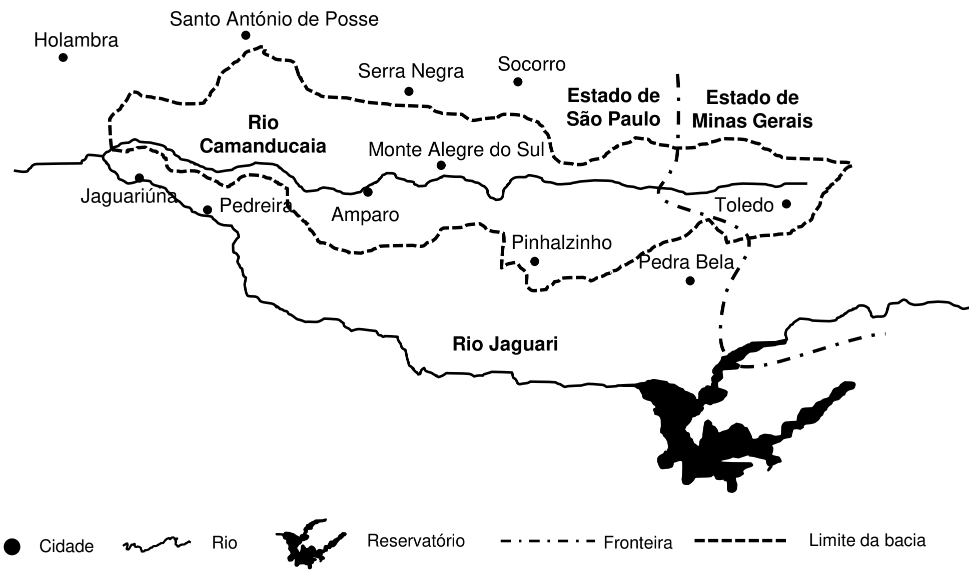


Figura 3.1.2. Bacia hidrográfica do Rio Camanducaia. Adaptado de Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2018).

O Rio Camanducaia possui três estações fluviométrica, cujas informações estão disponíveis no sítio do Sistema Integrado de Gerenciamento de Recursos Hídricos de São Paulo (Sistema de Gerenciamento de Recursos Hídricos, 2017):

1) 3D-001, denominada de Dal Bo, localizada no município de Jaguariúna, Estado de São Paulo, latitude 22° 40' 23'' e longitude 46° 58' 21'', com área de drenagem de 928 km^2;

2) 3D-002, denominada de Monte Alegre do Sul, localizada no município de Monte Alegre do Sul, Estado de São Paulo, latitude 22° 41' 44" e longitude 46° 40' 25", com área de drenagem de 387 km^2; e
3) 3D-017, denominada de Formiga, localizada no município de Toledo, Estado de Minas Gerais, latitude 22° 43' 40'' e longitude 46° 26' 18'', com área de drenagem de 102 km^2.

Os eventos de seca do Rio Camanducaia foram obtidos com base nos dados históricos de vazão média mensal da estação fluviométrica 3D-002, abrangendo o período de janeiro de 1944 a dezembro de 2016, totalizando um horizonte de análise de 73 anos com 876 registos. Na Figura 3.1.3 são apresentas as localizações dos postos fluviométricos na UGRH Nº 5 pertencentes ao Estado de São Paulo.

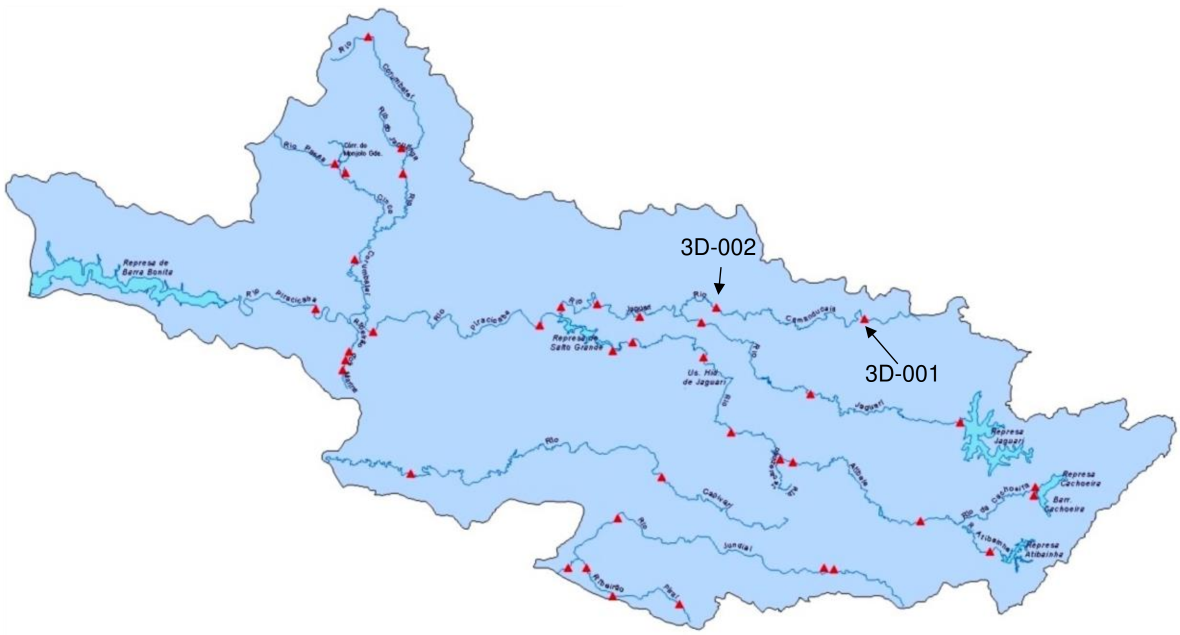
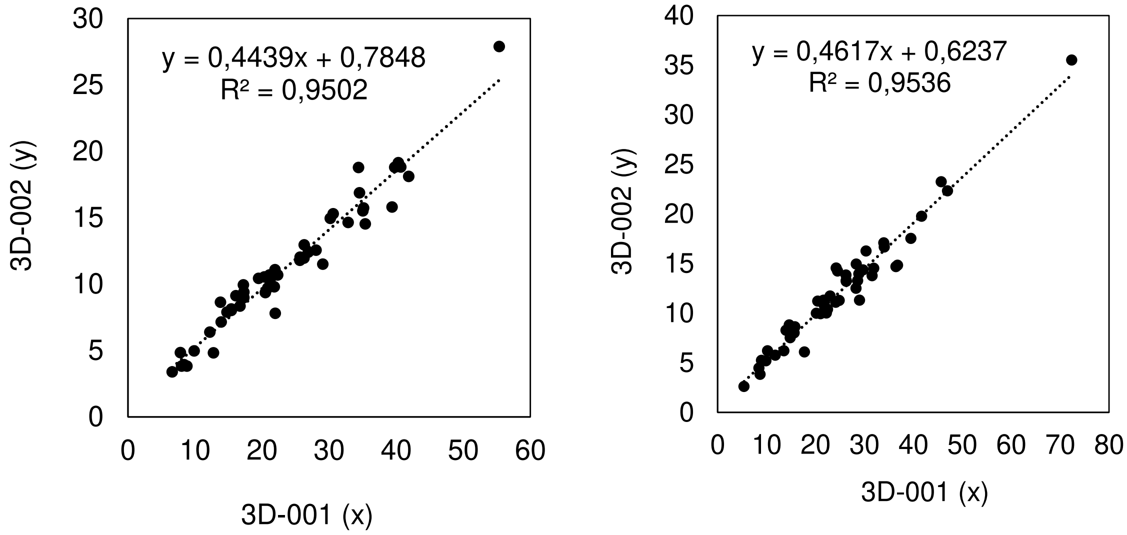


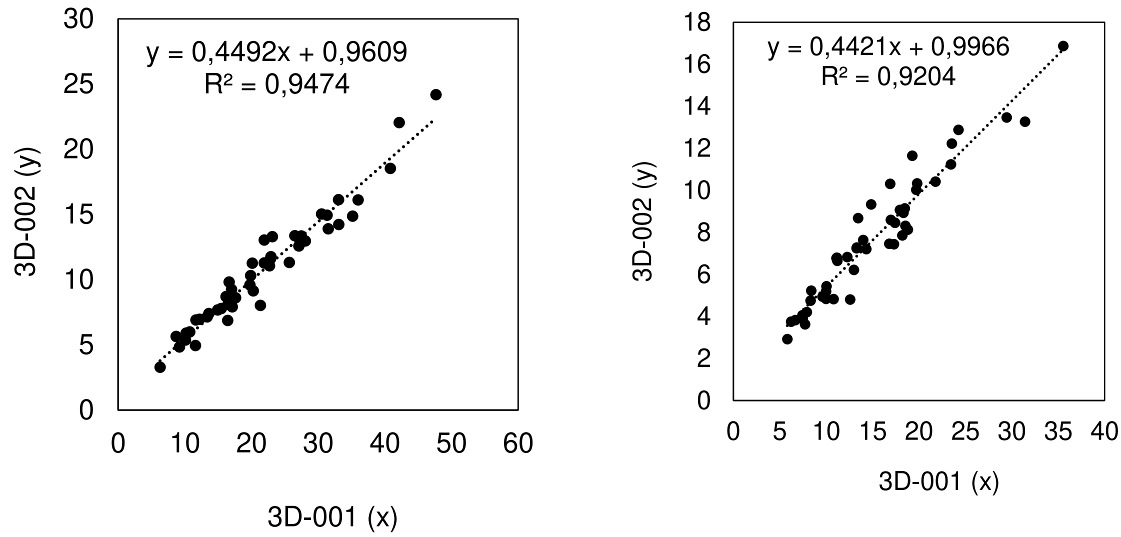
Figura 3.1.3. Postos fluviométricos da UGRH Nº 5 no Estado de São Paulo. Adaptado de Sistema de Gerenciamento de Recursos Hídricos (2017).

A estação fluviométrica 3D-002 foi selecionada por apresentar uma série histórica de dados com menor número de registros inexistentes, apenas oito valores mensais, de janeiro a maio de 1944, janeiro e fevereiro de 1994 e janeiro de 1999. O preenchimento dos dados inexistentes foi realizado com base em relações lineares entre as vazões das estações fluviométricas 3D-001 e 3D-002 apresentadas nas Figuras 3.1.4 a 3.1.6.



(a) (b)

Figura 3.1.4. Relação entre as vazões (em m^3/s): das estações fluviométricas 3D-001 e 3D-002 para os meses: (a) janeiro; (b) fevereiro.



(a) (b)

Figura 3.1.5. Relação entre as vazões (em m^3/s) das estações fluviométricas 3D-001 e 3D-002 para os meses: (a) março; (b) abril.

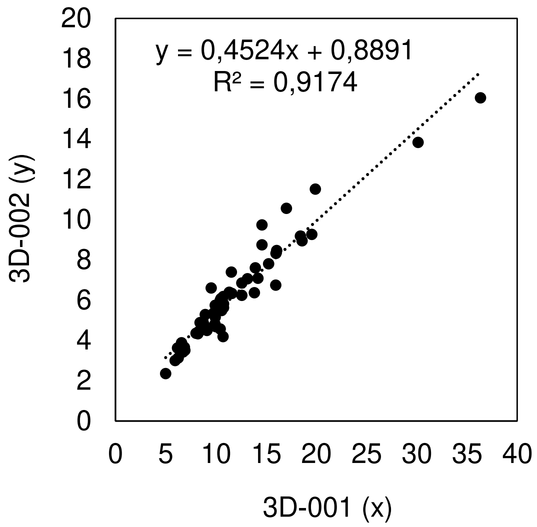


Figura 3.1.6. Relação entre as vazões (em m^3/s) das estações fluviométricas 3D-001 e 3D-002 para o mês de maio (m3/s).

Nas Figuras 3.1.7 e 3.1.8, são apresentadas a séries de vazões médias mensais do Rio Camanducaia para o período compreendido entre janeiro de 1944 a dezembro de 2016, e seu comportamento sazonal de longo termo para a média, mediana, valores máximo e mínimo.

Nas Tabelas 3.1.1 e 3.1.2 são apresentadas as principais estatísticas das vazões médias mensais do Rio Camanducaia para a estação fluviométrica 3D-002, como a média (Méd), desvio padrão (DP), máximo (Máx), mínimo (Mín) e mediana (Med).

Como pode ser observado na Figura 3.1.8 e na Tabela 3.1.1, o Rio Camanducaia apresenta o comportamento sazonal típico da região sudeste do país, onde se destacam dois períodos distintos, um de vazões baixas, compreendido entre os meses de maio a novembro, e outro de vazões elevadas, compreendido entre os meses de dezembro a abril. Cabe destacar na Figura 3.1.8 os elevados valores de vazões no mês junho em decorrência do El Ninõ de 1982-1983.

No Apêndice A, é apresentada a série histórica de vazões médias mensais do Rio Camanducaia para o período de janeiro de 1944 a dezembro de 2016 para a estação fluviométrica 3D-002.

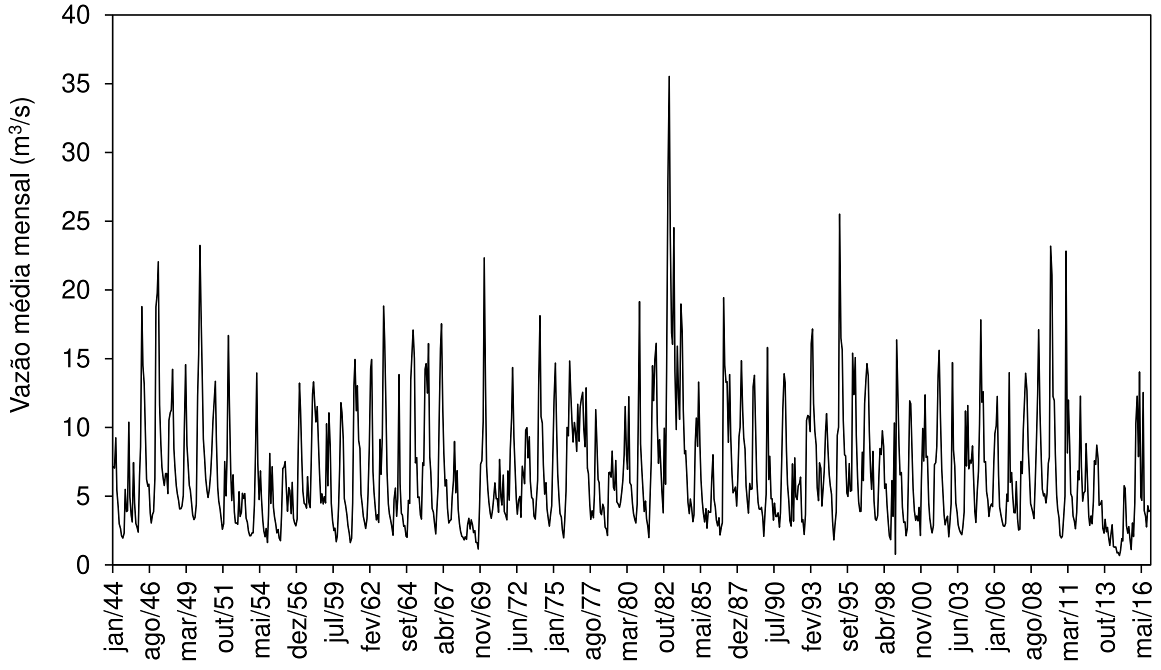


Figura 3.1.7. Vazões média mensais históricas do Rio Camanducaia na estação fluviométrica 3D-002.

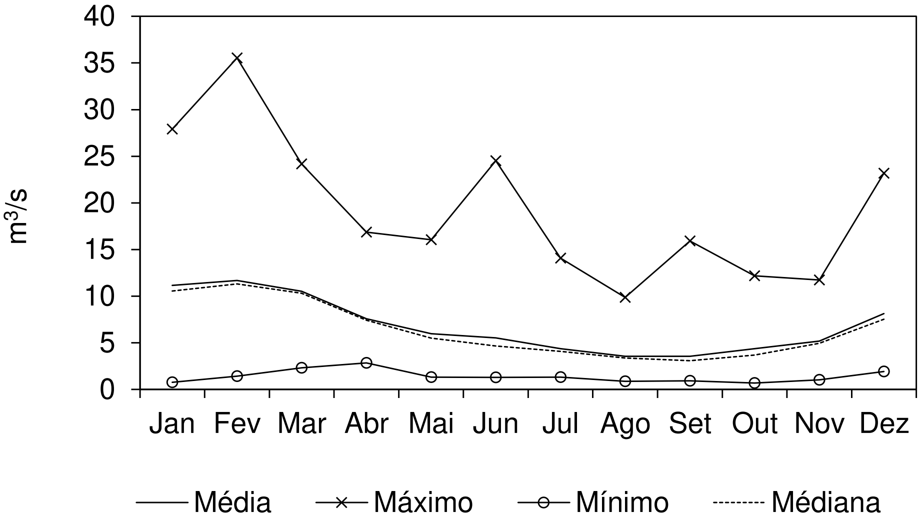


Figura 3.1.8. Comportamento sazonal histórico das vazões médias mensais de longo termo do Rio Camanducaia na estação fluviométrica 3D-002.

Tabela 3.1.1. Estatísticas mensais de longo termo para as vazões médias mensais (em m^3/s).do Rio Camanducaia na estação fluviométrica 3D-002.

Estatística	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Méd	11,15	11,68	10,53	7,58	5,98	5,54	4,37	3,56	3,57	4,37	5,18	8,13
DP	5,13	5,48	4,16	2,99	2,49	3,12	1,88	1,40	2,22	2,31	2,14	4,12
Máx	27,90	35,53	24,17	16,87	16,05	24,52	14,07	9,86	15,91	12,17	11,74	23,18
Mín	0,78	1,43	2,32	2,85	1,33	1,30	1,31	0,88	0,92	0,69	1,02	1,92
Med	10,54	11,31	10,31	7,43	5,49	4,66	4,07	3,36	3,09	3,68	4,95	7,54

Tabela 3.1.2. Estatísticas de longo termo para as vazões médias mensais do Rio Camanducaia na estação fluviométrica 3D-002.

Estatística	Valor (m^3/s)
Méd	6,80
DP	4,49
Máx	35,53
Mín	0,69
Med	5,44

Com relação ao uso da água, segundo o Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2011), em torno de 47% da demanda consuntiva de água superficial da bacia hidrográfica do Rio Camanducaia é destinada para irrigação. Verifica-se o predomínio de irrigação de culturas nobres (como flores, olerícolas e frutas) e também a fertirrigação da cana-de-açúcar no município de Camanducaia.

Na Tabela 3.1.3, são apresentadas as demandas consuntivas de água superficial da bacia do Rio Camanducaia segundo o Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2011).

Parte da demanda consuntiva de água superficial retorna aos Rio Camanducaia na forma de lançamento de efluentes, sendo estes decorrentes do usos urbano, industrial e rural (aqüicultura e outros). O lançamento de efluentes na bacia hidrográfica do Rio Camanducaia representa aproximadamente 40% da demanda consultiva de água superficial, isto é, 0,36 m^3/s segundo o Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2011).

Tabela 3.1.3. Demandas consuntivas de água superficial da bacia do Rio Camanducaia.

Tipo de demanda	Valor (m^3/s)
Industrial	0,16
Abastecimento urbano	0,31
Irrigação	0,43

Com relação ao balanço hídrico (oferta e demanda) da bacia hidrográfica do Rio Camanducaia, segundo o Comitê das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí (2011), a situação pode ser considerada como "confortável", pois a disponibilidade hídrica do Rio Camanducaia, mais o lançamento de efluentes, é superior à captação de água superficial (Tabela 3.1.4).

A disponibilidade hídrica tem como base a vazão $Q7\text{-}10$, vazão mínima com 10 anos de recorrência e duração de sete dias.

Tabela 3.1.4. Balanço hídrico na bacia hidrográfica do Rio Camanducaia (em m^3/s).

Disponibilidade hídrica			Captação superficial	Lançamento de efluentes	Balanço hídrico
Q_{7-10}	Reversão para a cidade de Serra Negra	Total			
3,6	-0,1	3,5	-0,85	0,36	3,01

Mesmo diante de uma situação considerada como "confortável", em termos de balanço hídrico de águas superficiais, a bacia hidrográfica do Rio Camanducaia sofreu com restrições de captação de água em agosto de 2015 devido a episódios de seca.

Desta forma, restrições de captação foram estabelecidas para a bacia hidrográfica do Rio Camanducaia com base na Resolução Conjunta Nº 50 da ANA e do Departamento de Águas e Energia Elétrica (DAEE) do Estado de São Paulo, do dia 21 de janeiro de 2015, denominada de REC 50/2015.

3.2. CARACTERIZAÇÃO DOS EVENTOS DE SECA

Tendo como base a teoria dos runs apresentada em Yevjevich (1967), dada uma série temporal de vazões $Q = \{q_t, t \in [0, T]\}$ e uma vazão de referência q_0, também chamada de nível de truncamento, um evento de seca irá ocorrer quando $q_t < q_0$.

Desta forma, para cada evento de seca, a teoria dos runs permite obter suas durações e suas severidades, ou déficits, conforme apresentado na Figura 3.2.1. Logo, quando as vazões são inferiores à vazão de referência, $q_t < q_0$, a duração e a severidade do evento de seca, representadas pelas variáveis aleatórias X_1 e X_2, podem ser definidas como:

$$x_{1i} = t_{final\,i} - t_{inicial\,i} \qquad \text{para } q_t < q_0 \tag{3.2.1}$$

$$x_{2i} = \sum_{t_{inicial\,i}}^{t_{final\,i}} q_0 - q_t \qquad \text{para } q_t < q_0 \tag{3.2.2}$$

onde x_{1i}, x_{2i}, $t_{inicial\,i}$ e $t_{final\,i}$ são a duração, a severidade, o tempo de início e o tempo de término do evento de seca i, respectivamente.

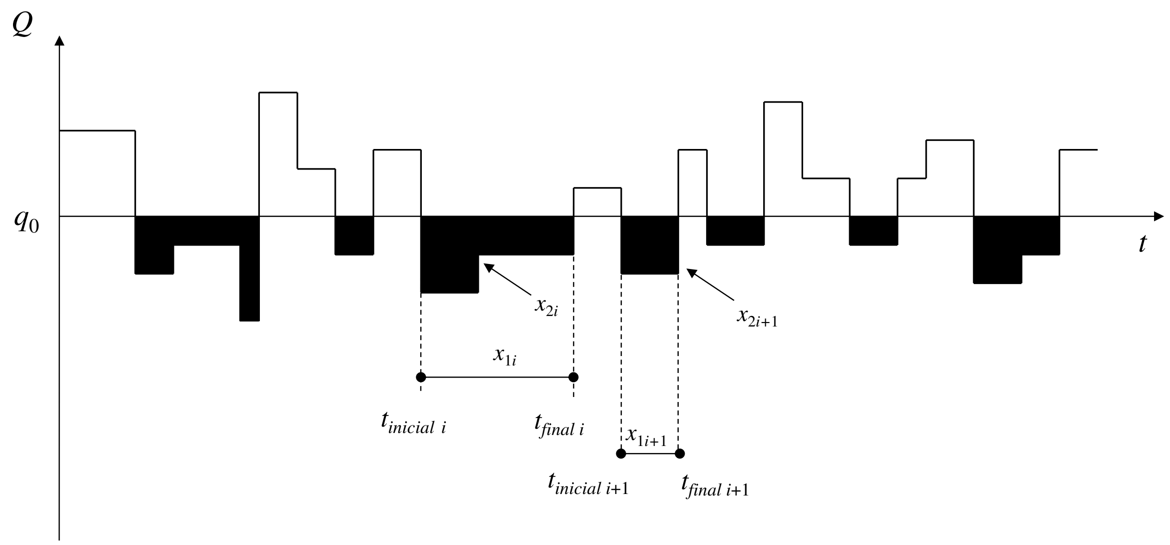


Figura 3.2.1. Caracterização dos eventos de seca. Adaptado de Yevjevich (1967).

Como pode-se observar, a aplicação da teoria dos runs nos permite obter um conjunto com n eventos de seca (observações), composto por dois subconjuntos, o das durações $X_1 = \{x_{1i},\ i = 1, \cdots, n\}$ e os das severidades $X_2 = \{x_{2i},\ i = 1, \cdots, n\}$, formados pelas principais variáveis relacionadas às secas, onde cada evento i é caracterizado individualmente por sua duração e severidade.

