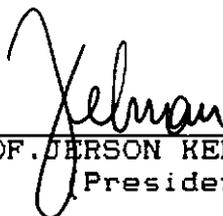


APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS
DE REAMOSTRAGEM EM
HIDROLOGIA

Fernanda da Serra Costa

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA CIVIL.

Aprovado por:



PROF. JERSON KELMAN, Ph.D
(Presidente)



PROF. BASÍLIO BRAGANÇA PEREIRA, Ph.D



DR. GERSON COUTO DE OLIVEIRA, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
OUTUBRO DE 1988

COSTA, FERNANDA DA SERRA

Aplicação de Técnicas Estatísticas de Reamostragem em Hidrologia (Rio de Janeiro) 1988.

VIII, 108 p. 29,7cm (COPPE-UFRJ, M.Sc., Engenharia Civil, 1988)

Tese - Universidade Federal do Rio de Janeiro, COPPE.

I. Assunto: Incerteza em Hidrologia I.COPPE/UFRJ,

II. Título (série)

AGRADECIMENTOS

Ao professor Jerson Kelman, pela dedicação e orientação segura e eficiente desta tese.

Ao engenheiro Jorge Machado Damázio, do CEPEL, pela grande participação no desenvolvimento desta tese, através de inúmeras críticas sugestões e revisão dos manuscritos.

Ao CNPq e CAPES pelo suporte financeiro.

Ao CEPEL, na pessoa do Engenheiro Leslie Afrânio Terry, chefe do Departamento de Sistemas, pelo apoio e incentivo prestados.

Ao professor Basílio Bragança Pereira, da COPPE, e ao Engenheiro Paulo Roberto Holanda Sales, da ELETROBRÁS, pelas sugestões, acompanhadas de forte incentivo.

A Ana Maria Costa Danielli pela dedicação e paciência no trabalho de datilografia, e as demais secretárias do Departamento de Sistemas do CEPEL pela ajuda prestada.

A amiga Maria Elvira Pineiro Maceira pelas sugestões, apoio e incentivo no decorrer deste trabalho.

Finalmente, aos meus pais e irmãos pelo apoio, confiança e carinho com que sempre acompanharam meus estudos.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários pra obtenção do grau de Mestre em Ciências (M.Sc.)

APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS DE REAMOSTRAGEM EM
HIDROLOGIA

Fernanda da Serra Costa

NOVEMBRO/1988

Orientador: Jerson Kelman

Programa: Civil

Nesta tese é abordado o problema da incerteza na estimação de parâmetros de modelos probabilísticos quando aplicados em Hidrologia.

A forma adotada para consideração desta incerteza é o uso de Intervalos de Confiança para os parâmetros em questão. Como o intervalo de confiança é função do desvio padrão dos estimadores dos parâmetros, são apresentados três metodologias alternativas ao uso da Estatística Clássica para o cálculo do estimador do desvio padrão destes parâmetros: o método Bayesiano, o Bootstrap e o Jackknife. Os dois últimos métodos são conhecidos como Técnicas de Reamostragem.

O método de Bootstrap é aplicado para analisar o problema da quantidade de informação a considerar na modelagem de fenômenos hidrológicos, especificamente na determinação do pico da cheia centenária.

Mostra-se que as Técnicas de Reamostragem são métodos razoavelmente precisos podendo ser adotados na construção de intervalos de confiança.

Abstract of this Thesis presented to COPPE/UFRJ as partial fulfillment of the requirements for the degree of Master of Sciences (M.Sc.)

APPLICATION OF RESAMPLING STATISTICAL TECHNIQUES IN
HYDROLOGY

Fernanda da Serra Costa

NOVEMBER/1988

Thesis Supervisor: Jerson Kelman
Departament: Civil

This thesis focus on the uncertainty regarding the numerical values of model parameters usually employed in Hydrology.

Confidence intervals are adopted in order to consider this uncertainty. Three methodologies are proposed to estimate the standard deviation of these parameters: the Bayesian, the Bootstrap and the Jackknife. Both Bootstrap and Jackknife are called Resampling Methods.

The Bootstrap is applied to decide the amount of information needed for modeling the one hundred year flood.

It is shown that the resampling techniques are reasonably accurate methods and can be applied to define confidence intervals.

ÍNDICE

CAPÍTULO I -	INTRODUÇÃO	01
CAPÍTULO II -	DESCRIÇÃO DAS METODOLOGIAS	05
	II.1 - Introdução	05
	II.2 - Metodologia Bayesiana	05
	II.2.1 - Distribuições Conjugadas Naturais	08
	II.2.2 - O Caso do Processo Normal	08
	II.3 - Técnicas de Reamostragem	14
	II.3.1 - O Jackknife	15
	II.3.2 - O Bootstrap	19
	II.4 - Sumário	22
CAPÍTULO III -	ESTUDOS COMPARATIVOS DOS MÉTODOS	23
	III.1 - Introdução	23
	III.2 - Características dos Estudos Realizados	24
	III.3 - Estudos de Casos	25
	III.3.1 - CASO 1: Desvio Padrão do Estimador da Vazão Média	25