Fica evidente que o nível de truncamento da série de vazões influencia diretamente a quantidade de eventos observados de seca, os valores de suas durações e severidades e, consequentemente, suas distribuições de probabilidade e a estrutura de dependência entre ambas as variáveis (duração e severidade).

A definição dos níveis de truncamento utilizados neste trabalho teve como base dois critérios:

1) Os níveis de truncamento devem representar valores de vazões que possam comprometer ou que possam sinalizar um comprometimento da captação de água destinada ao uso humano ou dessedentização de animais, uso industrial, irrigação ou outros usos não consuntivos; e
2) A aplicação do nível de truncamento deve gerar uma quantidade mínima de eventos de seca que nos permita uma “razoável” análise estatística dos mesmos. Que neste trabalho, foi considerada de no mínimo dez eventos.

Assim, foram selecionados quatro níveis de truncamento, representados pelas vazões médias mensais com probabilidade de serem superadas em 98%, 95%, 92% e 90% do tempo (percentis), denominadas de $P98$, $P95$, $P92$ e $P90$, apresentadas na Tabela 3.2.2 e obtidas com base na curva de permanência dos registos históricos de vazões da estação fluviométrica 3D-002.

Os níveis de truncamento foram considerados constantes ao longo do tempo (situação mais crítica).

O objetivo de se utilizar mais de um nível de truncamento, é a possibilidade de se ampliar a análise do comportamento dos eventos de seca diante de situações distintas de demanda pelo uso da água.

Tabela 3.2.2. Níveis de truncamento para a vazões médias mensais do Rio Camanducaia para a estação fluviométrica 3D-002.

Nível de truncamento	Valor (m³/s)
$P98$	1,77
$P95$	2,12
$P92$	2,49
$P90$	2,70

Os níveis de truncamento adotados, também procuraram guardar uma relação com os valores da REC 50/2015 que estabelece os níveis de “Alerta” e “Restrição” para a captação de água superficial do Rio Camanducaia, sendo os níveis verificados com base nas vazões médias horárias de três dias consecutivos da estação fluviométrica 3D-001, calculados todas as segundas e quintas-feiras. A REC 50/2015 define que o estado o estado de “Alerta” é caracterizado por vazões inferiores a 2 m^3/s e superiores a 1,5 m^3/s, enquanto que, o estado de “Restrição” é caracterizado por vazões inferiores ou iguais a 1,5 m^3/s.

A aplicação da teoria dos runs pode gerar eventos de seca muito próximos e desta forma, mutuamente dependentes. Isso ocorre quando o valor da vazão excede a vazão de referência em curtos intervalos de tempo, onde o intervalo de tempo Δx_{1i} e o volume Δx_{2i}, caracterizados pelo término de um evento i e início do evento $i+1$, sejam “relativamente pequenos” (Figura 3.2.2).

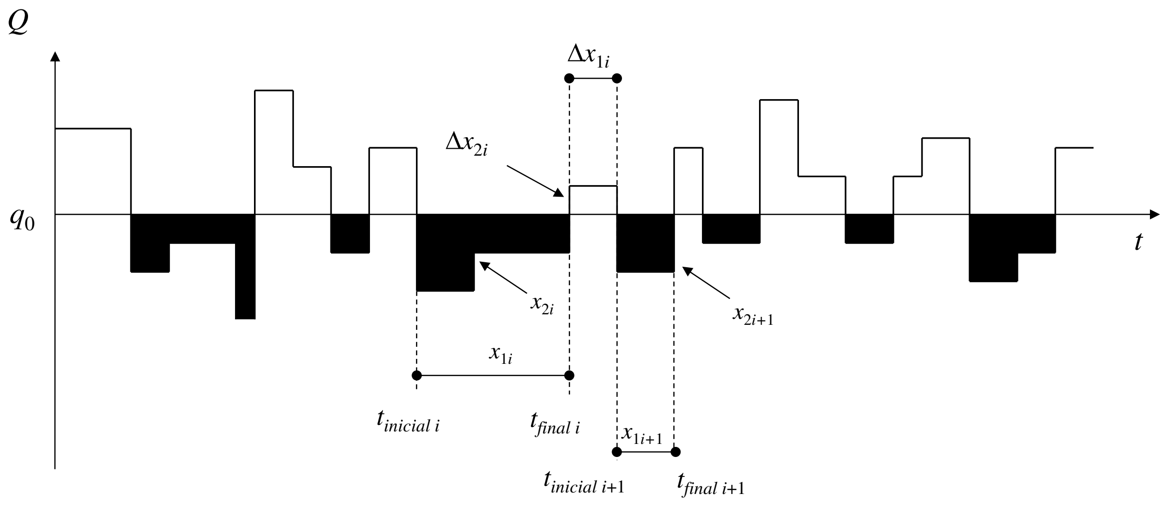


Figura 3.2.2. Eventos consecutivos. Adaptado de Zelenhasic e Salvai (1987).

Neste caso, pode-se representar os eventos i e $i+1$ como um único evento de seca i com duração e severidade:

$$\hat{x}_{1i} = x_{1i} + x_{1i+1} \tag{3.2.3}$$

$$\hat{x}_{2i} = x_{2i} + x_{2i+1} \tag{3.2.4}$$

Segundo Zelenhasic e Salvai (1987), a transição entre a condição de eventos mutuamente dependentes e independentes não é bem definida em termos práticos. Desta forma, os autores propuseram em seu trabalho duas alternativas:

1) Considerar somente uma das severidades, isto é, o maior valor entre x_{2i} e x_{2i+1}, eliminando desta forma o problema de dependência entre eventos de seca; ou
2) Considerar as duas severidades x_{2i} e x_{2i+1}, arbitrando um valor limite para o intervalo de tempo Δx_{1i} para definir a condição de dependência entre os eventos.

Zelenhasic e Salvai (1987) destacam que a segunda alternativa se apresenta como a mais próxima da realidade, pois o conjunto de dados analisados passa a ser composto por todos os eventos de seca hidrológica ocorridos no horizonte de tempo considerado. Desta forma, com a aplicação da proposta de Zelenhasic e Salvai (1987), obtêm-se novas amostras (ajustadas) para as durações e severidades, $\hat{X}_1$ e $\hat{X}_2$, respectivamente.

A hipótese de independência das observações em séries das durações e severidade dos eventos de seca do Rio Camanducaia foram verificadas utilizando os testes de Box-Pierce e Ljung para um nível de significância de 95%, de acordo com Box e Pierce (1970) e Ljung e Box (1978), respectivamente.

Os testes de Box-Pierce e Ljung são utilizados para examinar a hipótese de independência entre as observações de uma série temporal. Ambos os testes são aplicados aos resíduos de uma série temporal após o ajustes de um modelo Auto

Regressivo de Médias Móveis, ARMA(p,q), sendo p a ordem da parte auto regressiva e q a ordem da parte de média móvel do modelo. As hipóteses dos testes são:

1. H_0: os resíduos são independentes e identicamente distribuídos (hipótese nula ou verdadeira); ou
2. H_1: os resíduos não são independentes e identicamente distribuídos (hipótese alternativa).

Dada uma série temporal com n observações, as estatísticas Q dos testes de Box-Pierce e Ljung são apresentadas nas Equações (3.2.5) e (3.2.6), respectivamente.

$$Q = n(n+2)\sum_{j=1}^{K}\frac{\hat{r}_j^2}{n-j} \quad (3.2.5)$$

$$Q = n\sum_{j=1}^{K}\hat{r}_j^2 \quad (3.2.6)$$

sendo K o número de defasagens adotado para a função de auto correlação que está sendo testado (lag) e $\hat{r}_j$ a auto correlação estimada no lag j.

Os testes rejeitam a hipótese nula, H_0, se $Q > \chi^2_{1-\alpha,h}$, sendo $\chi^2_{1-\alpha,h}$ o valor de uma distribuição de probabilidade Qui-quadrado com nível de significância $1-\alpha$ e h gráus de liberdade. O grau de liberdade é calculado com base nos parâmetros do modelo ARMA e no número de lags que está sendo testado, $h = K - p - q$.

No Apêndice B são apresentadas as funções do ambiente computacional e linguagem de programação R para a obtenção e caracterização dos eventos de seca (B.1), tendo como base a teoria dos runs, aplicação dos testes de Box-Pierce e Ljung (B.2) e cálculo dos coeficientes de correlação de Pearson, τ de Kendal e ρ de Spearman (B.3).

3.3. DISTRIBUIÇÕES DE PROBABILIDADE DAS DURAÇÕES E SEVERIDADES

Tendo como base a revisão bibliográfica apresentada no Capítulo 2, foram avaliadas seis distribuições de probabilidade uni-variadas contínuas para a representação do comportamento probabilístico das durações e severidades dos eventos de seca do Rio Camanducaia (Tabela 3.3.1): Gama com dois parâmetros (GAM2), Exponencial (EXP), Weibull (WBL), Log-Normal (LOGN), Pareto Generalizada (GP) e Generalizada de Valores Extremos (GEV).

Tabela 3.3.1. Distribuições de probabilidade testadas das para representar o comportamento probabilístico das durações e severidades dos eventos de seca do Rio Camanducaia.

Distribuição de probabilidade	**Função densidade de probabilidade**	**Parâmetros**
GAM2	$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$	α: forma β: escala
EXP	$f(x) = \lambda e^{-\lambda x}$	λ: taxa
WBL	$f(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$	α: forma β: escala
LOGN	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$	μ: média σ: desvio padrão
GP	$f(x) = \frac{1}{\beta}\left(1 + \frac{\alpha(x-\xi)}{\beta}\right)^{(-\frac{1}{\alpha}-1)}$	α: forma ξ: locação β: escala
GEV	$f(x) = \frac{1}{\beta} t(x)^{\alpha+1} e^{-t(x)}$ $\text{sendo } t(x) = \begin{cases} \left(1 + \alpha\left(\frac{x-\xi}{\beta}\right)\right)^{-\frac{1}{\alpha}} & \text{se } \alpha \neq 0 \\ e^{-\frac{(x-\xi)}{\beta}} & \text{se } \alpha = 0 \end{cases}$	α: forma ξ: locação β: escala

Na Tabela 3.3.2 são apresentadas as expressões teóricas dos valores esperados das distribuições de probabilidade da Tabela 3.3.1.

Tabela 3.3.2. Valores esperados das distribuições de probabilidade.

Distribuição de probabilidade	$\boldsymbol{E}[\,\cdot\,]$
GAM2	$\alpha \cdot \beta$
EXP	${}^{1}/_{\lambda}$
WBL	$\beta \cdot \Gamma\left(1 + {}^{1}/_{\alpha}\right)$
LOGN	$e^{\left(\mu + {}^{\sigma^2}/_{2}\right)}$
GP	$\xi + \left({}^{\beta}/_{1-\alpha}\right)$ para $\alpha < 1$
GEV	$\begin{cases} \xi + {}^{\beta(g-1)}/_{\alpha} \text{ para } \alpha \neq 0 \text{ e } \alpha < 1 \\ \xi + \beta \cdot \gamma \text{ para } \alpha = 0 \\ \infty \text{ para } \alpha \geq 1 \end{cases}$ sendo $g = \Gamma(1-\alpha)$ e γ é a constante de Euler.

O Método da Máximo Verossimilhança (MLE), foi utilizado para estimar os parâmetros das distribuições de probabilidade a partir dos dados observados de durações e severidades. O MLE estima os valores dos parâmetros de um modelo estatístico maximizando a probabilidade dos dados observados por meio da função de verossimilhança.

Considerando uma amostra de uma variável aleatória com n observações, $X = (x_1, \cdots, x_n)$, independentes e identicamente distribuídas, com distribuição de probabilidade cuja função densidade de probabilidade é expressa como $f(x|\theta)$, sendo θ seu parâmetro, tal que $\theta \in \Theta$, $\Theta \subset \mathbb{R}^r$, $\mathbb{R} = [-\infty, +\infty]$, $r \in \mathbb{N}$, $\mathbb{N} = [0,1,2,3, \cdots]$. A função densidade de probabilidade conjunta pode ser expressa como:

$$f(x_1, \cdots, x_n|\theta) = f(x_1|\theta) \times \ldots \times f(x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \qquad (3.1.1)$$

Considerando os valores $x_1, \cdots, x_n$ fixos, a Equação (3.3.1) pode ser interpretada como sendo uma função do parâmetro θ, denominada de função de verossimilhança ($\mathcal{L}$):

$$\mathcal{L}(\theta|x_1, \cdots, x_n) = f(x_1|\theta) \times \ldots \times f(n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) \qquad (3.3.2)$$

A aplicação do logaritmo na Equação (3.3.2) transforma o produto das densidades de probabilidade na soma dos logaritmos das densidades:

$$\hat{l}(\theta|x_1, \cdots, x_n) = \log \mathcal{L}(\theta|x_1, \cdots, x_n) = \sum_{i=1}^{n} \log f(x_i|\theta) \qquad (3.3.3)$$

Assim, o MLE consiste em obter o valor de θ que maximiza a função de verossimilhança. Este valor é denominado de estimador de máximo verossimilhança ($\hat{\theta}$):

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \hat{l}(\theta|x_1, \cdots, x_n) \qquad (3.3.4)$$

Após aplicação do MLE, o teste de Kolmogorov-Smirnov (KS), teste KS, foi utilizado para verificar a hipótese de adequação das distribuições de probabilidade aos dados observados.

O teste KS é um teste não paramétrico sobre a igualdade de distribuições de probabilidade contínuas e unidimensionais. O teste é utilizado para comparar a distribuição de probabilidade empírica de uma amostra de dados observados com uma distribuição de probabilidade de referência.

Considerando uma amostra de uma variável aleatória com n observações, $X = (x_1, \cdots, x_n)$, a estatística D_n do teste KS corresponde a máxima distância vertical, em valor absoluto, entre a distribuição de probabilidade acumulada teórica (F) e empírica ($\hat{F}$) dos dados:

$$D_n = \sup_n \left| F(x) - \hat{F}(x) \right| \tag{3.3.5}$$

As hipóteses do teste KS são:

1. H_0: $F = \hat{F}$ (hipótese nula); ou
2. H_1: $F \neq \hat{F}$ (hipótese alternativa).

Para avaliação da hipótese nula, H_0, foi realizado um bootstrap paramétrico adaptado de Stute et al. (1993). Considerando a amostra observada de uma variável aleatória com n observações, $X = (x_1, \cdots, x_n)$, independentes e identicamente distribuídas, com distribuição de probabilidade acumulada $F(x|\theta)$:

1) Calcular a estatística D_n para a amostra observada;
2) A partir de $F(\cdot\,|\theta)$, gerar m pseudo observações da variáveis aleatórias X, cada uma com n observações, $X_1 = (x_{11}, \cdots, x_{n1}), \cdots, X_{m=}(x_{1m}, \cdots, x_{nm})$;
3) Calcular a estatística teste para cada pseudo observação (D_{nm});
4) Calcular o p-valor aproximado como:

$$\frac{1}{m} \sum_{i=1}^{m} \mathbf{1}(D_{ni} > D_n) \tag{3.3.6}$$

sendo $\mathbf{1}$ a função indicador, tal que $\mathbf{1} = 1$ se $D_{ni} < D_n$, para $i \in [1, \cdots, m]$, caso contrário $\mathbf{1} = 0$.

O p-valor representa a probabilidade de se obter uma estatística do teste igual ou mais extrema que aquela observada a partir da amostra, sob a hipótese nula.

Desta forma, o p-valor adotado neste trabalho para verificar a hipotese de adequação das distribuições de probabilidade aos dados observados das durações e severidades (hipotese nula) foi 0,05, que corresponde a um nível de significância de 95%. Logo, se o p-valor calculado pelo bootstrap paramétrico foi maior ou igual a 0,05, por exemplo 0,3, significa que, a hipotese nula deveria ser rejeitar se o nível de significância do teste tivesse sido escolhido como menor que 70%. Como o nível de significância adotado é de 95% (p-valor igual a 0,05), a hipotese nula não pode ser rejeitada.

Para selecionar a distribuição de probabilidade com melhor ajuste aos dados observados de durações e severidades foi utilizado o Critério de Informação de Akaike (AIC), calculado como (Akaike, 1974):

$$AIC = -2\log\mathcal{L}\left(\hat{\theta}\middle|x_1,\cdots,x_n\right) + 2p \qquad (3.3.7)$$

sendo $\mathcal{L}\left(\hat{\theta}\middle|x_1,\cdots,x_n\right)$ o máximo valor da função de verossimilhança e p o número de parâmetros da distribuição de probabilidade.

No Apêndice B são apresentadas as funções utilizadas e implementadas no ambiente computacional e linguagem de programação R para ajustes das distribuições de probabilidade uni-variadas (B.4 e B.5) e aplicação do teste de Kolmogorov-Smirnov (B.6).

3.4. CÓPULAS

Considerando os vetores de variáveis aleatórias $\boldsymbol{X} = (X_1,\cdots,X_k)$, independentes e identicamente distribuídas, com distribuições de probabilidade marginais contínuas $F_1,\cdots,F_k$, e $\boldsymbol{U} = (U_1,\cdots,U_k)$, para $\boldsymbol{X} \in \mathbb{R}^k$ e $\boldsymbol{U} \in {\mathbb{Z}^+}^k$, em que $\mathbb{Z}^+ = [0,1]$, existe um único valor de x_i, para $i \in [1,\cdots,k]$, tal que $F_i(x_i) = u_i$. Logo:

$$x_i = F^{(-1)}(u_i) \qquad (3.4.1)$$

Assim, sendo $\boldsymbol{X} = (X_1,\cdots,X_k)$ um vetor de variáveis aleatórias com distribuições de probabilidade contínuas $F_1,\cdots,F_k$, pode-se aplicar a transformação $\mathbb{R}^k \longmapsto \mathbb{Z}^k$ para se obter um vetor com os componentes de uma distribuição de probabilidade uniforme entre [0,1]:

$$(x_1, \cdots, x_k) \longmapsto \big(F_1(x_1), \cdots, F_k(x_k)\big) \tag{3.4.2}$$

Sklar (1959) apresentou as cópulas como funções de distribuição de probabilidade conjunta que associam distribuições de probabilidade marginais no intervalo [0,1].

Sendo $\boldsymbol{U} = (U_1, \cdots, U_k)$ um vetor de variáveis aleatórias com distribuição de probabilidade uniforme no intervalo [0,1], uma cópula C pode ser definida como uma distribuição de probabilidade conjunta, onde:

$$C(u_1, \cdots, u_k) = P(U_1 \leq u_1, \cdots, U_k \leq u_k) \tag{3.4.3}$$

Desta forma, uma cópula k-dimensional pode ser escrita como uma função $C: [0,1]^k \to [0,1]$, ou $C: \mathbb{Z}^{+^k} \to \mathbb{Z}^+$, isto é, uma função com domínio $\mathbb{Z}^{+^k}$ e contradomínio $\mathbb{Z}^+$, devendo satisfazer:

1) $\lim_{u_j \to 0} C\big(u_1, \cdots, u_{j-1}, u_j, u_{j+1}, \cdots, u_k\big) = 0$;
2) $\lim_{\forall i \in [1,\cdots,k] \neq j, u_i \to 1} C\big(u_1, \cdots, u_{j-1}, u_j, u_{j+1}, \cdots, u_k\big) = \big(1, \cdots, 1, u_j, 1, \cdots, 1\big) = u_i$; e
3) C não decresce em cada argumento (componente).

Teorema de Sklar: Sendo H a distribuição de probabilidade conjunta do vetor de variáveis aleatórias $\boldsymbol{X} = (X_1, \cdots, X_k)$, e $F_1, \cdots, F_k$ suas distribuições de probabilidade marginais contínuas. Então existe uma cópula k-dimensional C tal que para todo $\boldsymbol{X} \in \mathbb{R}^k$:

$$H(x_1, \cdots, x_k) = C\big(F_1(x_1), \cdots, F_k(x_k)\big) \tag{3.4.4}$$

Se $F_1, \cdots, F_k$ são contínuas, então C é única.

O Teorema de Sklar baseia-se em princípios que remontam o trabalho de Fréchet (1951), o qual afirma que distribuições de probabilidade marginais podem ser acopladas via função para descrever uma distribuição de probabilidade conjunta.

Corolário: Sendo H a distribuição de probabilidade conjunta do vetor de variáveis aleatórias $\boldsymbol{X} = (X_1, \cdots, X_k)$, e $F_1, \cdots, F_k$ suas distribuições de probabilidade marginais contínuas com inversas $F_1{}^{(-1)}, \cdots, F_k{}^{(-1)}$, respectivamente. Então, para $\boldsymbol{X} \in \mathbb{R}^k$ e $\boldsymbol{U} \in \mathbb{Z}^{+k}$ existe uma cópula C tal que:

$$C(u_1, \cdots, u_k) = C(\boldsymbol{u}) = H(F^{(-1)}(u_1), \cdots, F^{(-1)}(u_k)) \tag{3.4.5}$$

A Equação (3.4.5) permite estudar o fenômeno de dependência das variáveis sem fixar suas distribuições de probabilidade marginais.

A cópula C contém todas as informações da distribuição de probabilidade conjunta das variáveis, independente de suas distribuições de probabilidade marginais. Em resumo, as cópulas contêm as informações de como as variáveis dependem uma das outras. Mais detalhes sobre a teoria das cópulas são apresentados em Nelsen (2006) e Joe (1997).

Muitos modelos multivariados para representar a dependência entre as variáveis aleatórias são gerados por famílias de cópulas paramétricas, indexadas por um parâmetro θ, também chamado de parâmetro de dependência, que pode ser um valor real ou vetor (Genest et al., 1995).

Existe uma série de famílias de cópulas para representar a dependência entre variáveis aleatórios, como por exemplo, as cópulas Elípticas (Normal ou Gaussiana, e t-Student), Arquimedianas (Clayton, Gumbel, Frank, Ali-Mikhail-Haq and Joe), de Valores Extremos (Gumbel, Galambos, Husler-Reiss, Twan and t-EV), entre outras (Plackett and Farlie-Gumbel-Morgestern).

Considerando o vetor de variáveis aleatórias $\boldsymbol{X} = (X_1, \cdots, X_k)$, independentes e identicamente distribuídas, com distribuição de probabilidade conjunta H, distribuições de probabilidade marginais contínuas $F_1, \cdots, F_k$, e cópula C com parâmetro θ, sendo que $\theta \in \Theta$, $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}$. Segundo Genest et al. (1995), a estimativa de θ pode ser realizada de duas formas. Na primeira, considera-se que modelos probabilísticos uni-variados estão disponíveis para as distribuições de probabilidade marginais. Neste caso, usa-se diretamente a função de verossimilhança, estimando-se simultaneamente os parâmetros das distribuições de probabilidade marginais e da cópula. Na segunda, a estimativa de θ é realizada em duas etapas distinta, onde na primeira estima-se os parâmetros das distribuições de probabilidade marginais, e na segunda o parâmetro da cópula.

A estimativa das distribuições de probabilidade marginais pode ser realizada de forma paramétrica, com base no estimador de máximo verossimilhança, ou de forma não-paramétrica, a partir das distribuições de probabilidade marginais empíricas ($\hat{F}$) das variáveis aleatórias:

$$\hat{F}_i(x) = \frac{1}{n+1}\sum_{j=1}^{n} \mathbf{1}\left(x_{ij} \leq x\right) \tag{3.4.6}$$

sendo n é o número de observações para cada variável aleatória e $\mathbf{1}$ é a função indicador, tal que $\mathbf{1} = 1$ se $x_{ij} \leq x$ para $i \in [1, \cdots, k]$ e $j \in [1, \cdots, n]$, caso contrário $\mathbf{1} = 0$.

Sendo c a função densidade da cópula C e $\hat{u}_i = \hat{F}_i(x_i)$, para $i \in [1, \cdots, k]$, as pseudo observações que compõem a variáveis aleatórias $\hat{U} = (\hat{u}_1, \cdots, \hat{u}_k)$, a função de pseudo verossimilhança pode ser definida como:

$$\mathcal{L}(\theta|\hat{u}_1, \cdots, \hat{u}_k) = \prod_{i=1}^{k} c(\hat{u}_1, \cdots, \hat{u}_k|\theta) \tag{3.4.7}$$

Aplicando o logaritmo na Equação (3.4.7) temos:

$$\hat{l}(\theta|\hat{u}_1, \cdots, \hat{u}_k) = \log \mathcal{L}(\theta|\hat{u}_1, \cdots, \hat{u}_k) = \sum_{i=1}^{k} \log c(\hat{u}_1, \cdots, \hat{u}_k|\theta) \tag{3.4.8}$$

Desta forma, o estimador de máximo pseudo verossimilhança da cópula C é calculado como:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \hat{l}(\theta|\hat{u}_1, \cdots, \hat{u}_k) \tag{3.4.9}$$

Neste trabalho, utilizou-se as estimativas das distribuições de probabilidade marginais realizadas de forma paramétrica para a determinação do estimador de máximo pseudo verossimilhança da cópula.

Considerando um vetor de variáveis aleatória $\boldsymbol{X} = (X_1, \cdots, X_k)$, $\boldsymbol{X} \in \mathbb{R}^k$, com distribuição de probabilidade conjunta H e distribuições de probabilidades marginais contínuas $F_1, \cdots, F_k$. Segundo o Teorema de Sklar, existe uma cópula C tal que $H(x_1, \cdots, x_k) = C(F_1(x_1), \cdots, F_k(x_k))$, onde C é única no intervalo [0,1]. Quando C é desconhecida, considera-se para $\boldsymbol{X}$ um modelo paramétrico de cópula pertencente a uma classe $\mathcal{C}_0 = \{C(\cdot\,|\theta): \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^r$, $r \in \mathbb{N}$.

Desta forma, após estimativa de θ, deve-se testar a hipótese nula $H_0: C \in \mathcal{C}_0$. Os testes de qualidade de ajuste, ou teste GOF (Goodness-of-Fit), são amplamente discutidos e utilizados na verificação da qualidade de ajuste de modelos paramétricos de cópulas como pode ser visto em Wang e Wells (2000), Fermanian (2005), Genest et al. (2006), Genest e Favre (2007), Genest e Rémillard (2008), Genest et al. (2009), Kojadinovic e Yan (2010), Kojadinovic e Yan (2011), Kojadinovic et al. (2011) e Zhang et al. (2016).

Sendo $\boldsymbol{U} = (U_1, \cdots, U_k)$ um vetor de variáveis aleatórias com n observações cada uma, tal que $\boldsymbol{U} \in \mathbb{Z}^{+k}$ e $U_1 = (u_{11}, \cdots, u_{k1}), \cdots, U_{n=}(u_{1n}, \cdots, u_{kn})$. Para todo $u = (u_1, \cdots, u_k) \in \mathbb{Z}^{+k}$, pode-se definir a cópula empírica $\hat{C}$ como:

$$\hat{C}(u_1, \cdots, u_k) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(u_{1i} \leq u_1, \cdots, u_{ki} \leq u_k) \qquad (3.4.10)$$

sendo $\mathbf{1}$ a função indicador, tal que $\mathbf{1} = 1$ se $u_{ij} < u_i$ para $i \in [1, \cdots, k]$ e $j \in [1, \cdots, n]$, caso contrário $\mathbf{1} = 0$.

Dado que a cópula empírica é não-paramétrica, $\hat{C}$ é considerada como referência para testar a hipótese $H_0: C \in \mathcal{C}_0$. Assim, $\hat{C}$ é um estimador consistente da cópula desconhecida C (Genest et al., 2009; Kojadinovic et al., 2011).

Como sugerido em Fermanian (2005), Quessy (2005) e Genest e Rémillard (2008), é natural que o teste GOF compare a cópula empírica $\hat{C}$ com o modelo paramétrico $C(.\,|\theta)$ adotado para C, assumindo a hipótese nula $H_0: C \in \mathcal{C}_0$.

Neste contexto, como apresentado em Kojadinovic et al. (2011), Genest et al. (2009), tendo como base o teste de um processo empírico tal que $\sqrt{n}\{\hat{C}(u_1,\cdots,u_k) - C(u_1,\cdots,u_k|\theta)\}$, foi proposto um teste GOF baseado na estatística de Cramér-von Mises (S_n) onde:

$$\begin{aligned} S_n &= \int_{[0,1]^k} n\{\hat{C}(u_1,\cdots,u_k) - C(u_1,\cdots,u_k|\theta)\}^2 \mathrm{d}\hat{C}(u_1,\cdots,u_k) \\ &= \sum_{i=1}^{n}\{\hat{C}(u_{1,i},\cdots,u_{ki}) - C(u_{1,i},\cdots,u_{ki}|\theta)\}^2 \end{aligned} \tag{3.4.11}$$

O p-valor do teste GOF baseado na estatística de Cramér-von Mises pode ser obtido por um bootstrap paramétrico conforme apresentado em Genest e Rémillard (2008) e Genest at al. (2009) e detalhado na sequência.

Considerando um vetor de variáveis aleatórias $\boldsymbol{X} = (X_1,\cdots,X_k)$, com distribuição de probabilidade conjunta H e distribuições de probabilidades marginais contínuas $F_1,\cdots,F_k$, sendo que cada variável aleatória possui n observações, tal que $\boldsymbol{X} \in \mathbb{R}^k$ e $X_1 = (x_{11},\cdots,x_{k1}),\cdots,X_{n=}(x_{1n},\cdots,x_{kn})$, tem-se o vetor de variáveis aleatórias $\boldsymbol{U} = (U_1,\cdots,U_k)$, tal que $\boldsymbol{U} \in \mathbb{Z}^{+^k}$ e $U_1 = (u_{11},\cdots,u_{k1}),\cdots,U_{n=}(u_{1n},\cdots,u_{kn})$. Assumindo a hipótese nula $H_0{:}\, C \in \mathcal{C}_0$, tal que $\mathcal{C}_0 = \{C(\cdot\,|\theta){:}\, \theta \in \Theta\}$, $\Theta \complement \mathbb{R}^r$, $r \in \mathbb{N}$:

1) Calcular a estatística S_n para amostra observada;
2) A parir de $C(\cdot\,|\theta)$ gerar m pseudo observações de $\boldsymbol{U}$; $\widehat{\boldsymbol{U}}_m = (\widehat{U}_{1m},\cdots,\widehat{U}_{km})$, cada uma contendo n observações, $\widehat{U}_{1m} = (\hat{u}_{11m},\cdots,\hat{u}_{k1m}),\cdots,\widehat{U}_{nm=}(\hat{u}_{1nm},\cdots,\hat{u}_{knm})$, tal que $\widehat{\boldsymbol{U}}_m \in \mathbb{Z}^{+^k}$;
3) Para cada pseudo observação, assumir a hipótese nula $H_0{:}\, C \in \mathcal{C}_0$, estimar o parâmetro θ e calcular a estatística do teste (S_{nm});
4) Calcular o p-valor aproximado como:

$$\frac{1}{m}\sum_{i=1}^{m} \mathbf{1}(S_{ni} > S_n) \tag{3.4.12}$$

sendo $\mathbf{1}$ a função indicador, tal que $\mathbf{1} = 1$ se $S_{ni} < S_n$, para $i \in [1, \cdots, m]$, caso contrário $\mathbf{1} = 0$.

Assim, como para as distribuições de probabilidade uni-variadas, o p-valor adotado para verificar a hipotese de adequação das cópulas aos dados observados de durações e severidades foi 0,05, sendo o AIC utilizado para selecionar as cópulas de melhor adequação.

Tendo como base a revisão bibliográfica apresentada no Capítulo 2, neste trabalho foram avaliadas seis cópulas para modelar a dependência entre as durações e severidades dos eventos de seca do Rio Camanducaia e, consequentemente, a distribuição de probabilidade conjunta entre elas, sendo cinco cópulas Arquimedianas, Clayton (CLA), Frank (FRA), Gumbel (GUM), Joe (JOE) e Ali-Mikhail-Haq (AMH), e uma cópula Elíptica do tipo Gaussiana (GAU).

As cópulas avaliadas neste tralho abrangem uma variedade de tipos de dependência entre as variáveis, possibilitando a investigação de como as durações e severidades dos eventos de seca do Rio Camanducaia se comportam mutuamente.

As cópulas Elípticas são obtidas diretamente do Teorema de Sklar, especificamente quando a função de distribuição conjunta H é elíptica. Elas possuem simetria radial com coeficientes de caudas superior e inferior iguais. Portanto, não possuem dependência nas caudas e, por isso, não são indicadas para modelar a relação entre variáveis aleatórias de eventos extremos.

As cópulas Arquimedianas podem representar uma grande variedade de dependência entre variáveis aleatórias, incluindo assimetria e dependências nas caudas, contornando as principais limitações das cópulas Elípticas. No entanto, estes tipos de cópulas não derivam diretamente do Teorema de Sklar e, desse modo, é necessário considerar algumas premissas para que essas funções sejam de fato cópulas.

As cópulas Arquimedianas são construídas a partir de uma função $\varphi: [0,1] \mapsto [0, \infty]$, chamada de função geradora da cópula, contínua e estritamente decrescente, tal que $\varphi(1) = 0$ e $\varphi(0) \leq \infty$. Então sua pseudo inversa é definida é como:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), 0 \le t \le \varphi(0) \\ 0, \varphi(0) \le t \le +\infty \end{cases} \quad (3.4.13)$$

sendo $\varphi^{[-1]}: [0,1] \to [0,\infty]$.

A pseudo inversa $\varphi^{[-1]}$ é contínua e não crescente em [0,∞] e estritamente decrescente em [0,$\varphi(0)$]. Além disso, $\varphi^{[-1]}\big(\varphi(t)\big) = t$ para todo $t \in [0,1]$ e:

$$\varphi\left(\varphi^{[-1]}(t)\right) = \begin{cases} t, 0 \le t \le \varphi(0) \\ \varphi(0), \varphi(0) \le t \le +\infty \end{cases} = \min(t, \varphi(0)) \quad (3.4.14)$$

Se $\varphi(0) = \infty$, então $\varphi^{[-1]} = \varphi^{-1}$.

Assim, uma cópula Arquimediana é definida como:

$$C(u_1, \cdots, u_k) = \varphi^{-1}(\varphi(u_1), \cdots, \varphi(u_k)) \quad (3.4.15)$$

Por simplificação, considerando uma cópula C bi-dimensional, $C: [0,1]^2 \to [0,1]$, se C é uma cópula Arquimediana com função geradora φ, então:

1) C é simétrica, isto é, $C(u_1, u_2) = C(u_2, u_1)$; e
2) C é associativa, isto é, $C(C(u_1, u_2), w) = C(u_1, C(u_2, w))$.

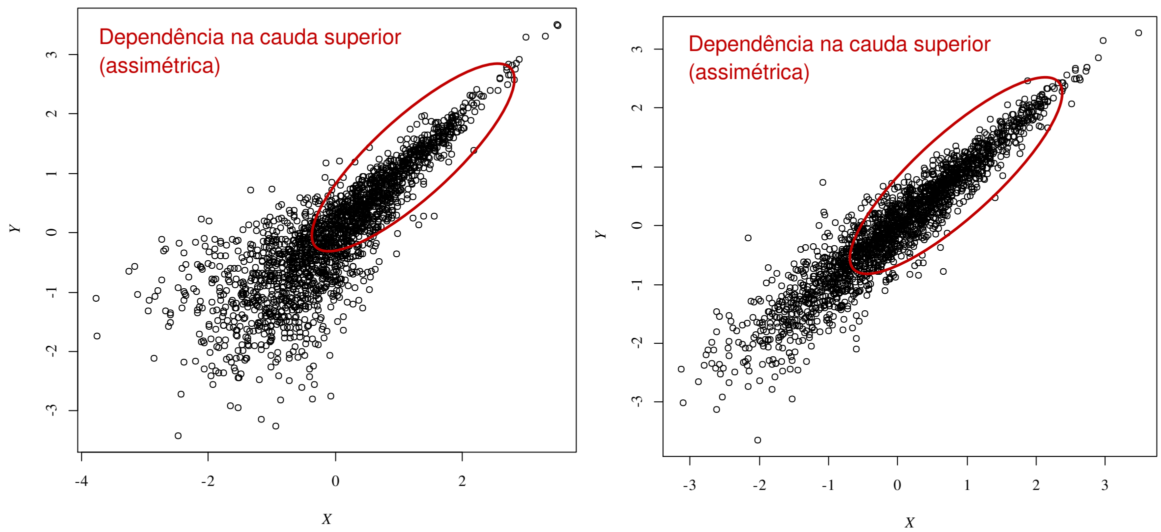
Na Tabela 3.4.1, são apresentadas as cópulas Arquimedianas avaliadas para representar o comportamento conjunto das durações e severidades dos eventos de seca do Rio Camanducaia, juntamente com suas funções geradoras e seus parâmetros. Na mesma tabela, também é apresentada a cópula GAU, sendo Φ^{-1} a inversa da distribuição de probabilidade Normal Padronizada.

Considerando valores pré-definidos para o parâmetro θ, nas Figuras 3.4.1 a 3.4.3, são apresentadas 2.000 simulações de duas variáveis aleatórioa X e $Y \in \mathbb{R}$, ambas com distribuição de probabilidade Normal Padronizada ($\mu = 0$ e $\sigma = 1$) e distribuição de probabilidade conjunta representada pelas cópulas da Tabela 3.4.1. Como pode-se observar, as simulações exemplificam visualmente como cada cópula “codifica” a dependência entre X e Y.

Tabela 3.4.1. Cópulas avaliadas na modelagem das durações e severidades dos eventos de seca do Rio Camanducaia (Nelsen, 2006; Silva e Lopez, 2008).

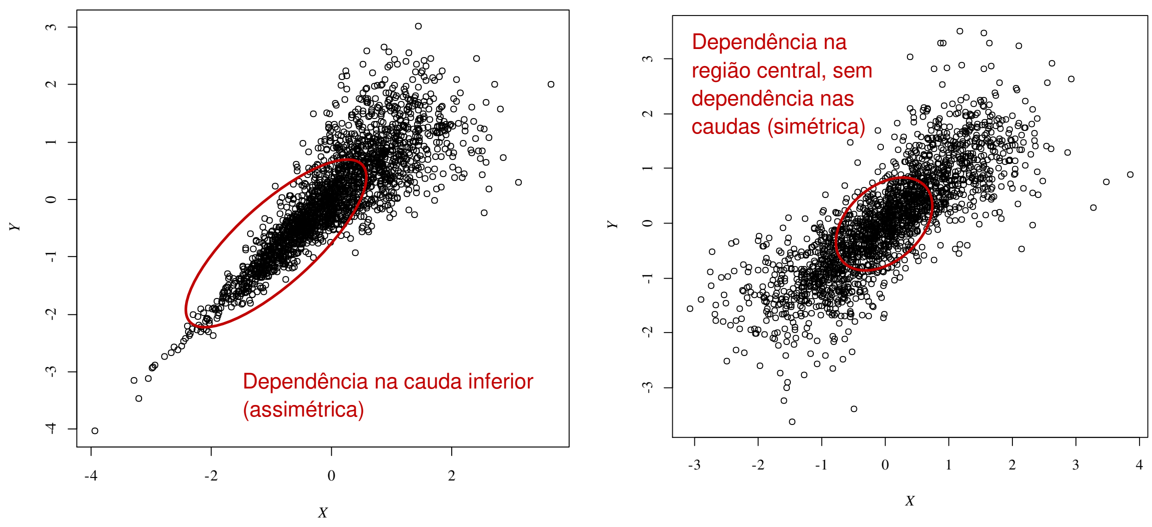
Tipo	Cópula	$\varphi(t)$	$C(u_1, u_2)$	$\theta \in$
Arquimediana	CLA	$\frac{1}{\theta}(t^{-\theta}-1)$	$[\max(u_1^{-\theta}+u_2^{-\theta}-1{,}0)]^{-1/\theta}$	$[-1,\infty)\backslash\{0\}$
	GUM	$(-\ln t)^{\theta}$	$e^{\left(-((-\ln u_1)^{\theta}+(-\ln u_2)^{\theta})^{1/\theta}\right)}$	$[1,\infty)$
	FRA	$-\ln\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$	$-\frac{1}{\theta}\ln\left(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$	$(-\infty,\infty)\backslash\{0\}$
	JOE	$-\ln\left(1-(1-t)^{\theta}\right)$	$1-\left((1-u_1)^{\theta}+(1-u_2)^{\theta}-(1-u_1)^{\theta}(1-u_2)^{\theta}\right)^{1/\theta}$	$[1,\infty)$
	AMH	$\ln\left(\frac{1-\theta(1-t)}{t}\right)$	$\frac{u_1 u_2}{1-\theta(1-u_1)(1-u_2)}$	$[-1,1)$
Elíptica	GAU	Não se aplica	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}e^{\left(\frac{-x^2-2\theta xy+y^2}{2(1-\theta^2)}\right)}dxdy$	$[-1,1]$

Nota: Φ^{-1} inversa da distribuição de probabilidade Nomal Padronizada.



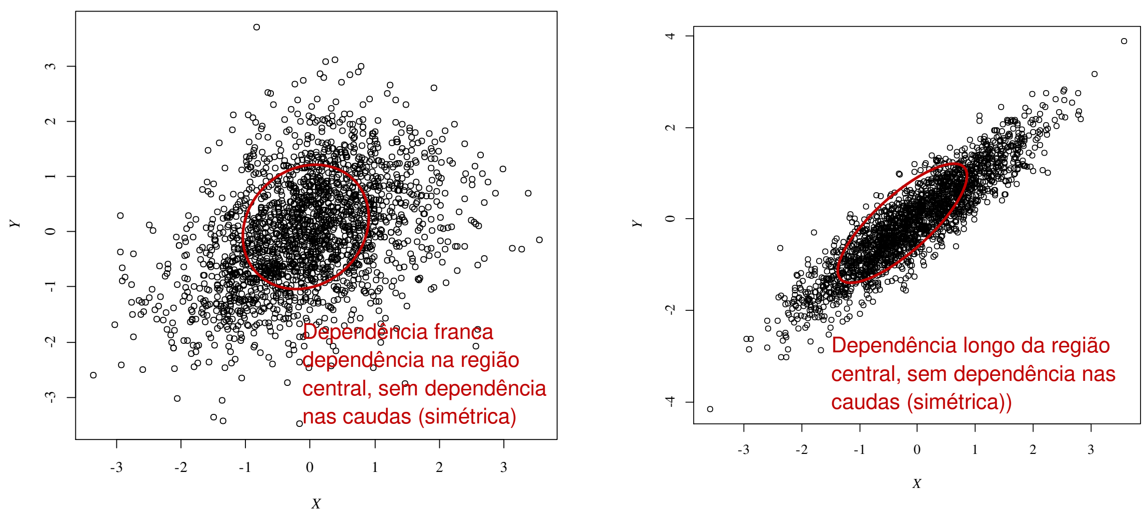
(a) (b)

Figura 3.4.1. Variáveis aleatórias com distribuição de probabilidade Normal Padronizada simuladas com as cópulas: (a) JOE com $\theta = 5$; (b) GUM com $\theta = 5$.



(a) (b)

Figura 3.4.2. Variáveis aleatórias com distribuição de probabilidade Normal Padronizada simuladas com as cópulas: (a) CLA com $\theta = 5$; (b) FRA com $\theta = 10$.



(a) (b)

Figura 3.4.3. Variáveis aleatórias com distribuição de probabilidade Normal Padronizada simuladas com as cópulas: (a) AMH com $\theta = 0{,}9$; (b) GAU com $\theta = 0.9$.

As cópulas CLA, JOE e GUM possuem estrutura assimétrica em relação a sua diagonal secundária, apresentando forte concentração de probabilidade em uma das suas extremidades (caudas superior ou inferior). No caso das cópulas JOE e

GUM, percebe-se forte estrutura de dependência na cauda superior. No caso da cópula CLA, a forte estrutura de dependência é observada na cauda inferior.

As cópulas GAU, FRA e AMH são simétricas em relação a sua diagonal secundária, isto é, não apresentam uma concentração evidente de probabilidade em suas extremidades (cauda superior e inferior). Considerando uma mesmo valor de θ, a cópula AMH registra franca estrutura dependência na região central quando comparada com a cópula GAU, que apresenta uma estrutura mais forte de dependência ao longo de sua diagonal principal. Desta forma, a cópula AMH pode ser utilizada para capturar estruturas com fraca entre as variáveis

As cópulas com dependência nas caudas são as mais indicadas para modelar a relação entre variáveis aleatórias quando o estudo tem interesse, principalmente, nos eventos extremos.

Na Tabela 3.4.2, são apresentados as condições do parâmetro θ, considerando os intervalos apresentados na Tabela 3.4.1, para as quais ocorrem as situações particulares de dependência perfeita e a independência entre as variáveis aleatórias.

Na Tabela 3.4.3, são apresentadas as principais características das cópulas Arquimedianas JOE, GUM, CLA, FRA e AMH, e Elíptica GAU quanto à assimétrica e a dependência das variáveis aleatórias nas caudas.

Tabela 3.4.2. Tipos de dependência entre as variáveis aleatórias para as cópulas JOE, GUM, CLA, FRA, AMH e GAU em função do parâmetro θ.

Cópula	Dependência perfeita	Independência
JOE	$\theta \to \infty$	$\theta \to 1$
GUM	$\theta \to \infty$	$\theta \to 1$
CLA	$\theta \to \infty$	$\theta \to 0$
FRA	$\theta \to \pm\infty$	$\theta \to 0$
AMH	$(*)$	$\theta \to 0$
GAU	$\theta \to \pm 1$	$\theta \to 0$

Nota: $(*)$ A dependência perfeita não é alcançada, mas aumenta para $\theta \to \pm 1$.

Tabela 3.4.3. Características das cópulas JOE, GUM, CLA, FRA e AMH e GAU quanto à assimétrica e dependência nas caudas.

Cópula	Simetria	Dependência nas caudas
JOE	Assimétrica	Superior
GUM	Assimétrica	Superior
CLA	Assimétrica	Inferior
FRA	Simétrica	Sem dependência
AMH	Simétrica	Sem dependência
GAU	Simétrica	Sem dependência

Intuitivamente, as cópulas assimétricas com dependência na cauda superior são mais adequadas para representar os eventos de seca.

No Apêndice B são apresentadas as funções do ambiente computacional e linguagem de programação R para preparação dos dados para ajuste das cópulas (B.7), ajustes das cópulas (B.8), cálculo do AIC (B.9) e teste GOF (B.10).

3.5. PERÍODO DE RETORNO DOS EVENTOS DE SECA

O período de retorno, também denominado de período de recorrência ou tempo de recorrência, pode ser definido como o intervalo de tempo estimado para a ocorrência de um determinado evento com características específicas. Nos estudos de seca, um conceito amplamente utilizado para o período de retorno é o período de tempo médio entre a ocorrência de eventos com durações e severidades maiores ou iguais a valores de referência. Shiau e Shen (2001) apresentaram uma expressão matemática teórica para o período de retorno das severidades dos eventos de seca. Os mesmos conceitos foram aplicados por Bonaccorso et al. (2003) para as durações.

Dada uma série temporal de vazões $Q = \{q_t, t \in [0, T]\}$ e uma vazão de referência q_0, sendo o vetor de variáveis aleatórias $\boldsymbol{X} = (X_1, X_2)$, que correspondem às durações e severidades dos eventos de seca, respectivamente, com distribuições de probabilidade contínua F_1 e F_2, tal que $X_1 = (x_{11}, \cdots, x_{1n})$ e $X_2 = (x_{21}, \cdots, x_{2n})$. Na Figura 3.5.1 são apresentados:

1) O número de eventos de seca até a ocorrência do próximo evento com duração maior ou igual a um valor arbitrário x_1, denominado de N; e
2) Os intervalos de tempo entre inícios de eventos de seca consecutivos, denominado de $L = \{l_i, i \in [1,2,\cdots,N]\}$; e
3) O intervalo de tempo entre inicios de dois eventos de seca sucessivos com durações maiores ou iguais a x_1, denominado de $Z = \{z_j, j \in [1,2,\cdots,N]\}$.

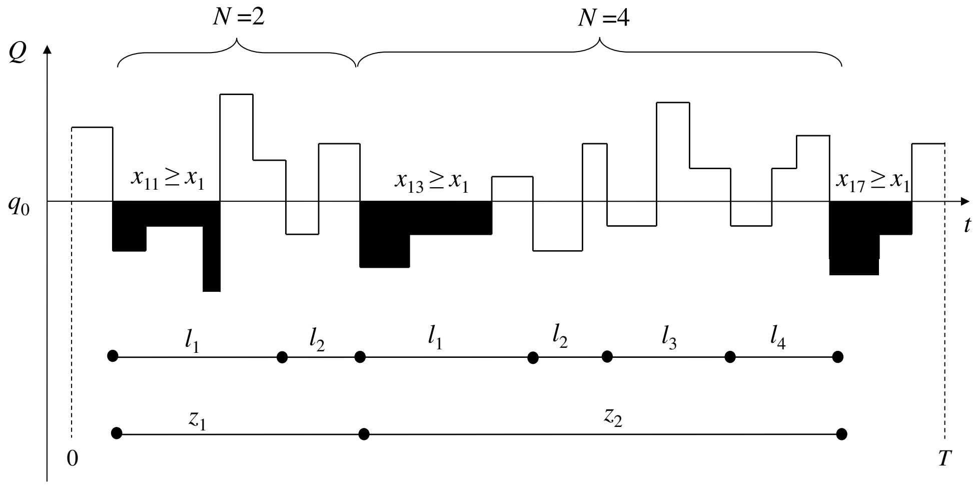


Figura 3.5.1. Intervalos de tempo entre eventos de seca para $X_1 \geq x_1$. Adaptado de Bonaccorso et al. (2003).

Tendo como base a Figura 3.5.1, o intervalo de tempo entre eventos de seca com durações maiores ou iguais a x_1 pode ser expresso como:

$$Z = \sum_{i=1}^{N} l_i \quad (3.5.1)$$

Com base na Equação (3.5.1), o período de retorno das durações dos eventos de seca, denominado de T_{X_1}, pode ser expresso pelo valor esperado de Z, onde:

$$T_{X_1} = E[Z] = E[\sum_{i=1}^{N} l_i] = E[N] \cdot E[l_i] \quad (3.5.2)$$

Assumindo que os intervalos de tempo L possuem a mesma distribuição de probabilidade, a Equação (3.5.2) pode ser escrita como:

$$T_{X_1} = E[Z] = E[N] \cdot E[L] \tag{3.5.3}$$

Obviamente, N é uma varável aleatória que depende da distribuição de probabilidade das durações. Sendo F_1 a distribuição de probabilidade das durações, a probabilidade dos eventos de seca com durações maiores ou iguais a x_1, $P(X_1 \geq x_1)$, pode ser expressa como $1 - F_1(x_1)$ como apresentado na Figura 3.5.2.

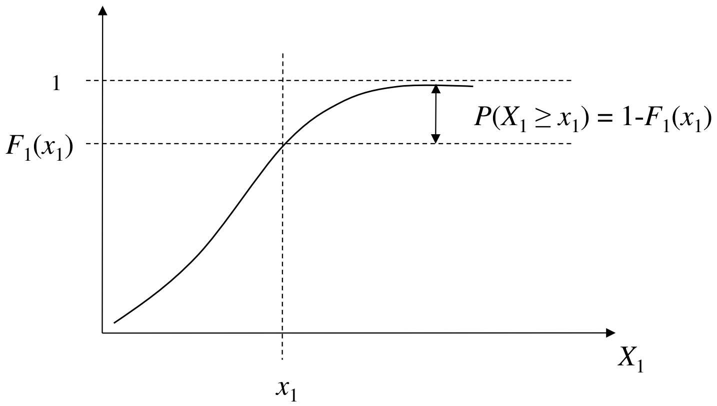


Figura 3.5.2. Representação da probabilidade dos eventos de seca com $X_1 \geq x_1$.

Considerando dois eventos de secas sucessivos com durações maiores ou iguais x_1, existem entre eles $N - 1$ eventos com durações menores que x_1. Assim, o número de eventos de secas com duração maiores ou iguais a pode ser representado por uma distribuição de probabilidade Geométrica:

$$P(N = n) = [1 - F_1(x_1)] \cdot F_1(x_1)^{n-1} \tag{3.5.4}$$

Assim, o valor esperado de N é expresso como:

$$E[N] = \frac{1}{1 - F_1(x_1)} \tag{3.5.5}$$

Com base nas Equações (3.5.3) e (3.5.5), o período de retorno dos eventos de seca com durações maiores ou iguais a x_1, pode ser expresso tendo-se o conhecimento prévio do valor esperado do tempo entre inícios de eventos de seca consecutivos e da distribuição de probabilidade das durações:

$$T_{X_1} = \frac{E[L]}{P(X_1 \geq x_1)} = \frac{E[L]}{1 - F_1(x_1)} \tag{3.5.6}$$

De maneira similar, pode-se obter o período de retorno das severidades dos eventos de seca, onde $W = \{w_j, j \in [1,2,\cdots,N]\}$ representa o intervalo de tempo entre inícios de dois eventos de seca sucessivos com severidades maiores ou iguais a x_2.

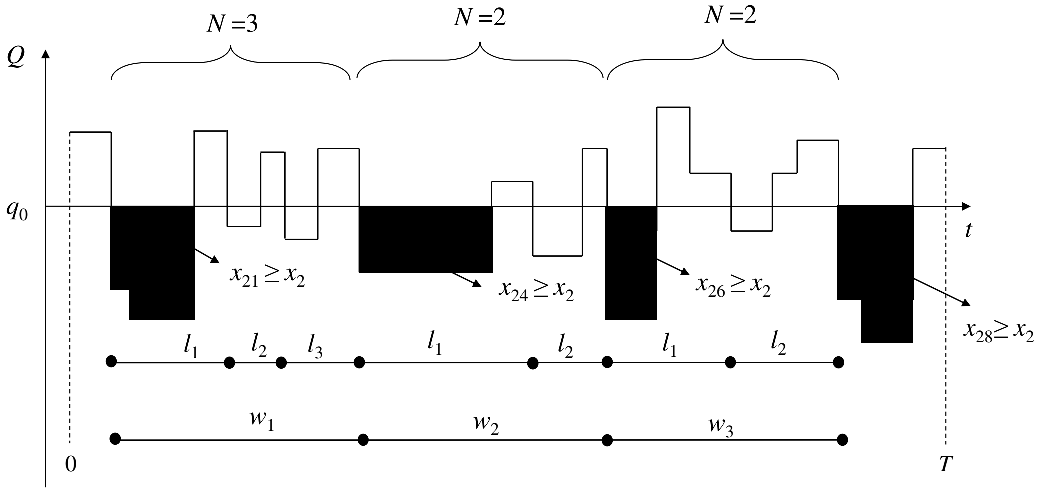


Figura 3.5.2. Intervalo de tempo entre eventos de seca com $X_2 \geq x_2$. Adaptado de Shiau e Shen (2001).

Com base na Figura 3.5.2, o W pode ser expresso como:

$$W = \sum_{i=1}^{N} l_i \tag{3.5.7}$$

Com base na Equação (3.5.7), o período de retorno das severidades dos eventos de seca, denominado de T_{X_2}, pode ser expresso pelo valor esperado de W, onde:

$$T_{X_2} = E[W] = E[\sum_{i=1}^{N} l_i] = E[N] \cdot E[l_i] \quad (3.5.8)$$

De maneira análoga às durações, assumindo que os intervalos de tempo L possuem a mesma distribuição de probabilidade, a Equação (3.5.8) pode ser escrita como:

$$T_{X_2} = E[W] = E[N] \cdot E[L] \quad (3.5.9)$$

Considerando que entre dois eventos de secas sucessivos com severidades maiores ou iguais x_2, existem $N - 1$ eventos com severidades menores que x_2.

Logo, o número de eventos de secas com severidades maiores ou iguais a x_2 pode ser representado por uma distribuição de probabilidade Geométrica:

$$P(N = n) = [1 - F_2(x_2)] \cdot F_2(x_2\,)^{n-1} \quad (3.5.10)$$

sendo F_2 a distribuição de probabilidade das severidades, $F_2(x_2\,)$ a probabilidade dos eventos de seca terem severidades menores que x_2 e 1-$F_2(x_2)$ a probabilidade dos eventos de secas terem severidades maiores ou iguais a x_2 (Figura 3.5.3).

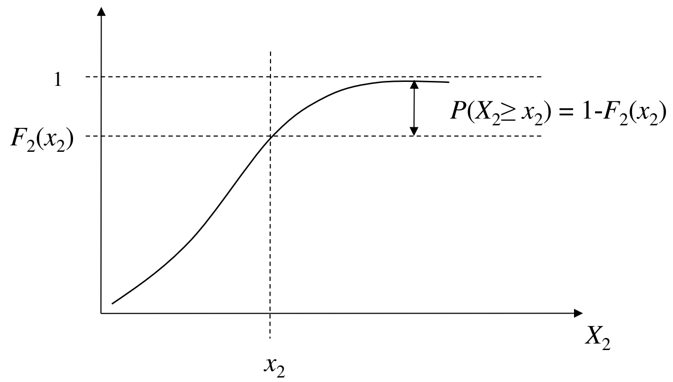


Figura 3.5.3. Representação da probabilidade dos eventos de seca com $X_2 \geq x_2$.

O valor esperado de N é expresso como:

$$E[N] = \frac{1}{1 - F_2(x_2)} \quad (3.5.11)$$

Logo, com base nas Equações (3.5.9) e (3.5.11) têm-se:

$$T_{X_2} = \frac{E[L]}{P(X_2 \geq x_2)} = \frac{E[L]}{1 - F_2(x_2)} \quad (3.5.12)$$

sendo T_{X_2} o período de retorno dos eventos de seca com severidades maiores ou iguais a x_2, F_2 a distribuição de probabilidade das severidades e $P(X_2 \geq x_2)$ a probabilidade dos eventos de seca com severidades maiores ou iguais a x_2.

A análise separada das variáveis aleatórias não revela a dependência entre elas. Assim, a partir dos casos uni-variados dos períodos de retorno, Shiau (2003) apresentou uma expressão analítica para o período de retorno conjunto das durações e severidades dos eventos de seca para valores de $X_1 \geq x_1$ e $X_2 \geq x_2$, na qual várias combinações de valores de X_1e X_2 podem resultar em um mesmo período de retorno:

$$\begin{aligned} T_{X_1X_2} &= \frac{E[L]}{P(X_1 \geq x_1, X_2 \geq x_2)} = \frac{E[L]}{1 - F_{X_1}(x_1) - F_{X_2}(x_2) + F_{X_1X_2}(x_1, x_2)} \\ &= \frac{E[L]}{1 - F_{X_1}(x_1) - F_{X_2}(x_2) + C\big(F_{X_1}(x_1), F_{X_2}(x_2)\big)} \end{aligned} \quad (3.5.13)$$

sendo $T_{X_1X_2}$ o período de retorno conjunto das durações e severidades dos eventos de seca, $P(X_1 \geq x_1, X_2 \geq x_2)$ a probabilidade das durações e severidades serem maiores ou iguais a x_1 e x_2, respectivamente, $F_{X_1X_2}$ a distribuição de probabilidade conjunta da durações e severidades e C a cópula que associa as distribuições de probabilidade das durações e severidades.

4. RESULTADOS E DISCUSSÕES

4.1. EVENTOS OBSERVADOS

Na Tabela 4.1.1, são apresentadas as principais características dos eventos de seca observados do Rio Camanducaia, obtidas com base na teoria dos runs, a partir dos dados históricos de vazões médias mensais da estação fluviométrica 3D-002, para cada nível de truncamento da Tabela 3.2.2.

Os valores apresentados na Tabela 4.1.1 correspondem ao número de eventos de seca, suas durações e severidades média, máxima e mínima, e as medidas de correlação entre durações e severidades calculadas a partir do τ de Kendall, ρ de Spearman e coeficiente de correlação de Pearson (r). O coeficiente r mede o grau de associação linear entre duas variáveis, enquanto que o ρ de Spearman mede o grau de associação não linear entre variáveis monótonas. O τ de Kendall é uma alternativa de medida de correlação ao ρ de Spearman.

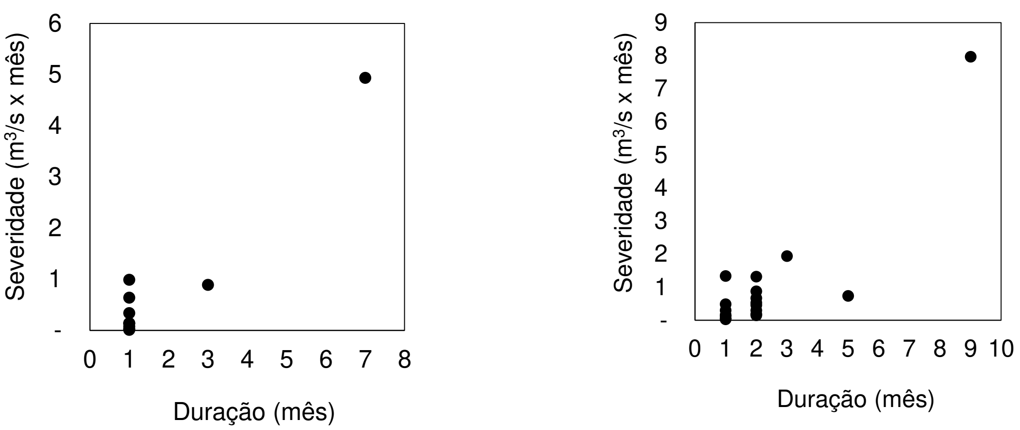
Tabela 4.1.1. Principais características dos eventos de seca observados do Rio Camanducaia.

Características dos eventos de seca		**Nível de truncamento**			
		$\boldsymbol{P98}$	$\boldsymbol{P95}$	$\boldsymbol{P92}$	$\boldsymbol{P90}$
Número de eventos		10	21	37	40
Duração (mês)	Méd	1,80	2,05	1,89	2,20
	Máx	7	9	9	9
	Mín	1	1	1	1
Severidade (m^3/s x mês)	Méd	0,82	0,85	1,07	1,41
	Máx	4,94	7,97	11,30	13,19
	Mín	0,01	0,03	0,08	0,02
τ de Kendall		0,55	0,58	0,78	0,63
ρ de Spearman		0,62	0,67	0,77	0,78
r		0,96	0,89	0,87	0,86

Como pode-se observar, à medida que se eleva o valor do nível de truncamento, eleva-se o número de ocorrências de eventos e os respectivos valores esperados das severidades. Para as durações, no que diz respeito aos seus valores esperados, estes também tendem aumentar com a elevação do nível de truncamento, mas particularmente para o nível de truncamento $P92$ há uma redução do valor esperado, pois com a alteração de $P95$ para $P92$ ocorre um aumento da soma das durações de 63%, enquanto que o aumento do número de eventos é de 76%. Este fato, faz com que o valor esperado das durações para o nível de truncamento $P92$ seja inferior ao $P95$.

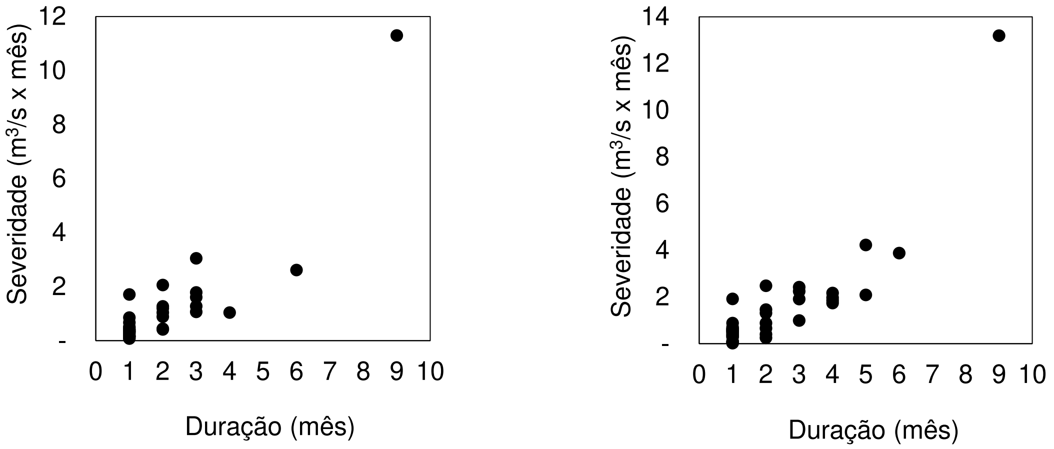
Na Tabela 4.1.1, observa-se também que os valores dos coeficientes r, ρ de Spearman e τ de Kendall indicam um grau de associação positiva entre as durações e severidade, onde o coeficiente r tende a diminuir com a elevação do nível de truncamento, enquanto que ρ de Spearman apresenta comportamento oposto.

Nas Figuras 4.1.1 e 4.1.2, são apresentados os gráficos de dispersão das durações e severidades dos eventos de seca observados do Rio Camanducaia para cada nível de truncamento. Como pode-se observar, tendo como base a série histórica de vazões, o número obtido de eventos de seca é restrito, variando de 10 a 40 eventos a depender no nível de truncamento utilizado.



(a) **(b)**

Figura 4.1.1. Eventos de seca observados no Rio Camanducaia para a estação fluviométrica 3D-002 considerando os níveis de truncamento: (a) $P98$; (b) $P95$.



(a) **(b)**

Figura 4.1.2. Eventos de seca observados no Rio Camanducaia para a estação fluviométrica 3D-002 considerando os níveis de truncamento: (a) $P92$; (b) $P90$.

Os comportamentos das durações e severidades observados nas Figuras 4.1.1 a 4.1.2, corroboram como o valores dos coeficientes r e ρ de Spearman apresentados na Tabela 4.1.1, indicando que, com a elevação do nível de truncamento o grau de associação entre as durações e severidades pode não ser linear.

Os testes de Box-Pierce e Ljung foram utilizados para verificar a hipótese de independência das observações em série para as durações e severidades e na Tabela 4.1.2 são apresentados os seus p-valores.

Tabela 4.1.2. p-Valores dos testes de Box-Pierce e Ljung para as observações e série das durações e severidades dos eventos de seca observados do Rio Camanducaia.

Variável	**Teste**	**Nível de truncamento**			
		$\boldsymbol{P98}$	$\boldsymbol{P95}$	$\boldsymbol{P92}$	$\boldsymbol{P90}$
Duração	Box-Pierce	0,57	0,67	0,82	0,91
	Ljung	0,52	0,65	0,81	0,91
Severidade	Box-Pierce	0,51	0,60	0,78	0,77
	Ljung	0,44	0,68	0,77	0,77

Como pode-se observar, todos os p-valores são maiores que 0,05, indicando que a hipótese de independência das observações em série não pode ser rejeitada para um nível de significância de 95%.

Desta forma, as alternativas propostas por Zelenhasic e Salvai (1987) para tratar a condição de eventos mutualmente dependentes não são necessárias.

4.2. ANALISE UNI-VARIADA DAS DURAÇÕES E SEVERIDADES

Nas Figuras 4.2.1 a 4.2.4, são apresentadas as comparações entre as Funções de Distribuição Acumulada (CDF) teóricas (GAM2, EXP, WBL, LOGN, GP e GEV) e empíricas, calculadas a partir das durações e severidades observadas dos eventos de seca do Rio Camanducaia.

Como pode-se observar principalmente na Figura 4.2.1(a), as CDF teóricas aparentemente não são tão próximas da CDF empírica, podendo indicar pouca adequação aos dados observados.

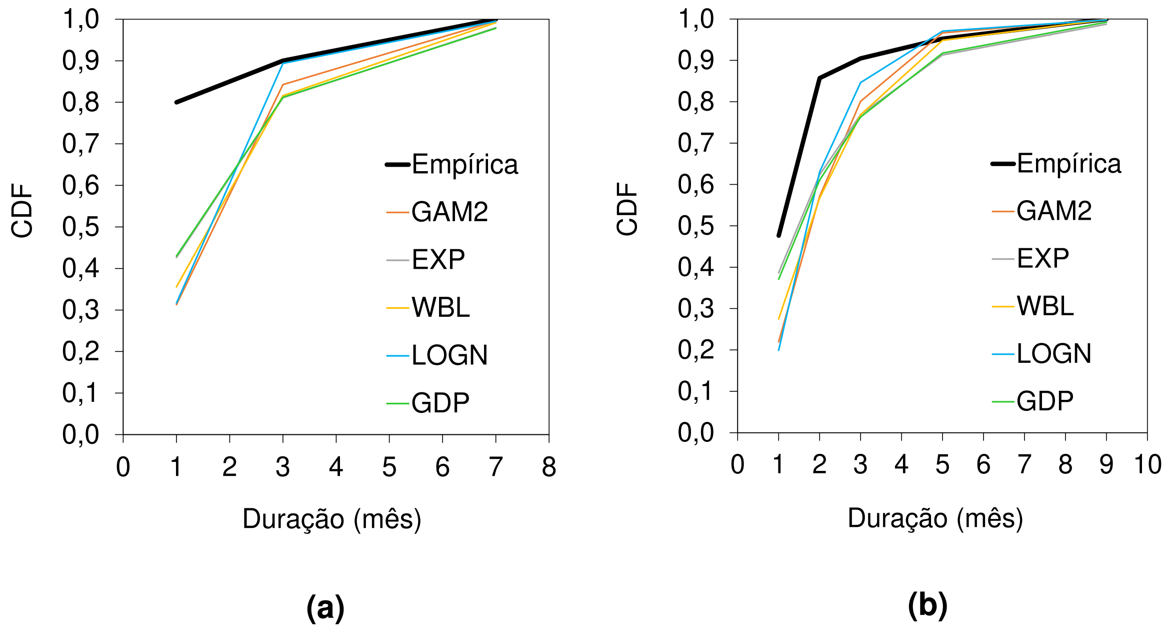
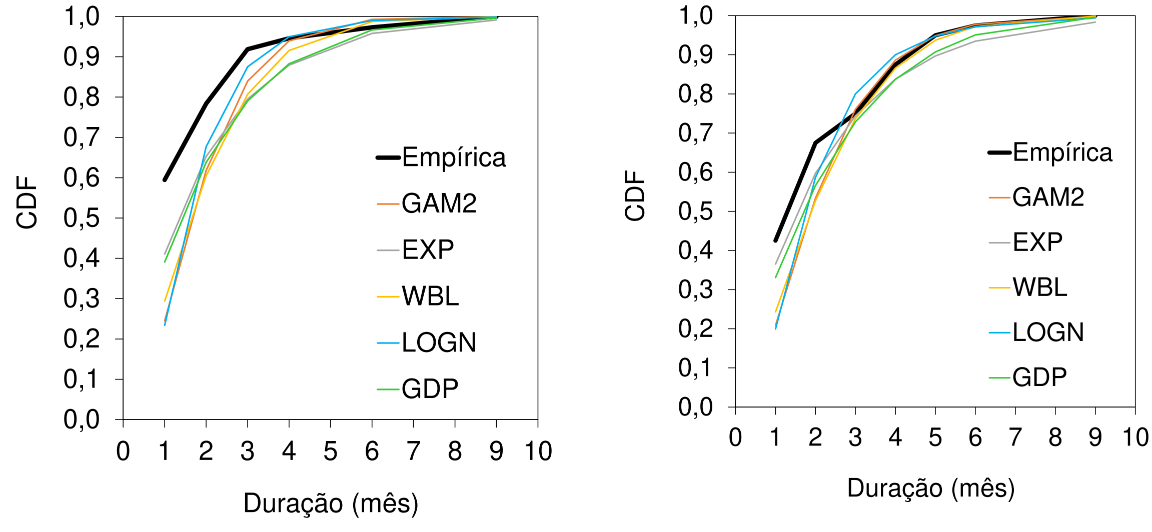
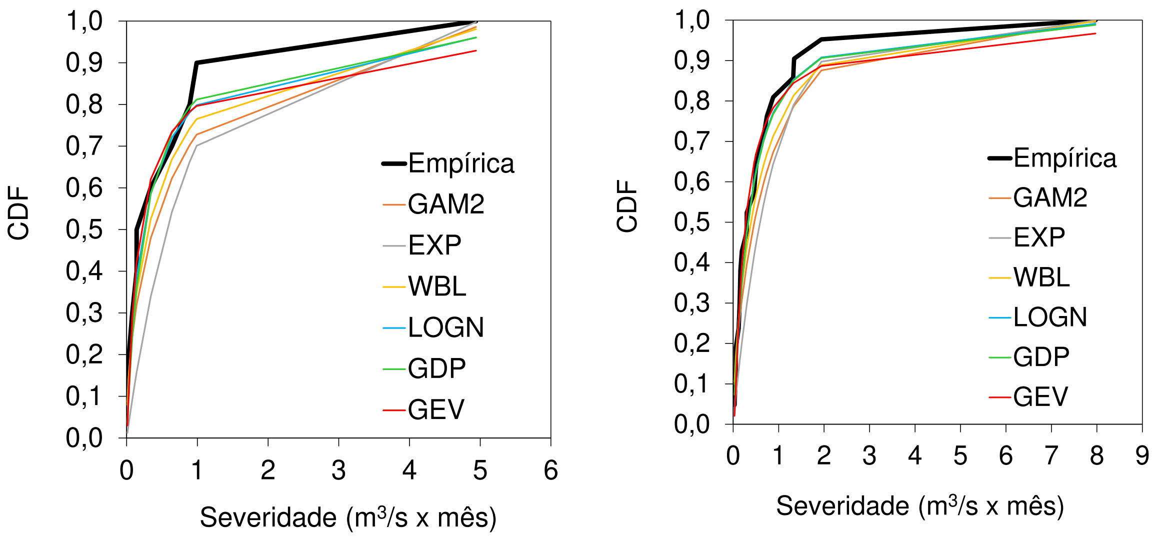


Figura 4.2.1. CDF das durações observadas dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P98$; (b) $P95$.



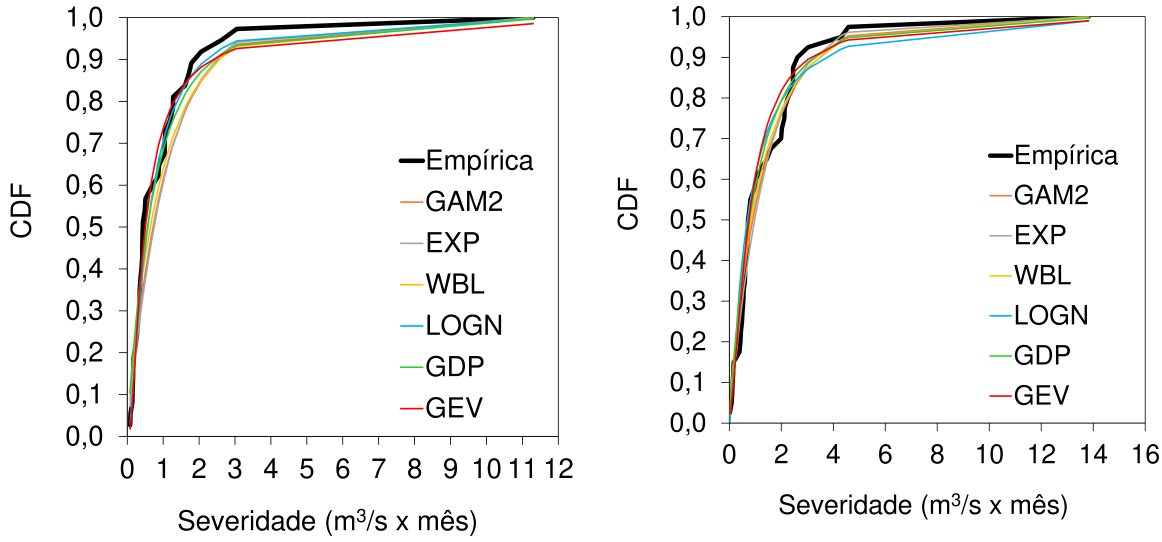
(a) (b)

Figura 4.2.2. CDF das durações observadas dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P92$; (b) $P90$.



(a) (b)

Figura 4.2.3. CDF das severidades observadas dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P98$; (b) $P95$.



(a) **(b)**

Figura 4.2.4. CDF das severidades observadas dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P92$; (b) $P90$.

Para as durações, especificamente para a distribuição de probabilidade GEV não foi possível obter um estimador de máximo verossimilhança para os quatro níveis de truncamento. Este fato pode estar associado a natureza discreta dos dados.

As Figuras 4.2.3 e 4.2.4 mostram que as CDF teóricas e empíricas das severidades são razoavelmente próximas, indicando que as seis distribuições de probabilidade teóricas podem representar adequadamente o comportamento probabilístico das severidades para os quatro níveis de truncamento analisados.

Nas Tabelas 4.2.1 e 4.2.2, são apresentados os valores do AIC e os p-valores do teste KS para cada distribuição de probabilidade aplicada às durações e severidades dos eventos de seca do Rio Camanducaia.

Exceto para a distribuição de probabilidade GEV, os p-valores do teste KS apresentados na Tabela 4.2.1 mostram que a hipótese de adequação das distribuições de probabilidade GAM2, EXP, WBL, LOGN e GP, aos dados observados das durações, não pode ser rejeitada para o nível de truncamento $P95$. As únicas distribuições de probabilidade cuja hipótese de adequação aos dados observados das durações que não podem ser rejeitadas para os quatro níveis de truncamento são a EXP e GP.

As distribuições de probabilidade teóricas com melhor ajuste às durações dos eventos de seca do Rio Camanducaia são EXP para os níveis de truncamento $P98$, $P92$ e $P90$, e LOGN para o nível de truncamento $P95$.

Tabela 4.2.1. Valores do AIC e p-valores do teste KS para cada distribuição de probabilidade testada para durações observadas dos eventos de seca do Rio Camanducaia.

Distribuição de probabilidade	Nível de truncamento							
	$P98$		$P95$		$P92$		$P90$	
	AIC	KS p-valor	AIC	KS p-valor	AIC	KS p-valor	AIC	KS p-valor
GAM2	33,69	0,01	68,64	0,05	111,34	0,00	134,95	0,00
EXP	**33,76**	**0,08**	74,10	0,18	**123,18**	**0,14**	**145,08**	**0,25**
WBL	34,95	0,03	72,20	0,05	117,47	0,00	139,07	0,00
LOGN	29,47	0,01	**62,21**	**0,06**	100,82	0,00	127,28	0,00
GP	35,75	0,06	75,92	0,09	124,63	0,06	145,49	0,07
GEV	$(*)$	$(*)$	$(*)$	$(*)$	$(*)$	$(*)$	$(*)$	$(*)$

Nota: Os valores em negrito correspondem à distribuição de probabilidade selecionada. $(*)$ Avaliação não foi possível.

Tabela 4.2.2. Valores do AIC e p-valores do teste KS para cada distribuição de probabilidade testada para severidades observadas dos eventos de seca do Rio Camanducaia.

Distribuição de probabilidade	Nível de truncamento							
	$P98$		$P95$		$P92$		$P90$	
	AIC	KS p-valor	AIC	KS p-valor	AIC	KS p-valor	AIC	KS p-valor
GAM2	15,77	0,84	38,50	0,78	83,05	0,20	110,33	0,52
EXP	18,03	0,16	37,29	0,19	81,27	0,13	109,25	0,25
WBL	14,90	0,95	33,46	0,98	81,50	0,48	109,21	0,76
LOGN	**14,17**	**0,997**	**28,79**	**0,996**	**71,00**	**0,75**	109,13	0,89
GP	14,65	0,84	30,13	0,97	76,58	0,74	**106,75**	**0,84**
GEV	17,37	0,95	31,32	0,89	72,67	0,95	111,28	0,82

Os p-valores do teste KS apresentados na Tabela 4.2.2, indicam que para as severidades dos eventos de seca a hipótese de adequação das seis distribuições de probabilidade avaliadas não pode ser rejeitada para nenhum dos níveis de truncamento, corroborando com a análise visual das Figuras 4.2.3 e 4.2.4.

Assim, de acordo com os valores do AIC apresentados na Tabela 4.2.2, as distribuições de probabilidade com melhor ajuste às severidades dos eventos de seca do Rio Camanducaia são LOGN para os níveis de truncamento $P98$, $P95$ e $P92$, e GP para o nível de truncamento $P90$.

Na Tabela 4.2.3, são apresentadas, para cada nível de truncamento as distribuições de probabilidade e seus parâmetros estimados, selecionadas para representar o comportamento probabilístico uni-variado das durações e severidades dos eventos de seca do Rio Camanducaia.

Tabela 4.2.3. Distribuições de probabilidade e seus parâmetros estimados, selecionadas para representar o comportamento probabilístico das durações e severidades dos eventos de seca do Rio Camanducaia.

Nível de truncamento	**Duração**		**Severidade**	
	Distribuição de probabilidade	**Parâmetros estimados**	**Distribuição de probabilidade**	**Parâmetros estimados**
$P98$	EXP	$\hat{\lambda}$ 0,556	LOGN	$\hat{\mu}$ -1,462 $\hat{\sigma}$ 1,737
$P95$	LOGN	$\hat{\mu}$ 0,498 $\hat{\sigma}$ 0,588	LOGN	$\hat{\mu}$ -1,127 $\hat{\sigma}$ 1,347
$P92$	EXP	$\hat{\lambda}$ 0,529	LOGN	$\hat{\mu}$ -0,569 $\hat{\sigma}$ 1,057
$P90$	EXP	$\hat{\lambda}$ 0,455	GP	$\hat{\xi}$ 0 $\hat{\alpha}$ 0,262 $\hat{\beta}$ 1,023

Desta forma, para os quatro níveis de truncamento as distribuições de probabilidade EXP e LOGN são as mais adequadas para representar, uni-variadamente, as durações e severidades dos eventos de seca do Rio Camanducaia, exceto para o nível de truncamento $P90$, onde a distribuição de probabilidade GP se mostrou mais adequada para representar as severidades.

4.3. ANALISE CONJUNTA DAS DURAÇÕES E SEVERIDADES

A Tabela 4.1.1 apresenta as principais características das durações e severidades dos eventos de seca do Rio Camanducaia. Os valores do τ de Kendall, ρ de Spearman e r indicam significante nível de dependência entre as durações e severidades, sugerindo sua modelagem de forma conjunta.

Tendo como base as distribuições de probabilidade uni-variadas da Tabela 4.2.3, as cópulas apresentadas na Tabela 3.4.1 foram testadas para representar conjuntamente o comportamento probabilístico das durações e severidades dos eventos de seca do Rio Camanducaia.

Na Tabela 4.3.1, são apresentados os valores do AIC e os p-valores do teste GOF para cada cópula da Tabela 3.4.1.

Tabela 4.3.1. Valores do AIC e p-valores do teste GOF para cada cópula testada na análise bivariada das durações e severidades dos eventos de seca do Rio Camanducaia.

Cópula	**Nível de truncamento**							
	$P98$		$P95$		$P92$		$P90$	
	AIC	**GOF p-valor**	**AIC**	**GOF p-valor**	**AIC**	**GOF p-valor**	**AIC**	**GOF p-valor**
CLA	-1,02	($*$)	-11,04	0,10	-9,04	0,00	-12,16	0,00
FRA	-4,52	0,28	-12,96	0,34	-24,82	0,47	-37,62	0,10
GUM	-7,04	0,79	-14,96	0,88	-29,33	0,35	-41,14	0,36
JOE	**-8,14**	**0,90**	-12,74	0,45	**-29,85**	**0,17**	**-43,20**	**0,28**
AMH	-1,65	0,10	-9,80	0,03	-9,32	0,00	-12,31	0,00
GAU	-5,16	0,36	**-15,89**	**0,56**	-25,79	0,31	-33,44	0,17

Nota: Os valores em negrito correspondem às cópulas selecionadas. ($*$) Avaliação não foi possível.

De acordo com os p-valores do teste GOF, a hipótese de adequação das cópulas FRA, GUM, JOE e GAU às durações e severidades dos eventos de seca do Rio Camanducaia, não pode ser rejeitada para os quatro níveis de truncamento analisados. A hipótese de adequação das cópulas CLA e AMH não pode ser rejeitada somente para os níveis de truncamento $P95$ e $P98$, respectivamente.

De forma geral, pode-se dizer que as cópulas assimétricas com dependência na calda superior do tipo JOE e GUM, e as cópulas simétricas sem dependência nas caudas do tipo FRA e GAU, mostram-se mais adequadas na representação conjunta das durações e severidades dos eventos de seca do Rio Camanducaia.

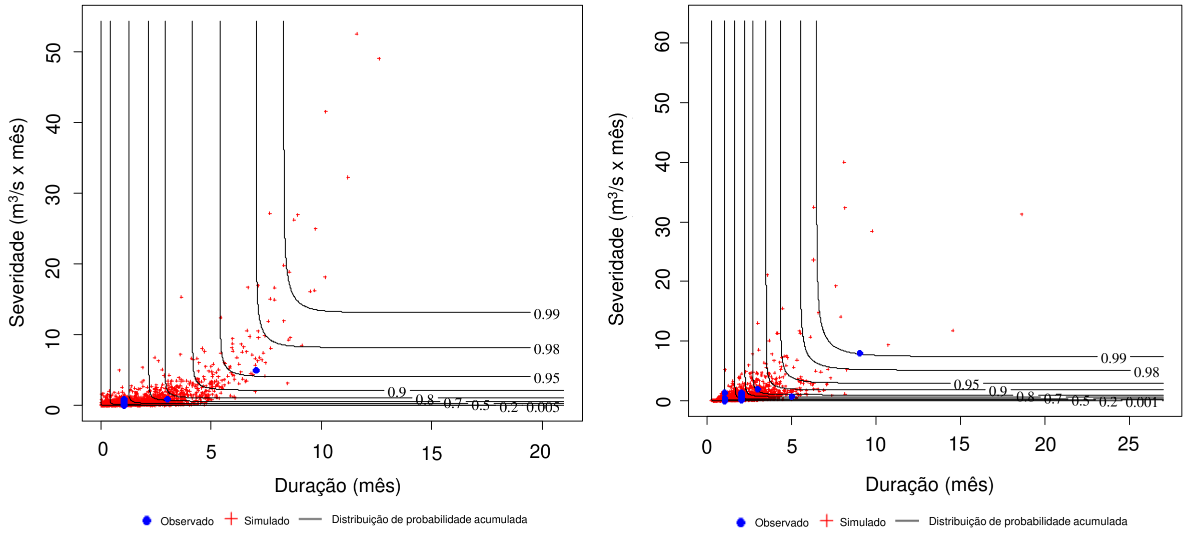
Os valores do AIC da Tabela 4.3.1 indicam que as cópulas com melhor ajuste para representar conjuntamente as durações e severidades dos eventos de seca do Rio Camanducaia são JOE, para os níveis de truncamento $P98$, $P92$ e $P90$, e GAU para o nível de truncamento $P95$.

Na Tabela 4.3.2 são apresentadas as cópulas selecionadas e seus respectivos parâmetros estimados.

Tabela 4.3.2. Cópulas e parâmetros estimados, utilizadas na representação conjunta do comportamento probabilístico das durações e severidades dos eventos de seca do Rio Camanducaia.

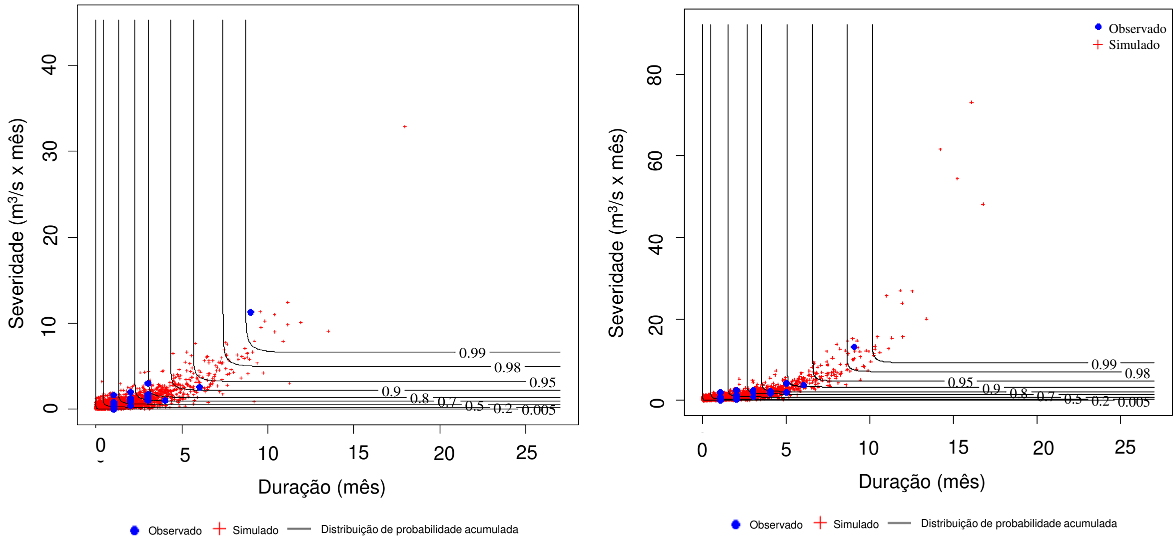
Nível de truncamento	Cópula selecionada	Parâmetro estimado ($\hat{\theta}$)
$P98$	JOE	2,772
$P95$	GAU	0,729
$P92$	JOE	2,895
$P90$	JOE	3,735

Nas Figuras 4.3.1 e 4.3.2, são apresentadas as durações e severidades observadas dos eventos de seca do Rio Camanducaia, suas distribuições de probabilidade conjunta, representada pelas isolinhas, e os 2.000 eventos simulados (pseudo-observações) a partir das distribuições de probabilidade uni-variadas da Tabela 4.2.3 e das cópulas da Tabela 4.3.2. Como pode-se observar, o acoplamento das distribuições de probabilidade uni-variadas das durações e severidades permite a geração de um maior número de eventos de seca, contendo uma combinação maior de pares duração-severidade em comparação aos valores observados obtidos a partir da série histórica de vazões médias mensais.



(a) (b)

Figura 4.3.1. Eventos de seca observados e simulados para o Rio Camanducaia e distribuição de probabilidade conjunta (isolinhas) para os níveis de truncamento: (a) $P98$; (b) $P95$.



(a) (b)

Figura 4.3.2. Eventos de seca observados e simulados para o Rio Camanducaia e distribuição de probabilidade conjunta (isolinhas) para os níveis de truncamento: (a) $P92$; (b) $P90$.

É importante ressaltar que o limitado número de observações de eventos de seca obtido a partir dos dados históricos é uma barreira para a mensuração dos riscos de falhas no abastecimento de água para uso humano e outros.

Para os níveis de truncamento analisados, a maioria dos eventos de seca observados com base nos dados históricos possuem duração de um a dois meses, representando de 70% a 86% dos eventos. Este fato pode levar os tomadores de decisão a uma falsa percepção de segurança hídrica em função dos pares duração-severidade.

Por exemplo, para o nível de truncamento $P98$, tendo como base os dados observados a partir do histórico de vazões 1944 a 2016, 80% dos eventos de seca tem duração de um a dois meses, com somente dois eventos com durações de três e sete meses e severidades de 0,89 e 4,94 m^3/s x mês. Entretanto, considerando a amostra de eventos simulados (pseudo-observações), um número de 352 eventos pode ocorrer com durações entre três a sete meses, com severidades variando de 0,01 a 16,69 m^3/s x mês. Para o nível de truncamento $P95$, somente três eventos de seca são observados a partir dos dados históricos com durações de três a nove meses, com severidades entre 0,74 e 7,97 m^3/s x mês. Considerando os eventos simulados (pseudo-observações), 292 eventos podem ocorrer com durações de três a nove meses, com severidades variando de 0,12 a 40,01 m^3/s x mês.

Na Tabela 4.3.3, são apresentados os intervalos com 95% de confiança para os valores esperados das durações e severidade dos eventos de seca do Rio Camanducaia, assim como também os intervalos de confiança para as medidas de correlação entre durações e severidades, representadas pelo τ de Kendall e ρ de Spearman.

Os intervalos de confiança foram calculados por simulação de Monte Carlo, considerando 100 amostras, cada uma com 2.000 eventos de secas, obtidos a partir das distribuições de probabilidade uni-variadas da Tabela 4.2.3 e das cópulas da Tabela 4.3.2. O teste de Shapiro-Wilks, de acordo com Royston (1982a), Royston (1982b) e Royston (1995), foi utilizado para verificar a hipótese de normalidade dos valores esperados das durações e severidades, τ de Kendall e ρ de Spearman para as amostras geradas. A hipótese de normalidade não pode ser rejeitada para um nível de significância de 95%, pois os p-valores do teste são todos superiores a 0,05 conforme apresentado na Tabela 4.3.4.

Tabela 4.3.3. Intervalos com 95% de confiança para os valores esperados das durações e severidades, τ de Kendall e ρ de Spearman entre as durações e severidades dos eventos de seca do Rio Camanducaia.

Estatística	**Nível de truncamento**			
	$P98$	$P95$	$P92$	$P90$
Duração média (mês)	1,77-1,87	1,90-2,00	1,81-1,96	2,12-2,28
Severidade média (m^3/s x mês)	0,88-1,28	0,73-0,88	0,94-1,04	1,31-1,45
τ de Kendall	0,46-0,51	0,50-0,54	0,48-0,53	0,57-0,61
ρ de Spearman	0,63-0,70	0,69-0,74	0,66-0,71	0,76-0,80

Tabela 4.3.4. p-Valores dos testes de Shapiro-Wilks para verificar a hipótese de normalidade dos valores esperados das durações e severidades, τ de Kendall e ρ de Spearman das amostras gerada nas simulações de Monte Carlo.

Variável	**Nível de truncamento**			
	$P98$	$P95$	$P92$	$P90$
Duração	0,08	0,26	0,90	0,10
Severidade	0,54	0,64	0,46	0,06
τ de Kendall	0,29	0,18	0,42	0,90
ρ de Spearman	0,14	0,14	0,26	0,58

Na Tabela 4.3.5, são apresentados os valores esperados, máximos e mínimo das durações e severidades dos eventos de seca observados e dos 2.000 eventos simulados, além dos respectivos valores de τ de Kendall e ρ de Spearman entre as durações e severidades.

Comparando as Tabelas 4.3.3 e 4.3.5, observa-se que os valores esperados das durações e severidades observados a partir dos dados históricos, em alguns casos, não estão contidos nos intervalos de confiança calculados a partir das simulações de Monte Carlo. Este fato já era esperado, uma vez que os parâmentos das distribuições de probabilidade foram estimados pelo MLE. Como o MLE busca maximizar a probabilidade dos dados observados que, diferentemente do Método dos Momentos busca equalizar os momentos teóricos e empíricos, é natural que ocorra diferença entre os valores esperados teórico e observados (momento de primeira ordem).

Na Tabela 4.3.6, são apresentados os valores esperados observados e teóricos das durações e severidades dos eventos de seca do Rio Camanducaia. Os valores teóricos foram calculados com base suas expressões da Tabela 3.3.2 correspondentes às distribuições de probabilidade da Tabela 4.2.3.

Tabela 4.3.5. Valores esperado, máximo e mínimo das durações e severidades dos eventos de seca observados e simulados para o Rio Camanducaia e os respectivos valores de τ de Kendall e ρ de Spearman entre as durações e severidades.

Variável		**Estatística**	**Nível de truncamento**			
			$P98$	$P95$	$P92$	$P90$
Duração (mês)	Observada	Méd	1,80	2,05	1,89	2,20
		Máx	7	9	9	9
		Mín	1	1	1	1
	Simulada	Méd	1,77	1,94	1,90	2,25
		Máx	12,61	18,62	17,98	16,77
		Mín	0,0008	0,26	0,0001	0,003
Severidade (m^3/s x mês)	Observada	Méd	0,82	0,85	1,07	1,41
		Máx	4,94	7,97	11,30	13,19
		Mín	0,01	0,03	0,08	0,02
	Simulada	Méd	0,97	0,85	0,99	1,43
		Máx	52,49	40,01	32,87	73,08
		Mín	0,0008	0,001	0,02	0,0001
τ de Kendall	Observado	N/A	0,55	0,58	0,78	0,63
	Simulado	N/A	0,46	0,52	0,49	0,59
ρ de Spearman	Observado	N/A	0,62	0,67	0,77	0,78
	Simulado	N/A	0,64	0,70	0,67	0,78

N/A: Não se aplica.

Tabela 4.3.6. Valores esperado das durações e severidades observados e teóricos (das distribuições de probabilidade) dos eventos de seca do Rio Camanducaia.

Nível de truncamento	Valor esperado das durações (mês)		Valor esperado das severidades (m³/s x mês)	
	Observado	Teórico	Observado	Teórico
$P98$	1,80	1,80	0,82	1,05
$P95$	2,05	1,96	0,85	0,80
$P92$	1,89	1,89	1,08	0,99
$P90$	2,20	2,20	1,41	1,39

Os testes de Box-Pierce e Ljung foram utilizados para verificar a hipótese de independência das observações em série para as durações e severidades simuladas. De acordo com os p-valores, a hipótese de independência dos simulados não pode ser rejeitada para um nível de significância de 95% conforme apresentado na Tabela 4.3.7.

Tabela 4.3.7. p-Valores dos testes de Box-Pierce e Ljung para as observações e série das durações e severidades simuladas dos eventos de seca do Rio Camanducaia.

Variável	Teste	Nível de truncamento			
		$\boldsymbol{P98}$	$\boldsymbol{P95}$	$\boldsymbol{P92}$	$\boldsymbol{P90}$
Duração	Box-Pierce	0,92	0,59	0,28	0,08
	Ljung	0,92	0,59	0,28	0,08
Severidade	Box-Pierce	0,45	0,37	0,97	0,12
	Ljung	0,45	0,37	0,97	0,12

Assim, os resultados mostram que o método apresenta forte potencial para subsidiar a análise de risco de falhas no abastecimento de água, e consequentemente, auxiliar no planejamento dos recursos hídricos para enfrentamento dos eventos de seca em uma bacia hidrográfica sem regularização no sudeste brasileiro.

É importante enfatizar que, a depender do interesse do estudo, pode-se optar pela cópula que melhor se adeque ao tipo de análise desejada. Por exemplo, se o interesse é na simulação de eventos extremos, cópulas assimétricas com dependência nas caudas são mais adequadas. Casos contrário, se os eventos extremos não são relevantes, cópulas simétricas sem dependência na cauda representam a melhor opção.

No Apêndice B são apresentadas as funções do ambiente computacional e linguagem de programação R para simulação de pseudo observações por meio de cópulas (B.11) e teste de Shapiro-Wilks.

Nas Figuras 4.3.7 a 4.3.10, com objetivo de facilitar o entendimento do método, são apresentados os fluxogramas das etapas correspondentes a obtenção dos eventos de seca, analises uni e bivariada e geração de eventos. Os fluxogramas mostram de forma macro os principais procedimentos adotados em cada etapa do método, cujas explicações foram detalhadas ao longo do trabalho.

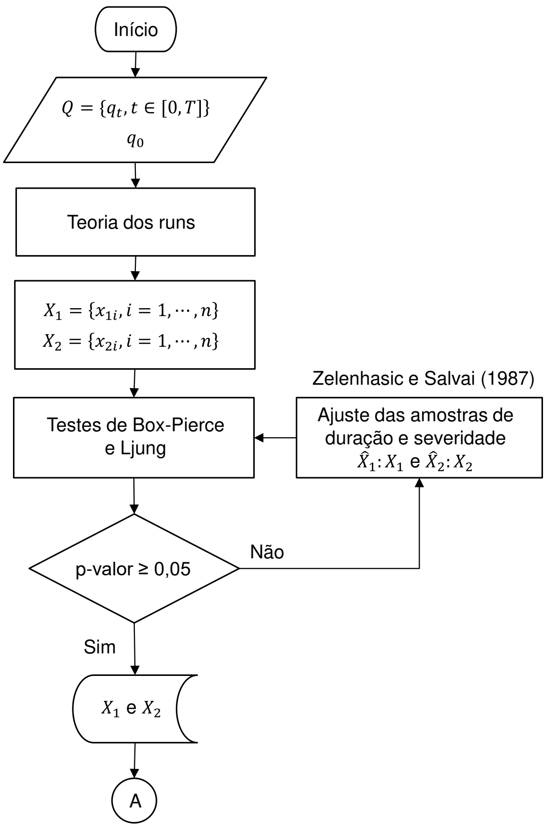


Figura 4.3.7. Fluxograma da etapa de obtenção dos eventos de seca.

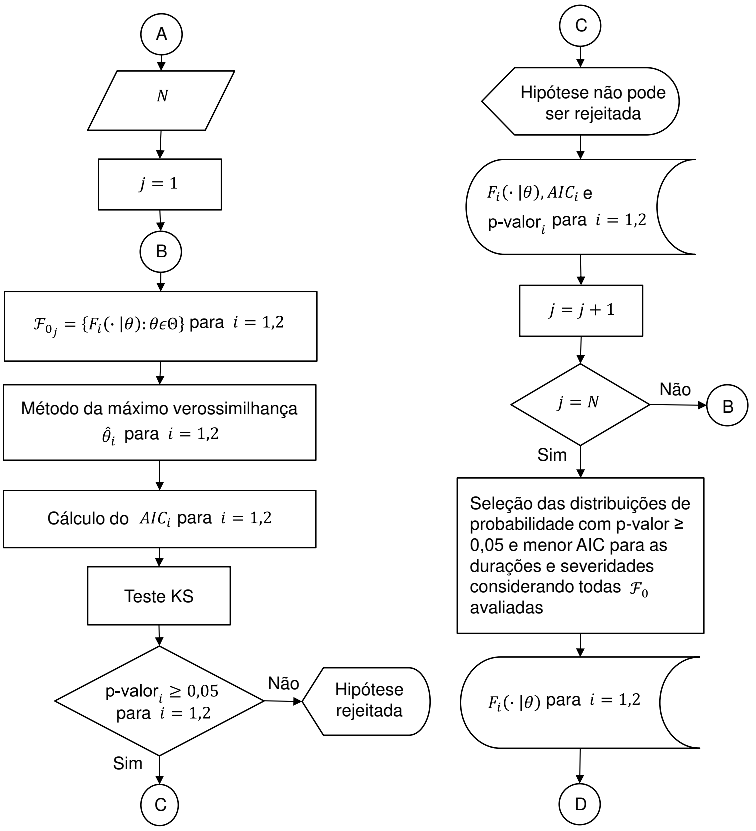


Figura 4.3.8. Fluxograma da etapa de análise uni-variada dos eventos de seca.

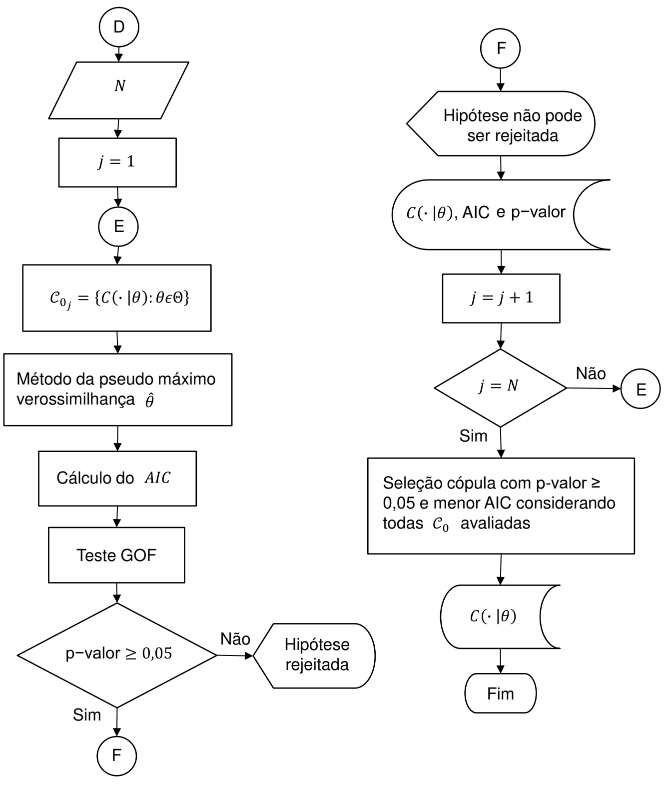


Figura 4.3.9. Fluxograma da etapa de análise bivariada dos eventos de seca.

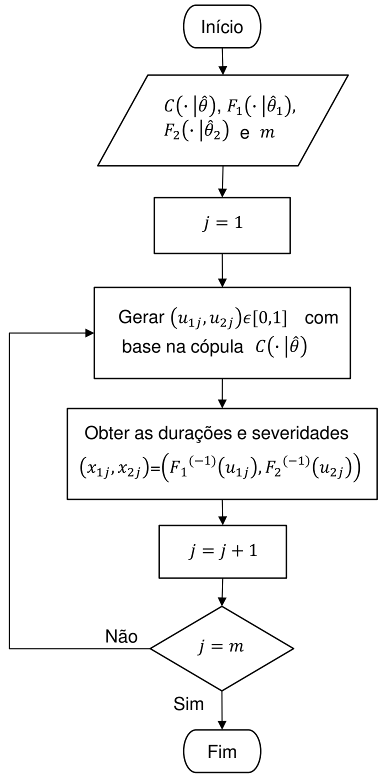


Figura 4.3.10. Fluxograma da etapa de geração de eventos de seca.

4.4. ANALISE DO PERÍODO DE RETORNO DAS DURAÇÕES E SEVERIDADES

Na Tabela 4.4.1, são apresentados os valores esperados observados do intervalo de tempo entre inícios de eventos de seca consecutivos (L) para o Rio Camanducaia, calculados com base na série histórica de vazões médias mensais. Como pode-se observar, o valor esperado de L reduz com o aumento do nível de truncamento indicando uma maior frequência de ocorrência dos eventos de seca.

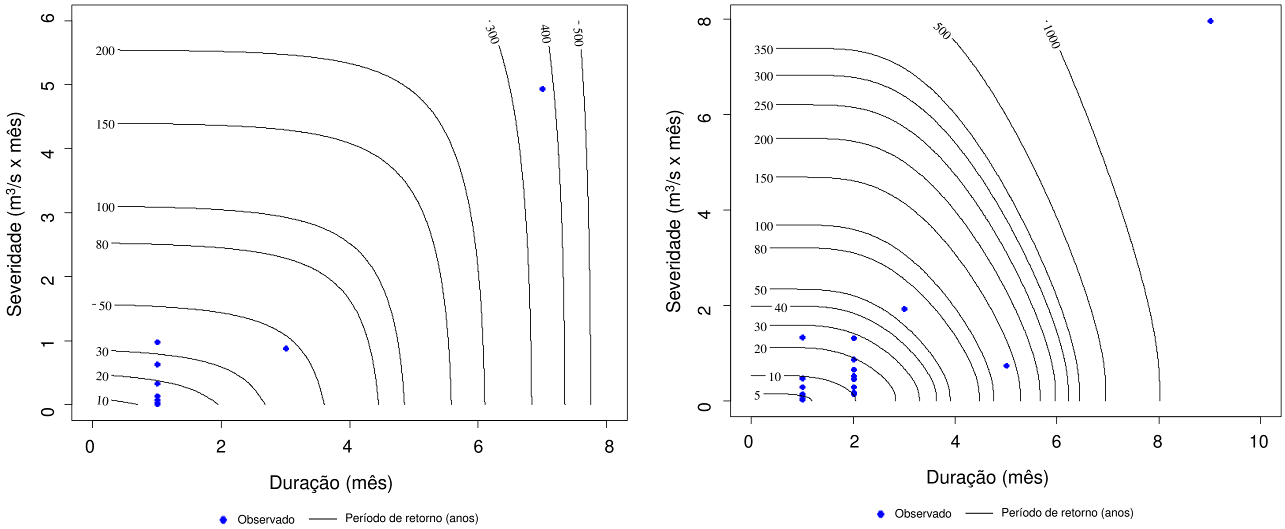
Tabela 4.4.1. Valor esperado de L para os eventos de seca do Rio Camanducaia.

Nível de truncamento	$\boldsymbol{E[L]}$ **(anos)**
$P98$	6,75
$P95$	3,75
$P92$	1,98
$P90$	1,83

Nas Figuras 4.4.1 e 4.4.2, são apresentados os períodos de retorno dos eventos de seca do Rio Camanducaia, expressos em anos (isolinhas), calculados com base na Equação (3.5.13), nas distribuições de probabilidade da Tabela 4.2.3 e nas cópulas da Tabela 4.3.2.

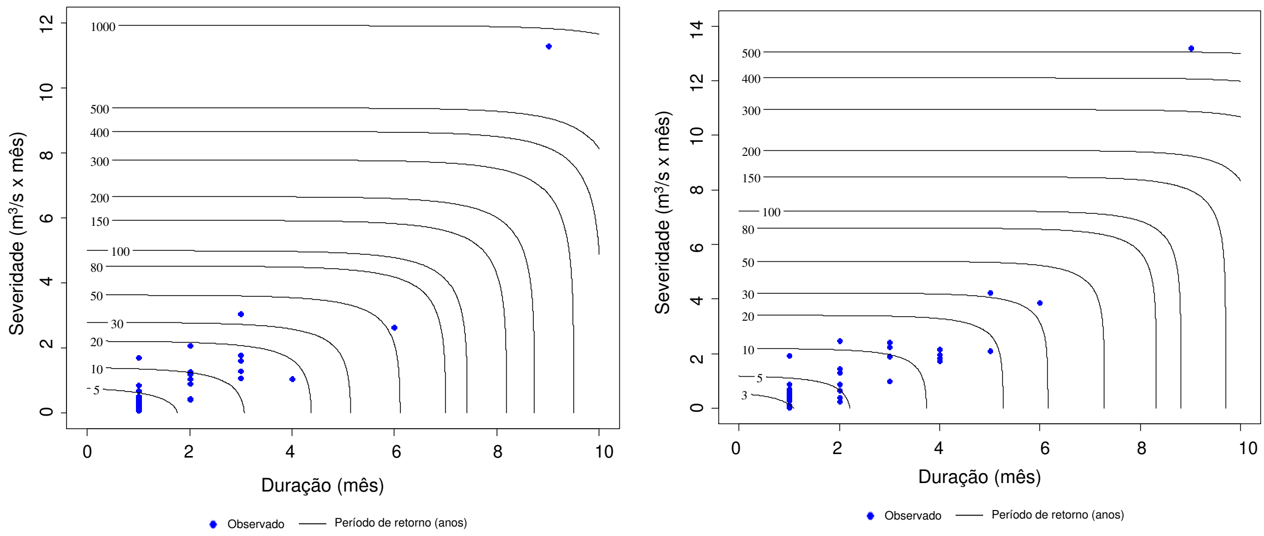
Como pode-se observar nas Figuras 4.4.1 e 4.4.2, para os níveis de truncamento analisados, a maioria dos eventos de seca observados do Rio Camanducaia possuem períodos de retorno inferiores a 50 anos, sendo 90% dos eventos para o nível de truncamento $P98$, 86% para o $P95$, 95% para o $P92$ e 98% para o $P90$. Os eventos observados, mais críticos em termos de duração e severidade, possuem períodos de retorno superiores a 300 anos, sendo estes, eventos com baixa probabilidade de ocorrência de acordo com as Figuras 4.3.1 e 4.3.2. Para os níveis de truncamento $P95$ e $P92$, o período de retorno dos eventos mais críticos são superior e próximo a 1.000 anos, respectivamente.

As Figuras 4.4.1 e 4.4.2 permitem efetuar um mapeamento dos comportamentos dos eventos de seca do Rio Camanducaia em termos de duração, severidade e período de retorno, podendo-se desta forma estabelecer intervalos para os pares duração-severidade em função do período de retorno desejado.



(a) (b)

Figura 4.4.1. Período de retorno em anos (isolinhas), dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P98$ (b) $P95$.



(a) **(b)**

Figura 4.2.2. Período de retorno em anos (isolinhas), dos eventos de seca do Rio Camanducaia para o nível de truncamento: (a) $P92$; (b) $P90$.

4.5. ANALISE SENSIBILIDADE DO PERÍODO DE RETORNO

Com base na Equação (3.5.13), pode-se observar que os períodos de retorno dos eventos de seca são influenciados pelas distribuições de probabilidade uni-variadas da durações e severidades, pela cópula que associa estas distribuições e pelo valor esperado do intervalo de tempo entre inícios de eventos de seca consecutivos.

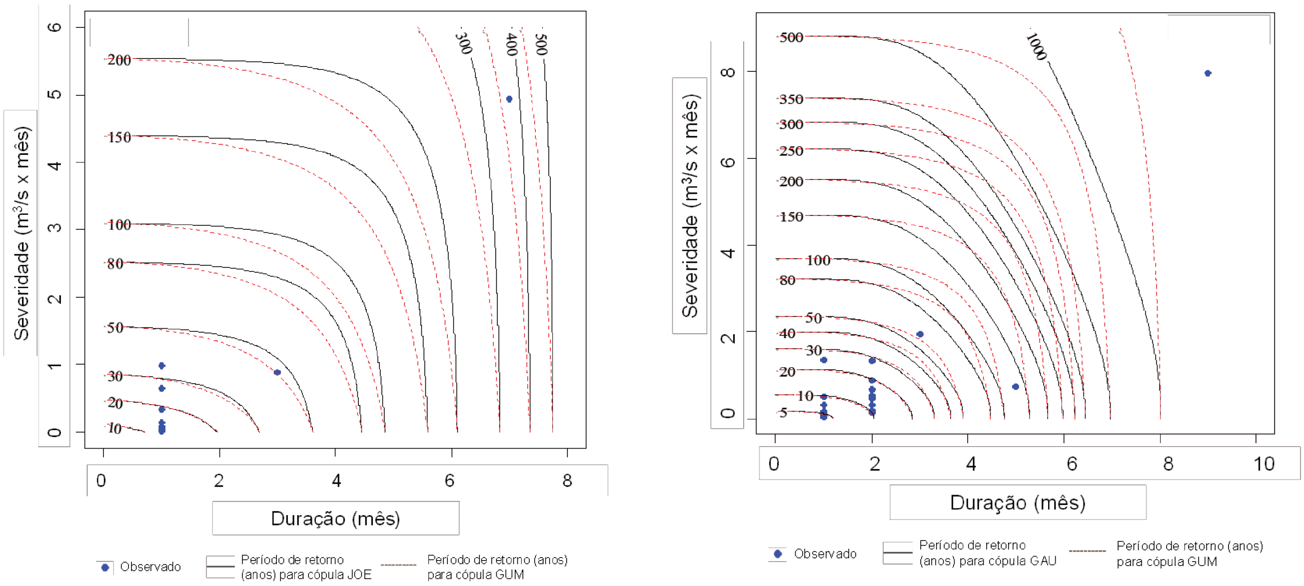
Considerando que as distribuições de probabilidade uni-variadas e o valor esperado do intervalo de tempo entre eventos não se altera, os períodos de retorno são influenciados somente pela cópula, isto é, pela forma como ela representa a estrutura de dependência das variáveis aleatórias (assimetria ou simetria e dependência ou não dependência nas caudas) e pela sua função matemática.

Para avaliar a sensibilidade dos períodos de retorno quanto às cópulas, levando em consideração a forma como elas representam a estrutura de dependência das durações e severidades e a sua função matemática, os resultados obtidos a partir das cópulas selecionadas da Tabela 4.3.2 foram comparados com os resultados calculados a partir de uma cópula GUM. A cópula GUM apresentou o segundo melhor ajuste para três níveis de truncamento como apresentado na Tabela 4.3.1.Na Tabela 4.5.1 são apresentadas as cópulas utilizadas na análise de sensibilidade dos períodos de retorno dos eventos de seca do Rio Camanducaia.

Tabela 4.5.1. Cópulas utilizadas para avaliar a sensibilidade dos períodos de retorno dos eventos de seca do Rio Camanducaia.

Nível de truncamento	Cópulas					
	Selecionadas			Analisadas		
	Tipo	**Simetria**	**Dependência nas caudas**	**Tipo**	**Simetria**	**Dependência nas caudas**
$P98$	JOE	Assimétrica	Superior	GUM	Assimétrica	Superior
$P95$	GAU	Simétrica	Nenhuma	GUM	Assimétrica	Superior
$P92$	JOE	Assimétrica	Superior	GUM	Assimétrica	Superior
$P90$	JOE	Assimétrica	Superior	GUM	Assimétrica	Superior

Nas Figuras 4.51 e 4.5.2, são apresentadas as comparações entre os períodos de retorno dos eventos de seca do Rio Camanducaia calculados para as cópulas da Tabela 4.5.1



(a) (b)

Figura 4.5.1. Sensibilidade do período de retorno em anos (isolinhas) dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P98$ (b) $P95$.

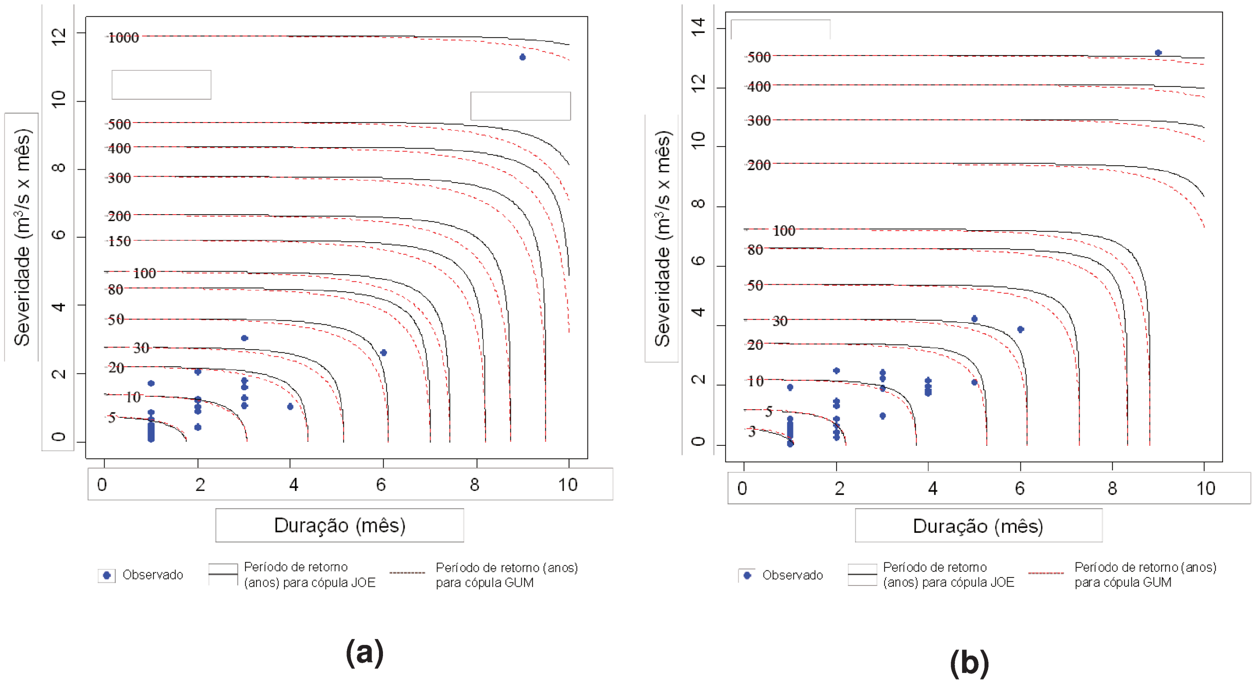


Figura 4.5.2. Sensibilidade do período de retorno em anos (isolinhas) dos eventos de seca do Rio Camanducaia para os níveis de truncamento: (a) $P92$ (b) $P90$.

Para os níveis de truncamento $P98$, $P92$ e $P90$, a comparação entre os períodos de retorno calculados a partir das cópulas JOE e GUM permite avaliar a influência da função matemática das cópulas, uma vez que as mesmas possuem igual estrutura de simetria e dependência nas caudas. Para o nível de truncamento $P95$, a comparação entre os períodos de retorno calculados a partir das cópulas GAU e GUM permite avaliar a influência da forma de representação da estrutura de dependência das durações e severidades.

As Figuras 4.5.1(a) e 4.5.2 apresentam as comparações entres os períodos de retorno calculadas pelas cópulas JOE e GUM. Como pode-se observar, a cópula GUM tende a superestimar os valores dos períodos de retorno, sendo este comportamento mais evidente para uma determinada região de pares duração-severidade. Pode-se observar também que, com o aumento do nível de truncamento, a região dos pares duração-severidade onde a cópula GUM superestima os períodos de retorno tende a ser mais restrita.

Na Figura 4.5.1(b), é apresentada a comparação entre os períodos de retorno calculados pelas cópulas GAU e GUM. Como pode-se observar, a cópula GAU tente a subestimar os períodos de retorno quando comparada à cópula GUM, assim como na análise anterior, existe uma região onde este comportamento é mais evidente.

Desta forma, considerando os resultados apresentados para o Rio Camanducaia para as cópulas GUM e JOE, ambas assimétricas com dependência na cauda superior, observa-se que a cópula GUM tendem a superestimar os períodos de retorno dos eventos de seca, sendo essa superestimava mais abrangente para pares duração-severidade considerando níveis de truncamento menores. Para o caso de cópulas que possuem formas distintas de representação da estrutura de dependência das variáveis aleatória, observa-se que, a cópula GAU (simétrica sem dependência nas caudas) tende a subestimar o período de retorno dos eventos de seca do Rio Camanducaia quando comparada à cópula GUM (assimétrica com dependência na cauda superior), sendo este comportamento mais evidente para uma determinada região de pares duração-severidade.

Os resultados apresentados evidenciam como a definição do tipo de cópula para representar a estrutura de dependência entre as durações e severidades dos eventos de seca pode afetar o cálculo dos períodos de retorno dos eventos.

5. CONCLUSÕES E RECOMENDAÇÕES

O presente trabalho investigou o comportamento dos eventos de seca do Rio Camanducaia, localizado na região sudeste do Brasil, tendo como base a série histórica de vazões médias mensais de 1944 a 2016 da estação fluviométrica 3D-002, localizada no munícipio de Monte Alegre do Sul, Estado de São Paulo, latitude 22° 41' 44" e longitude 46° 40' 25", com área de drenagem de 387 km^2.

A seguir são apresentadas as principais conclusões e resultados decorrentes da aplicação direta do método proposto considerando o estudo de caso:

1) Um reduzido número de eventos de seca é observado a partir dos registros históricos de vazões médias mensais. Estes poucos eventos podem prover uma limitada visão do risco de suprimento de água na bacia hidrográfica;
2) Nem todas as distribuições de probabilidade avaliadas (GAM2, EXP, WLB, LOGN, GP e GEV) performaram adequadamente na representação do comportamento probabilístico das durações dos eventos de seca. Especificamente, para a distribuição de probabilidade GEV não foi possível obter o estimador de máximo verossimilhança em nenhum dos quatro níveis de truncamento ($P98$, $P95$, $P92$ e $P90$);
3) De acordo com os p-valores do teste KS, somente as distribuições de probabilidade EXP e GP possuem hipótese de adequação não rejeitada aos dados de duração para os quatro níveis de truncamento. As distribuições de probabilidade com melhor ajuste às durações, de acordo com o AIC, foram a EXP, para os níveis de truncamento $P98$, $P92$ e $P90$, e a LOGN, para o nível de truncamento $P95$;
4) Para as severidades, considerando os quatro níveis de truncamento, a hipótese de adequação das seis distribuições de probabilidade não pode ser rejeitada de acordo com os p-valores do teste KS. As distribuições de probabilidade com melhor ajuste aos dados de severidade, de acordo com o AIC, foram a LOGN, para os níveis de truncamento $P98$, $P95$ e $P92$, e a GP, para o nível de truncamento $P90$;

5) De acordo com os p-valores do teste GOF, a hipótese das durações e severidades dos eventos de seca serem representadas conjuntamente pelas cópulas FRA, GUM, JOE e GAU não pode ser rejeitada para os quatro níveis de truncamento. A hipótese de adequação de uma cópula CLA não pode ser rejeitada somente para o nível de truncamento $P95$. Já, a hipótese de adequação de uma cópula AMH não pode ser rejeitada somente para o nível truncamento $P98$;
6) Observa-se que as cópulas assimétricas com dependência na cauda superior do tipo JOE e GUM, e simétricas sem dependência nas caudas do tipo GAU e FRA podem ser utilizadas para representar conjuntamente o comportamento probabilístico das durações e severidades dos eventos de seca para todos os quatro níveis de truncamento;
7) De acordo com o AIC, as cópulas com melhor ajuste na representação conjunta das durações e severidades foram a JOE, para os níveis de truncamento $P98$, $P92$ e $P90$, e GAU, para o nível de truncamento $P95$;
8) Não é possível estabelecer um único modelo probabilístico uni e bivariado para as durações e severidades dos eventos de seca. Condições como tamanho da amostra dos dados observados e nível de truncamento influenciam diretamente os resultados das análises;
9) As cópulas permitem gerar um grande número de eventos de seca, com diversas combinações de durações e severidades em comparação aos dados observados a partir da série histórica de vazões médias mensais;
10) A análise dos períodos de retorno com base nas cópulas JOE e GAU (as de melhor ajuste), mostram que grande parte dos eventos de seca observados a partir da série histórica de vazões, possuem recorrências inferiores a 50 anos, e eventos mais críticos, possuem recorrências superiores a 300 anos; e
11) A análise de sensibilidade mostrou que a cópula GUM tende a superestimar os valores dos períodos de retorno quando comparada à cópula JOE, sendo este comportamento mais evidente para uma determinada região de pares duração-severidade que tende a ser menor com o aumento do nível de truncamento. A análise de sensibilidade também mostrou que a cópula GAU tende a subestimar os períodos de retorno quando comparada à cópula GUM.

Desta forma, o método apresentado se mostra como uma ferramenta viável para ser aplicada nos estudos de planejamento e na gestão dos recursos hídricos de uma bacia hidrográfica, permitindo a geração de cenários de eventos de seca para a análise e mensuração dos riscos associados.

Como recomendações, o método pode ser estendido para a estação fluviométrica 3D-001, permitindo avaliar o comportamento espacial dos eventos de seca na bacia hidrográfica do Rio Camanducaia. As análises também podem ser realizadas em uma granularidade menor, como por exemplo, para as vazões médias diárias, obtendo-se resultados que podem ser utilizados na definição de um planejamento do uso da água no curto prazo, utilizando-se como níveis de truncamento as vazões de “Alerta” e “Restrição” de acordo com a REC 50/2015.

REFERÊNCIAS BIBLIOGRÁFICAS

1. AKAIKE, H. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, p. 716-723. 1974
2. AKYUZ, D. E.; BAYZIT, M.; ONOZ, B. Markov Chain Models for Hydrological Drought Characteristics. Journal of Hydrometeorology, 13, p. 298-309. 2012.
3. ALMEIDA, R.; BARBOSA, P. S. F. Caracterização do comportamento da estrutura de secas hidrológicas como subsídio a projetos e programas de mitigação de impactos. XV Simpósio Brasileiro de Recursos Hídricos. 2003.
4. ALMEIDA, R.; BARBOSA, P. S. F. Análise do comportamento probabilístico de secas hidrológicas. XXI Congresso Latino-americano de Hidráulica. 2004.
5. ARAÚJO J. C.; BRONSTERT, A. A method to assess hydrological drought in semiarid environments and its application to the Jaguaribe River basin, Brazil. Water International. 41, 2, p. 213–230. 2016.
6. AZAM, M.; MAENG, S. J.; KIM, H. S.; MURTAZAEV, A. Copula-Based Stochastic Simulation for Regional Drought Risk Assessment in South Korea. Water, 10, p. 2-29. 2018.
7. BERAN, M. A.; RODIER, J. A. Hydrological Aspects of Droughts: A Contribution to the International Hydrological Programme.Studies and Reports in Hydrology, UNESCO-WMO, Paris, 1985.
8. BONACCORSO, B.; CANCELLIERE, A.; ROSSI, G. An Analytical Formulation of Return Period of Drought Severity. Stochastic Environmental Researsh and Risk Assessment, 17, p. 157-174. 2003
9. BYZEDI, M.; SAGHAFIAN, B. Regional Analysis of Streamflow Drought: A Case Study for Southwestern Iran. World Academy Science, Engineering and Technology. International Journal of Energy and Power Engineering, 3, 9, p. 1700-1704. 2009.
10. BOX, G. E. P.; PIERCE, D. A. Distribution of residual correlations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65, p. 1509–1526. 1970.
11. CANCELLIERE, A.; SALAS, J. D. Drought Length Propeties for Periodic-Stochastic Hydrologic Data. Water Resouces Reserarch, 40, p. 1-13. 2004.

12. CHANG, T. J. Effects of Drought on Streamflow Characteristic. Journal of Irrigation and Drainage Engineering, 116, 3, p. 332-341. 1990.

13. CHANG, T. J.; STENSON, J. R. Is It Realistic to Define a 100-Year Drought for Water Management?, 26, 5, p. 823-829.1990.

14. COMITÊ DAS BACIAS HIDROGRÁFICAS DOS RIOS PIRACICABA, CAPIVARI E JUNDIAI. Plano das Bacias Hidrográficas dos Rios Piracicaba, Capivari e Jundiaí 2010 a 2020 – Relatório Síntese. São Paulo: Cobrape: Neoband Soluções Gráficas, 2011, 128p.

15. COMITÊ DAS BACIAS HIDROGRÁFICAS DOS RIOS PIRACICABA, CAPIVARI E JUNDIAI. Relatório da Situação dos Recursos Hídricos 2018 – UGRHI 05 – Bacias Hidrográficas dos Rios Piracicaba, Capivarí e Jundiaí. Versão Simplificada, 2018, 101p.

16. CUNHA, L. V.; VLACHOS, E.; YEVJEVICH, V. Drought, Enviroment and Society. In: YEVJEVICH, V.; CUNHA, L.; VLACHOS, E. (Ed.). Coping with Droughts. Litteton, Colorado: Water Resources Publications, 1983. p. 3-11.

17. DE MICHELE, C.; SALVADORI, G. A Generalized Pareto Intensity-Duration Model of Storm Rainfall Exploiting 2-Copulas. Journal of Geophysical Research, 108, p. 15-1-15-11. 2003.

18. DE MICHELE, C., SALVADORI, G., VEZZOLI, R., PECORA, S. Multivariate Assessment for Droughts: Frequency Analysis and Dynamic Return Period. Water Resources Research, 40, p. 6985-6994. 2013.

19. DODANGEH, E.; SHAHEDI, K.; SHIAU, J. T.; MIRAKBARI, M. Spatial Hydrological Drought Characteristics in Karkheh River Basin, Southwest Iran Using Copulas. Journal of Earth System Science, p. 1-20. 2017.

20. DRACUP, L.; LEE, K. S.; PAULSON Jr, E. G. On The Definition Drought. Water Resources Research, 16, 2, p. 297-302.1980a.

21. DRACUP, L.; LEE, K. S.; PAULSON Jr, E. G. On The Statistical Characteristics of Drought Events. Water Resources Research, 16, 2, p. 289-296.1980b.

22. EKANAYAKE, E. M. R. S. B., PERERA, K. Analysis of Drought Severity and Duration Using Copulas in Anaradhapura, Siri Lanka. British Journal of Environment & Climate Change, 4, p. 312-327. 2014.

23. FAN, L.; WANG, H.; WANG, C.; LAI, W.; ZHAO, Y. Exploration of Use of Copulas in Analyzing the Relationship between Precipitation and Meteorological Drought in Beijing, China. Advances in Meteorology, p. 1-11. 2017.
24. FERMANIAN, J. D. Goodness-of-fit tests for copulas. Journal of Multivariate Analysis, 95, p. 119–152, 2005.
25. FRÉCHET, M. R. Sur les tableaux de corrélation dont les marges sont données. Ann. Univ. Lyon Sci. 4, 53–84, 1951.
26. GANGULI, P.; REDDY, M. J. Risk Management of Drought in Gujarat Using Bivariate Copulas. Water Resources Management, 26, p. 3301-3327. 2012.
27. GENEST, C.; FAVRE, A. C. Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 12, p. 347–368. 2007.
28. GENEST, C.; GHOUDI, K.; RIVEST, L-P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika, 82, 3, p. 543-552, 1995.
29. GENEST, C.; QUESSY, J-F.; RÉMILLARD, B. Goodness-of-fit procedure for copula models based on the probability integral transformation. Scandinavian Journal of Statistic, 33, p. 337-366. 2006.
30. GENEST, C.; RÉMILLARD, B. Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. Annales de l'Institut Henri Poincare: Probabilites et Statistiques, 44, p. 1069-1127. 2008.
31. GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44, p. 199-214. 2009.
32. GONZÁLEZ, J.; VALDÉS, J. B. Bivariate Drought Recurrence Analysis Using Tree Ring Reconstructions. Journal Hydrology Engineering, 8, 5, p. 247-258. 2003
33. HAMDI, Y.; CHEBANA, F.; OUARDA, T. B. M. J. Bivariate Drought Frequency Analysis in the Medjerda River Basin Tunisia. Journal of Civil & Environmental Engineering, 6, 3, p. 1-11. 2016.
34. HORN, D. R. Characteristics and Spatial Variability of Droughts in Idaho. Journal of Irrigation and Drainage Division, ASCE 115, 1, p. 111-124. 1989.

35. HUDSON Jr., H. E.; HAZEN, R. Droughts and Low Streamflow. In: CHOW, V. T. (Ed.). Handbook of Applied Hydrology. New York: McGraw-Hill Book Company, 1964, p. 18-1-18-26.

36. JOE, H. Multivariate Models and Dependence Concepts; Chapman & Hall, London, 1997.

37. KWAK, J.; KIM, Y. S.; LEE, J. S.; KIM, H. S. Drought Severity-Duration-Frequency Analysis of Hydrological Drought Based on Copula Theory. Hydrology Days, p. 82-89. 2012.

38. KWAK, J.; KIM, S.; KIM, G.; SINGH, V. P.; PARK, J.; KIM, H. S. Bivariate Drought Analysis Using Streamflow Reconstruction with Tree Ring Indices in the Sacramento Basin, California, USA. Water, 8, 122, p. 1-16. 2016a.

39. KWAK, J.; KIM, S.; KIM, D.; KIM, H. Hydrological Drought Analysis Based on Copula Theory. In River Basin Management; Intech, p. 83-95. 2016b.

40. KOJADINOVIC, I.; YAN, J. Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. Journal of Statistical Software, 34, 1-20. 2010.

41. KOJADINOVIC, I.; YAN, J. A goodness-of-fit test for multivariate multiparameter copulas based on multiplier central limit theorems. Statistics and Computing, 21, p. 17-30. 2011.

42. KOJADINOVIC, I.; YAN, J.; HOLMES, M. Fast large-sample goodness-of-fit tests for copulas. Statistica Sinica, 21, p. 841-871. 2011.

43. LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. Biometrika, 65, p. 297–303. 1978.

44. MATHIER, L.; PERREAULT, L.; BOBÉE, B. The Use of Geometric and Gamma-Related Distributions for Frequency Analysis of Water Deficit. Stochastic Hydrology and Hydraulics, 6, p. 239-254. 1992.

45. MEDEIROS, G. C. S. Metodologia de Avaliação da Seca Hidrológica sob a Perspectiva da Demanda Hídrica, 2016, 79p. Dissertação de Mestrado, Universidade Federal do Rio Grande do Norte, 2016.

46. MILLAN, J.; YEVJAVICH, V. Probabilities of Observed Droughts. Hydrology Paper 50. Colorado State University. Fort Collins, Colorado, 1971.

47. MIRABBASI, R., FAKHERI-FARD, A., DINPASHOH, Y. Bivariate Drought Frequency Analysis Using the Copula Method. Theoretical and Applied Climatology, 108, p. 191–206. 2011.

48. MIRSHA, A. K.; SINGH, V. P. A. Review of Drought Concepts. Journal of Hydrology, 391, p. 203-216. 2010.

49. NALBANTIS, I. Evaluation of a Hydrological Drought Index. European Water, 23/24, p. 67-77. 2008.

50. NELSEN, R. An Introduction to Copulas; 2nd ed.; Springer, New York, USA, 2006.

51. PANU, U. S.; SHARMA, T. C. Analysis of Annual Hydrological Droughts: The Case of Norwest Ontario, Canada. Hydrological Sciences Journal, 54, 1, p. 29-42. 2009.

52. QUESSY, J.-F. Méthodologie et application des copules: tests d'adéquation, tests d'indépendance, et bornes sur la valeur-à-risque. Tese de Doutorado. Université Laval, Québec, Canada. 2005.

53. ROYSTON, P. An extension of Shapiro and Wilk's W Test for Normality to Large Samples. Applied Statistics,31, p. 115–124,.1982a.

54. ROYSTON, P. Algorithm AS 181: The W test for Normality. Applied Statistics, 31, p. 176–180. 1982b.

55. ROYSTON, P. Remark AS R94: A Remark on Algorithm AS 181: The W test for Normality. Applied Statistics, 44, p. 547–551. 1995

56. SALAS, J. D.; FU, C.; CANCELLIERE, A.; DUSTIN, D.; BODE, D.; PINEDA, A.; VICENT, E. Characterizing the Severity and Risk of Drought in the Poudre River, Colorado. Journal of Water Resources Planning and Management, 131, 5, p. 383-393. 2005.

57. SALVADORI, G.; DE MICHELE, C. Multivariate Real-Time Assessment of Drought via Copula-based Multi-site Hazard Trajectories and Fans. Journal of Hydrology, 526, p. 101-115. 2015.

58. SALVADORI, G.; DE MICHELE, C.; KOTTEGODA, N. T.; ROSSO, R. Extremes in Nature. An approach using Copulas. Water Science and Technology Library Series (Book 56), Springer, Dordrecht, 2007.

59. SANTOS, M. J. J. Caracterização e Monitoramento de Secas. Instituto da Água, Direção de Serviços de Recursos Hídricos, 1998. 26p.

60. SEN, Z. Wet and Dry Periods of Annual Flow Series. Journal of the Hydraulics Division, p. 1503-1514. 1976.

61. SEN, Z. Run-Sums of Annual Flow Series. Journal of Hydrology, 35, p. 311-324. 1977.

62. SEN, Z. Critical Drought Analysis of Periodc-Sthochastic Process. Journal of Hydrology, 46, p. 251-263. 1980a.

63. SEN, Z. Statistical Analysis of Hydrologic Critical Droughts. Journal of the Hydraulics Division, p. 99-115. 1980b.

64. SEN, Z. Critical Drought Analysis by Second-Order Markov Chain. Journal of Hydrology, 120, p. 183-202. 1990.

65. SHARNA, T. C. Estimation of Drought Severity on Independent and Dependent Hydrologic Series. Water Resources Management, 11, p.35-49. 1997.

66. SHARMA, T. C.; PANU, U. S. Analytical Procedures for Week Hydrological Droughts: A Case of Canadian Rivers. Hydrological Sciences Journal, 55, 1, p. 79-92. 2010.

67. SHARMA, T. C.; PANU, U. S. Modeling of Hydrological Drought Durations and Magnitudes: Experiences on Canadian Streamflows. Journal of Hydrology: Regional Studies, 1, p. 92-106. 2014.

68. SHE, D.; XIA, J. Copulas-Based Drought Characteristics Analysis and Risk Assessment across the Loess Plateau of China. Water Resources Management, 32, p. 547–564. 2018.

69. SHIAU, J. T. Return Period of Bivariate Distributed Extreme Hydrological Events. Stochastic Environmental Researsh and Risk Assessment, 17, p. 42-57. 2003

70. SHIAU, J. T. Fitting Drought Duration and Severity with Two-Dimensional Copulas. Water Resources Management 20, p. 795-815. 2006.

71. SHIAU, J. T.; FENG, S., NADARAJAH, S. Assessment of Hydrological Droughts for the Yellow River, China, Using Copulas. Hydrological Processes, 21, p. 2157-2163. 2007.

72. SHIAU, J. T.; MODARRES, R. Copula-based Drought Severity-Duration-Frequency Analysis in Iran. Meteorological Applications, 16, p. 481-489. 2009.

73. SHIAU, J. T.; SHEN, H. W. Recurrence Analysis of Hydrologic Droughts of Differing Severity. Journal of Water Resources Planning and Management, 27, 1, p. 30-40. 2001.

74. SILVA, R. F.; LOPES, H. F. Copula, marginal distributions and model selection: a Bayesian note. Statistics and Computing, 18, p. 313–320. 2008.

75. SISTEMA INTEGRADO DE GERENCIAMENTO DOS RECURSOS HÍDRICOS – SIGRH. (6 de fevereiro de 2017). São Paulo: SIGRH. Acesso: http://www.sigrh.sp.gov.br.

76. SKLAR, M. Fonctions de Répartition à n Dimensions et Leurs Marges; Université Paris 8: Saint-Denis, France, 1959.

77. SOLON, A. O.; CAMPOS, J. N. B. ; STUDART, T. M. A. C. Estimativa dos Valores Esperados para Durações Máximas de Secas Hidrológicas no Açude Castanhão - CE. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 14. 2001, Aracaju. SE, Anais... Aracaju: Associação Brasileira de Recursos Hídircos - ABRH, 2001.

78. STUTE, W.; Manteiga, W. G.; Quindimil, M. P. Boostrap based gonddness-of-fit tests. Metrika, 40, p. 243-256. 1993.

79. TESFAYE, Y. G. Seasonal Time Series Models and Their Application to the Modeling of River Flows, 2005, 173p. Dissertação de Mestrado, Universidade de Nevada, 2005.

80. TOSUNOGLUA, F.; ONOF, C. Joint modelling of drought characteristics derived from historical and synthetic rainfalls: Application of Generalized Linear Models and Copulas. Journal of Hydrology: Regional Studies, 14, p. 167-181. 2017.

81. ZHANG, S.; OKHRINB, O.; ZHOU, Q. M.; SONG, P. X-K. Goodness-of-fit test for specification of semiparametric copula dependence models. Journal of Econometrics, 193, p. 215-233. 2016.

82. ZHAO, P.; LÜ, H.; FU, G.; ZHU, Y.; SU, J.; WANG, J. Uncertainty of Hydrological Drought Characteristics with Copulas Functions and Probability Distributions: A Case Study of Weihe River, China. Water, 9, 334, p. 1-15. 2017.

83. ZELENHASIC, E.; SALVAI, A. A Method od Streamflow Drought Analysis. Water Resources Research, 23, 1, p. 156-168. 1987.

84. YAHIAOUI, A.; TOUAÏBIA, B.; BOUVIER, C. Frequency analysis of the Hydrological Drought Regime: Case of Oued Mina Catchment in western of Algeria. Revue Nature et Technologie, 1, p. 3-15. 2009.

85. YEVJEVICH, V. An Objective Approach to Definitions and Investigations of Continental Hydrologic Droughts. Hydrologic Paper 23. Colorado State University. Fort Collins, Colorado, 1967, 18p.

86. WANG, Y.; LI, C.; LIU, J.; FULIANG, Y.; QIU, Q.; TIAN, J.; ZANG, M. Multivariate Analysis of Joint Probability of Different Rainfall Frequencies Based on Copulas. Water, 9, p. 1-16. 2017.

87. WANG, W.; WELLS, M. T. Model selection and semiparametric inference for bivariate failure-time data (with discussion). Journal of the American Statistical Association, 95, p. 62-76. 2000.

88. WIJAYARATNE, L. H.; GOLUB, E. Multiyear Drought Simulation. Water Resources Bulletin, 27, 3, p. 387-395. 1991.

89. WILHITE, D. A.; GLANTZ, M. H. Understanding the Drought Phenomenon: The Role of Definitions. Water Internacional, 10, p. 111-120. 1985.

APÊNDICE A

Tabela A.1 – Vazões médias mensais da estação fluviométrica 3D-002 do rio Camanducaia.

Ano	Vazão média mensal (m^3/s)											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1944	7,15	7,05	9,24	5,44	4,24	2,99	2,66	2,13	1,96	2,31	5,49	3,89
1945	3,94	10,39	4,95	3,62	3,13	7,45	4,52	3,01	2,71	2,4	6,48	8,58
1946	18,8	14,52	13,16	10,02	6,24	5,73	5,88	3,79	3,06	3,68	3,87	6,3
1947	18,78	19,78	22,04	11,65	8,75	7,11	6,27	5,79	6,65	6,65	5,2	10,53
1948	11,09	11,28	14,23	8,46	7,09	5,79	5,19	4,68	4,08	4,09	4,29	5,07
1949	9,95	14,56	8,7	7,46	5,81	5,44	4,46	3,65	3,3	3,45	4,43	12,31
1950	15,3	23,25	18,53	13,47	9,18	7,69	6,3	5,4	4,92	5,51	6,64	8,12
1951	10,54	11,74	13,36	9,14	5,67	4,66	4,07	3,43	2,6	3	7,52	5,02
1952	7,9	16,68	11,75	6,82	4,69	6,56	3,85	3,06	3,05	2,96	5,32	3,53
1953	3,83	5,21	4,83	5,19	3,42	3,1	2,49	2,12	2,13	2,34	2,33	4,38
1954	8,02	13,96	7,39	4,75	6,85	4,86	3,28	2,48	2,06	2,66	1,63	3,59
1955	8,12	4,48	7,14	4,2	3,51	2,96	2,33	2,59	1,95	1,76	3,54	7,01
1956	7,14	7,53	5,36	3,9	5,63	5,43	3,73	6,01	3,4	3,08	2,85	3,34
1957	8,34	13,22	11,27	7,64	5,36	4,47	4,41	4,13	6,42	4,56	4,17	7,49
1958	12,42	13,31	11,58	10,41	11,52	9,22	6,62	4,52	5,17	4,43	4,96	4,51
1959	10,27	5,77	11,06	8,59	4,49	3,2	2,5	2,67	1,69	2,09	4,24	6,95
1960	11,8	11,09	9,13	4,83	4,33	3,64	2,86	2,29	1,63	1,94	4,5	13,09
1961	14,95	11,21	13,03	9,07	8,47	5,16	4,33	3,53	3,07	2,66	3,36	5,14
1962	8,64	14,26	14,94	6,79	5,29	4,11	3,29	3,66	3,09	9,13	6,6	9,45
1963	18,82	16,26	11,26	6,77	4,44	3,71	3,24	2,82	2,16	4,77	5,6	3,55
1964	4,84	13,85	5,9	3,75	3,62	2,8	2,86	2,08	2,01	4,7	4,47	13,45
1965	15,52	17,09	15,02	7,43	7,81	4,92	4,95	3,65	3,33	7,43	7,24	14,25
1966	14,64	12,5	16,11	7,86	6,74	4,12	3,84	3,07	2,26	3,93	5,72	10,52
1967	15,73	17,54	11,31	8,13	5,74	6,19	4,14	3,05	3,2	3,3	4,95	6,14
1968	8,98	5,25	6,86	4,06	3	2,46	2,08	2,04	1,83	2,05	1,86	2,9
1969	3,39	2,62	3,28	2,92	2,35	2,5	1,65	1,61	1,16	4,21	7,33	7,58
1970	10,44	22,33	12,96	7,2	5,48	4,56	3,76	3,39	4,03	4,79	5,98	4,81
1971	4,83	3,84	7,67	4,94	4,36	6,54	3,89	3,65	3,28	6,85	4,72	8,41
1972	10,68	14,36	9,25	7,25	4,8	3,7	4,72	4,99	3,47	7,19	6,69	5,92
1973	9,8	9,99	7,78	9,33	6,08	4,91	4,8	3,54	3,34	4,63	6,17	13,11
1974	18,12	10,8	10,31	7,28	5,15	5,99	4,07	3,33	2,83	3,72	4,31	8,53
1975	12,03	14,69	9,82	6,65	4,89	3,74	3,47	2,54	1,98	3,24	5,37	10
1976	9,4	14,83	12,58	10,33	8,95	10,36	9,76	8,26	11,69	9	11,43	12,11
1977	12,56	9,94	8,6	12,88	7,06	6,63	4,46	3,32	3,94	3,43	5,13	11,3

... continuação

Ano	Vazão média mensal (m^3/s)											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1978	9,12	6,22	5,99	3,82	3,66	4,41	4,17	2,76	2,6	2,14	6,71	6,75
1979	6,39	8,28	5,64	5,44	7,6	4,59	4,44	4,2	4,54	5,35	6,15	8
1980	11,51	8,02	6,96	12,23	5,99	5,8	4,45	3,73	3,31	3,05	4,41	10,23
1981	19,15	8,82	6,9	5,22	3,88	4,65	3,36	2,79	2	5,21	7,56	14,5
1982	11,96	14,96	16,13	10,31	7,39	9,11	7,09	5,33	3,8	9,94	5,87	14,08
1983	27,9	35,53	24,17	16,87	16,05	24,52	14,07	9,86	15,91	12,17	10,6	18,98
1984	16,87	11,3	8,09	8,31	6,37	4,56	3,74	4,8	4,2	3,15	3,64	8,91
1985	10,69	8,62	13,29	8,68	6,18	4,6	3,79	3,13	4,1	2,68	3,95	3,83
1986	3,83	6,1	8,01	4,8	4,19	3,27	2,85	3,43	2,19	2,75	3,09	19,44
1987	14,53	13,28	13,34	8,93	13,84	9,44	6,61	5,26	5,45	5,67	4,28	6,19
1988	9,37	10,01	14,86	11,24	9,27	8,43	5,74	4,6	3,77	5,93	5,49	5,53
1989	12,95	13,79	9,6	6,21	4,57	4,07	4,04	4,19	3,48	2,09	3,89	5,14
1990	15,81	6,21	7,91	4,82	4,81	3,27	4,49	3,52	3,52	3,8	2,75	3,79
1991	7,81	11,31	13,9	13,27	8,32	5,88	4,67	3,23	2,83	7,36	3,19	7,79
1992	4,97	4,75	5,75	5,93	6,39	3,13	3,31	2,22	3,66	10,52	10,88	10,8
1993	9,7	16,11	17,16	11,74	9,74	8,72	5,61	4,68	7,43	7,03	4,29	5,63
1994	6,86	9,7	11	8,53	6,6	5,8	5,13	3,17	1,82	2,85	3,86	9,42
1995	10,02	25,51	16,5	15,65	10,56	8,01	7,94	5,23	4,98	7,38	5,36	5,42
1996	15,41	12,39	15,07	9,45	6,33	4,6	3,88	3,91	8,2	6,15	11,74	13,48
1997	14,64	13,72	9,74	6,7	5,49	8,26	4,67	3,34	3,25	3,58	6,57	8,48
1998	8,03	9,76	8,62	5,58	5,9	4,35	3,02	2,1	1,84	6,15	3,52	10,33
1999	0,78	16,37	12,99	8,67	6,55	6,72	4,21	3,1	3,14	2,12	2,7	6,28
2000	11,93	11,73	8,24	5,58	3,93	3,25	3,57	3,2	4,2	2,15	6,2	9,92
2001	7,56	12,37	7,82	7,9	4,54	3,42	2,71	2,34	2,86	7,3	7,44	8,62
2002	13,8	15,61	10,75	6,98	5,48	3,64	2,92	3,33	3,54	2,06	3	4,43
2003	14,72	8,37	7,18	4,51	3,81	2,92	2,55	2,35	2,21	2,76	3,85	11,2
2004	7,21	11,6	7	7,62	7,41	8,66	6,41	3,88	3,09	5,27	6,8	10,25
2004	17,83	11,85	12,59	7,49	7,53	5,35	4,71	3,52	4,25	4,44	4,25	7,54
2006	9,56	10,11	12,25	6,56	4,22	3,66	3,16	2,84	2,8	3,02	5,16	4,62
2007	13,98	6	6,73	5,13	3,81	3,79	6,07	3,36	2,54	2,61	7,53	6,65
2008	9,85	12,1	13,95	12,68	8,41	6,76	4,47	4,18	3,78	3,37	6,15	8,69
2009	12,39	17,1	11,12	7,74	5,52	5,02	5,17	4,52	5,15	7,39	7,84	23,18
2010	21,09	12,25	11,96	8,07	5,1	4,06	3,44	2,14	1,97	2,08	3,49	5
2011	22,82	8,15	11,99	8,69	5,2	4,95	3,57	3,19	2,64	3,83	6,85	6,19
2012	12,27	7,15	4,65	5,23	5,37	8,83	6,34	3,62	2,88	3,56	3	4,6
2013	7,58	7,32	8,72	7,71	4,35	4,42	4,69	2,69	2,31	3,32	2,41	2,73
2014	1,93	1,43	2,32	2,93	1,33	1,3	1,31	0,88	0,92	0,69	1,02	1,92

... continuação

Ano	Vazão média mensal (m^3/s)											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2015	1,74	5,76	5,58	2,85	2,32	2,79	1,79	1,13	3,06	2,07	4,52	10,3
2016	12,27	7,88	14,04	4,95	4,7	12,55	3,97	3,57	2,77	4,3	3,89	3,96

APÊNDICE B

B.1. TEORIA DOS RUNS

Função: runs(x, x0)

- x: vetor numérico contendo a série temporal de dados; e
- x0: valor numérico correspondente ao nível de truncamento.

Valores: Matriz de dados $n \times 2$, onde n é o número de eventos. A primeira coluna da matriz contém os valores das severidades, e a segunda, as durações.

Na sequência são apresentados os detalhes de implementação da função.

```
runs <- function(x, x0, ...){
 y<-x-x0
 i<-1
 n<-0
 while (i<=length(x)){
  if (y[i]<0){
   n<-n+1
   while (y[i]<0&(i<=length(x))){
    i<-i+1}}
  if (i<=length(x)){
   if (y[i]>=0){
    i<-i+1}}}
 z<-matrix(nrow = n, ncol = 2)
 i<-1
 n<-0
 while (i<=length(x)){
  if (y[i]<0){
   n<-n+1
   s<-0
   d<-0
```

```
  while (y[i]<0&(i<=length(x))){
   z[n,1]<-y[i]+s
   s<-z[n,1]
   z[n,2]<-1+d
   d<-z[n,2]
   i<-i+1}}
 if (i<=length(x)){
  if (y[i]>=0){
   i<-i+1}}}
z<-abs(z)
z}
```

B.2. TESTE DE BOX-PIERCE E LJUNG

Função: Box.test(x, lag = 1, type = c("Box-Pierce", "Ljung-Box"))

- x: vetor numérico contendo a série temporal de dados;
- lag: defasagem testada; e
- type: tipo de teste.

Valor: p-Valor do teste.

B.3. COEFICIENTES DE CORRELAÇÃO

Função: cor(x, y, method=c("pearson","kendal","spearman"))

- x e y: vetores numéricos contendo as séries temporais de dados;
- method: tipo de correção.

Valor: Valor do coeficiente de correlação.

B.4. AJUSTES DAS DISTRIBUIÇÕES DE PROBABILIDADE

Biblioteca: library("fitdistrplus")

Função: summary(fitdist(x, distr, method = c("mle")))

- x: vetor numérico contendo a série temporal de dados;

- distr: distribuição de probabilidade: “gamma”, “exp”; “weibull” ou “lnorm”; e
- method: método de ajuste dos parâmetros (máximo verossimilhança).

Valores: Parâmetro(s) estimado(s) e AIC.

B.5. AJUSTE DAS DISTRIBUIÇÕES DE PROBABILIDADE DE VALORES EXTREMOS

Biblioteca: library(extRemes)

Função: fevd(x, threshold=0, type=c("GP", ”GEV”))

- x: vetor numérico contendo a série temporal de dados; e
- type: distribuição de probabilidade.

Valores: Parâmetro(s) estimado(s) e AIC.

B.6. KS TESTE USANDO BOOTSTRAP PARAMÉTRICO

Função: teste.ks(x, distr=c("gamma", "weibull", "lnorm", "exp", "gdp"), n)

- x: vetor numérico contendo a série temporal de dados;
- distr: distribuição de probabilidade; e
- n: número de pseudo observações.

Valores: p-Valor do teste.

Na sequência são apresentados os detalhes de implementação da função.

```
teste.ks<-function(x, distr=c("gamma", "weibull", "lnorm", "exp", "gdp"), n=2000, ...){
 random.ksstat<-matrix(nrow=n, ncol=1)
 if (distr=="gamma"||distr=="weibull"||distr=="lnorm"){
  library("fitdistrplus")
  p1<-fitdist(x, distr)$estimate[1]
  p2<-fitdist(x, distr)$estimate[2]
  if (distr=="gamma"){
   obs.ksstat<-ks.test(x, pgamma, shape=p1, rate=p2, alternative="gr")$stat
   for (i in 1:n){
```

```
   xx<-rgamma(length(x), shape=p1, rate=p2)
   random.ksstat[i,1]<-ks.test(xx, pgamma, shape=p1, rate=p2)$stat}}
 if (distr=="weibull"){
  obs.ksstat<-ks.test(x, pweibull, shape=p1, scale=p2, alternative="gr")$stat
  for (i in 1:n){
   xx<-rweibull(length(x), shape=p1, scale=p2)
   random.ksstat[i,1]<-ks.test(xx, pweibull, shape=p1, scale=p2)$stat}}
 if (distr=="lnorm"){
  obs.ksstat<-ks.test(x, plnorm, meanlog=p1, sdlog=p2, alternative="gr")$stat
  for (i in 1:n){
   xx<-rlnorm(length(x), meanlog=p1, sdlog=p2)
   random.ksstat[i,1]<-ks.test(xx, plnorm, meanlog=p1, sdlog=p2)$stat}}}
else{
 if (distr=="exp"){
  library("fitdistrplus")
  p1<-fitdist(x, distr)$estimate[1]
  obs.ksstat<-ks.test(x, pexp, rate=p1, alternative="gr")$stat
  for (i in 1:n){
   xx<-rexp(length(x), rate=p1)
   random.ksstat[i,1]<-ks.test(xx, pexp, rate=p1)$stat}}}
if (distr=="gdp"||distr=="gev"){
 library(fExtremes)
 library(extRemes)
 Fn<-ecdf(x)
 F_y<-sort(Fn(x))
 if (distr=="gdp"){
  p<-fevd(x, threshold=0, type="GP")$results$par
  p1<-p[1]
  p2<-p[2]
  F_x<-pgpd(sort(x), xi=p2, mu=0, beta=p1)
  obs.ksstat<-max(abs(F_x-F_y))
  for (i in 1:n){
```

```
   xx<-rgpd(length(x), xi=p2, mu=0, beta=p1)
   F_x<-pgpd(sort(xx), xi=p2, mu=0, beta=p1)
   Fn<-ecdf(xx)
   F_y<-sort(Fn(xx))
   random.ksstat[i,1]<-max(abs(F_x-F_y))}}
 if (distr=="gev"){
  p<-fevd(x, type="GEV")$results$par
  p1<-p[1]
  p2<-p[2]
  p3<-p[3]
  F_x<-pgev(sort(x), xi=p3, mu=p1, beta=p2)
  obs.ksstat<-max(abs(F_x-F_y))
  for (i in 1:n){
   xx<-rgev(length(x), xi=p3, mu=p1, beta=p2)
   F_x<-pgev(sort(xx), xi=p3, mu=p1, beta=p2)
   Fn<-ecdf(xx)
   F_y<-sort(Fn(xx))
   random.ksstat[i,1]<-max(abs(F_x-F_y))}}
 else{
  pvalue=NULL}}
pvalue<-length(random.ksstat[random.ksstat>obs.ksstat])/n
pvalue}
```

B.7. PREPARAÇÃO DOS DADOS PARA AJUSTE DAS CÓPULAS

Função: cop.data<-function(x, distr=c("gamma","weibull","lnorm","exp","gdp","gev"))

- x: matriz numérica contendo as variáveis. Cada coluna representa uma variável e as linhas suas observações; e
- distr: distribuição de probabilidade.

Valores: Matriz numérica contendo as probabilidades acumuladas das variáveis.

Na sequência são apresentados os detalhes de implementação da função.

```
cop.data<-function(x, distr=c("gamma","weibull","lnorm","exp","gdp","gev"), ...){
 n<-length(distr)
 prob<-matrix(nrow=length(x[,1]), ncol=n)
 for (i in 1:n){
  if (distr[i]=="gamma"){
   library("fitdistrplus")
   p1<-fitdist(x[,i], "gamma")$estimate[1]
   p2<-fitdist(x[,i], "gamma")$estimate[2]
   prob[,i]<-pgamma(x[,i], shape=p1, rate=p2)}
  if (distr[i]=="weibull"){
   library("fitdistrplus")
   p1<-fitdist(x[,i], "weibull")$estimate[1]
   p2<-fitdist(x[,i], "weibull")$estimate[2]
   prob[,i]<-pweibull(x[,i], shape=p1, scale=p2)}
  if (distr[i]=="lnorm"){
   library("fitdistrplus")
   p1<-fitdist(x[,i], "lnorm")$estimate[1]
   p2<-fitdist(x[,i], "lnorm")$estimate[2]
   prob[,i]<-plnorm(x[,i], meanlog=p1, sdlog=p2)}
  if (distr[i]=="exp"){
   library("fitdistrplus")
   p1<-fitdist(x[,i], "exp")$estimate[1]
   prob[,i]<-pexp(x[,i], rate=p1)}
  if (distr[i]=="gdp"){
   library(fExtremes)
   library(extRemes)
   p<-fevd(x[,i], threshold=0, type="GP")$results$par
   p1<-p[1]
   p2<-p[2]
   prob[,i]<-pgpd(x[,i], xi=p2, mu=0, beta=p1)}
  if (distr[i]=="gev"){
```

```
  library(fExtremes)
  library(extRemes)
  p<-fevd(x[,i], type="GEV")$results$par
  p1<-p[1]
  p2<-p[2]
  p3<-p[3]
  prob[,i]<-pgev(x[,i], xi=p3, mu=p1, beta=p2)}}
prob}
```

B.8. AJUSTE DAS CÓPULAS

Biblioteca: library(copula)

Função: fitCopula(copula, x, method="mpl",start=0)

- copula: tipo de cópula: claytonCopula(); frankCopula(); gumbelCopula(); joeCopula(); amhCopula() e normalCopula();
- x: matriz numérica contendo as probabilidades acumuladas das variáveis. Cada coluna representa uma variável. Matriz calculada pela função cop.data; e
- method: método de ajuste dos parâmetros (máximo pseudo verossimilhança).

Valores: Parâmetro(s) estimado(s) e AIC.

B.9. AIC DAS CÓPULAS AJUSTADAS

Função: cop.aic(copula, x)

- copula: tipo de cópula: “cla”, “fra”; “gum”; “joe, “amh” e “gau”; e
- x: matriz numérica contendo as probabilidades acumuladas das variáveis. Cada coluna representa uma variável. Matriz calculada pela função cop.data.

Valores: AIC.

Na sequência são apresentados os detalhes de implementação da função.

```
cop.aic<-function(copula, x, ...){
 library(copula)
```

```
if (copula=="cla"){
 fit<-fitCopula(claytonCopula(), x, method="mpl", start=0)
 clayton.cop<-claytonCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=clayton.cop)}
if (copula=="fra"){
 fit<-fitCopula(frankCopula(), x, method="mpl", start=0)
 frank.cop<-frankCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=frank.cop)}
if (copula=="gum"){
 fit<-fitCopula(gumbelCopula(), x, method="mpl", start=0)
 gumbel.cop<-gumbelCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=gumbel.cop)}
if (copula=="joe"){
 fit<-fitCopula(joeCopula(), x, method="mpl", start=0)
 joe.cop<-joeCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=joe.cop)}
if (copula=="amh"){
 fit<-fitCopula(amhCopula(), x, method="mpl", start=0)
 amh.cop<-amhCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=amh.cop)      }
if (copula=="gau"){
 fit<-fitCopula(normalCopula(), x, method="mpl", start=0)
 normal.cop<-normalCopula(coef(fit), dim=2)
 aic<--2*loglikCopula(coef(fit), u=x, copula=normal.cop)}
aic}
```

B.10. TESTE GOF

Função: gof.test<-function(x, copula, n)

- x: matriz numérica contendo as probabilidades acumuladas das variáveis. Cada coluna representa uma variável. Matriz calculada pela função cop.data;
- copula: tipo de cópula: “cla”, “fra”; “gum”; “joe, “amh” e “gau”; e

- n: número de pseudo observações.

Valores: p-Valor.

Na sequência são apresentados os detalhes de implementação da função.

```
gof.test<-function(x, copula, n, ...){
 library(copula)
 if (copula=="cla"){
  fit<-fitCopula(claytonCopula(), x, method="mpl", start=0)
  clayton.cop<-claytonCopula(coef(fit), dim=2)
  r<-gofCopula(clayton.cop, x, N=n)}
 if (copula=="fra"){
  fit<-fitCopula(frankCopula(), x, method="mpl", start=0)
  frank.cop<-frankCopula(coef(fit), dim=2)
  r<-gofCopula(frank.cop, x, N=n)}
 if (copula=="gum"){
  fit<-fitCopula(gumbelCopula(), x, method="mpl", start=0)
  gumbel.cop<-gumbelCopula(coef(fit), dim=2)
  r<-gofCopula(gumbel.cop, x, N=n)                 }
 if(copula=="joe"){
  fit<-fitCopula(joeCopula(), x, method="mpl", start=0)
  joe.cop<-joeCopula(coef(fit), dim=2)
  r<-gofCopula(joe.cop, x, N=n)}
 if (copula=="amh"){
  fit<-fitCopula(amhCopula(), x, method="mpl", start=0)
  amh.cop<-amhCopula(coef(fit), dim=2)
  r<-gofCopula(amh.cop, x, N=n) }
 if (copula=="gau"){
  fit<-fitCopula(normalCopula(), x, method="mpl", start=0)
  normal.cop<-normalCopula(coef(fit), dim=2)
  r<-gofCopula(normal.cop, x, N=n) }
r$p.value}
```

B.11. SIMULAÇÃO DE EVENTOS COM CÓPULAS

Função: cop.sim (x, cop, y, distr=c("gamma","weibull","lnorm","exp","gdp","gev"), s)

- x: matriz numérica contendo as probabilidades acumuladas das variáveis. Cada coluna representa uma variável. Matriz calculada com a função cop.data;
- copula: tipo de cópula: “cla”, “fra”; “gum”; “joe, “amh” e “gau”;
- y: matriz numérica contendo as variáveis observadas. Cada coluna representa uma variável;
- distr: distribuição de probabilidade: “gamma”; “weibull”, “lnorm”, “exp”, “gdp” ou “gev”; e
- n: número de pseudo observações.

Valores: Matriz numérica contendo as pseudo.observações.

Na sequência são apresentados os detalhes de implementação da função.

```
cop.sim<-function(x, cop, y, distr=c("gamma","weibull","lnorm","exp","gdp","gev"), s,
...){
 n<-length(distr)
 v<-matrix(nrow=s, ncol=n)
 library(copula)
 if (cop=="cla"){
  fit<-fitCopula(claytonCopula(), x, method="mpl", start=0)
  clayton.cop<-claytonCopula(coef(fit), dim=n)
  u<-rCopula(s,clayton.cop)
  p<-pCopula(u,clayton.cop)}
 if (cop=="fra"){
  fit<-fitCopula(frankCopula(), x, method="mpl", start=0)
  frank.cop<-frankCopula(coef(fit), dim=n)
  u<-rCopula(s,frank.cop)
  p<-pCopula(u,frank.cop)}
 if (cop=="gum"){
```

```
 fit<-fitCopula(gumbelCopula(), x, method="mpl", start=0)
 gumbel.cop<-gumbelCopula(coef(fit), dim=n)
 u<-rCopula(s,gumbel.cop)
 p<-pCopula(u,gumbel.cop)}
if (cop=="joe"){
 fit<-fitCopula(joeCopula(), x, method="mpl", start=0)
 joe.cop<-joeCopula(coef(fit), dim=n)
 u<-rCopula(s,joe.cop)
 p<-pCopula(u,joe.cop)}
if (cop=="amh"){
 fit<-fitCopula(amhCopula(), x, method="mpl", start=0)
 amh.cop<-amhCopula(coef(fit), dim=n)
 u<-rCopula(s,amh.cop)
 p<-pCopula(u,amh.cop)}
if (cop=="gau"){
 fit<-fitCopula(normalCopula(), x, method="mpl", start=0)
 normal.cop<-normalCopula(coef(fit), dim=n)
 u<-rCopula(s,normal.cop)
 p<-pCopula(u,normal.cop)}
 for (i in 1:n){
  if (distr[i]=="gamma"){
   library("fitdistrplus")
   p1<-fitdist(y[,i], "gamma")$estimate[1]
   p2<-fitdist(y[,i], "gamma")$estimate[2]
   v[,i]<-qgamma(u[,i], shape=p1, rate=p2)}
  if (distr[i]=="weibull"){
   library("fitdistrplus")
   p1<-fitdist(y[,i], "weibull")$estimate[1]
   p2<-fitdist(y[,i], "weibull")$estimate[2]
   v[,i]<-qweibull(u[,i], shape=p1, scale=p2)}
  if (distr[i]=="lnorm"){
   library("fitdistrplus")
```

```
  p1<-fitdist(y[,i], "lnorm")$estimate[1]
  p2<-fitdist(y[,i], "lnorm")$estimate[2]
  v[,i]<-qlnorm(u[,i], meanlog=p1, sdlog=p2)}
 if (distr[i]=="exp"){
  library("fitdistrplus")
  p1<-fitdist(y[,i], "exp")$estimate[1]
  v[,i]<-qexp(u[,i], rate=p1)}
 if (distr[i]=="gdp"){
  library(fExtremes)
  library(extRemes)
  p<-fevd(y[,i], threshold=0, type="GP")$results$par
  p1<-p[1]
  p2<-p[2]
  v[,i]<-qgpd(u[,i], xi=p2, mu=0, beta=p1) }
 if (distr[i]=="gev"){
  library(fExtremes)
  library(extRemes)
  p<-fevd(y[,i], type="GEV")$results$par
  p1<-p[1]
  p2<-p[2]
  p3<-p[3]
  v[,i]<-qgev(u[,i], xi=p3, mu=p1, beta=p2)}}
v}
```

B12. TESTE DE NORMALIDADE DE SHAPIRO-WILKS

Função: shapiro.test(x)

- x: vetor numérico contendo a série temporal de dados.

Valor: p-Valor.