III.3.2 - CASO 2: Desvio Padrão do Estimador da Vazão Decamilenar	39
III.3.3 - CASO 3 : Desvio Padrão do Tamanho do Reservatório para um Nível de Regulari- zação de 90%	55
III.4 - Conclusões	80
III.5 - Sumário	81
CAPÍTULO IV - COMPARAÇÃO DE METODOLOGIAS PARA MODELAGEM DE CHEIAS	82
IV.1 - Introdução	82
IV.2 - Metodologia 1: Bootstrap nos Máximo Anuais	84
IV.3 - Metodologia 2: Bootstrap nas Vazões Diárias	87
IV.4 - Estudos Realizados	91
IV.5 - Análise dos Resultados	94
IV.6 - Sumário	100
CAPÍTULO V - CONCLUSÕES	101
BIBLIOGRAFIA	103

CAPÍTULO I

INTRODUÇÃO

Modelos probabilísticos são desde muito tempo usados para descrever fenômenos hidrológicos (veja por exemplo FILLER [1]). A modelagem probabilística envolve duas etapas preliminares: a escolha do modelo e a estimação dos parâmetros do modelo a partir dos dados disponíveis. Consequentemente os hidrólogos envolvidos em modelagem probabilística deparam-se com dois tipos de incertezas nas etapas iniciais:

- . A incerteza na escolha do modelo que descreve o fenômeno em estudo (Exemplo: o pico da cheia sendo modelado por uma distribuição Gumbel ou Log-Pearson Tipo III);
- . A incerteza nos parâmetros, ou seja, incerteza na estimação dos parâmetros do modelo selecionado, causando o chamado "erro amostral de estimação".

Diversos pesquisadores já atentaram para o fato que os "erros amostrais de estimação", podem ser muito importantes. Como um exemplo prático podemos citar a estimação da curva de frequência de cheias, sujeita a erros amostrais que influenciam na decisão de projeto. Uma forma tradicional de se considerar a incerteza nos parâmetros de modelos probabilísticos é o uso dos chamados "Intervalos de Confiança", USWRC [2], ASHKAR et alia [3], cuja derivação formal pode ser feita através da abordagem estatística clássica. O intervalo de confiança de um parâmetro é usualmente dado em função do desvio padrão do seu estimador, sendo que em geral assume-se o estimador como normalmente distribuído. Assim, o intervalo de confiança de um parâmetro θ é dado por: $\hat{\theta} \pm z_{1-\alpha/2} S(\hat{\theta})$ ou

$$\hat{\theta} - z_{1-\alpha/2} S(\hat{\theta}) \leq \theta \leq \hat{\theta} + z_{1-\alpha/2} S(\hat{\theta}) \quad (1.1)$$

onde:

$\hat{\theta}$ - estimativa de θ

$z_{1-\alpha/2}$ - valor obtido da distribuição Normal padronizada

$S(\hat{\theta})$ - desvio padrão de $\hat{\theta}$

BENJAMIN et alia [4], descreve o método geral de derivação de intervalos de confiança.

O objetivo desta tese é avaliar a eficiência de quatro metodologias aqui propostas de cálculo do desvio padrão de estimadores. A precisão de cálculo do desvio padrão é fundamental para o cálculo do intervalo de confiança e conseqüentemente para consideração da incerteza paramétrica.

As metodologias propostas são o Método Clássico, o Método Bayesiano e duas Técnicas de Reamostragem: o Jackknife e o Bootstrap. A metodologia Bayesiana já é razoavelmente difundida na Hidrologia (DONALD et alia [5]). Os métodos que são apresentados como inovações são o Jackknife e o Bootstrap, técnicas desenvolvidas por QUENOUILLE [6] - TUKEY [7], EFRON [8], respectivamente, e que até alguns anos atrás tinham sua utilização limitada por exigirem o uso intensivo de computadores, apesar de sua simplicidade teórica.

A motivação para a apresentação destas novas técnicas, (Bayesiana, Jackknife e o Bootstrap) para estimação do desvio padrão de estimadores é que a Estatística Clássica nem sempre dispõe de fórmulas já derivadas e/ou facilmente encontráveis para a estimação de desvio padrão de estimadores, dependendo do parâmetro para qual se deseja avaliá-lo.

A descrição dos três métodos não clássicos faz parte do Capítulo II, enquanto que a avaliação da eficiência dos mesmos compreende o Capítulo III, onde são apresentados três casos exemplos.

O método Bootstrap, que é utilizado no Capítulo III como uma metodologia de cálculo do desvio padrão de estimadores pode também ser utilizado como uma forma de investigar um assunto que tem causado muitas discussões entre os pesquisadores da área de Hidrologia. Trata-se da divergência de opiniões quanto a escolha do modelo que descreve o fenômeno hidrológico. Alguns pesquisadores consideram mais conveniente observar apenas um aspecto específico do problema, construindo modelos mais simples, com um número reduzido de parâmetros, que entretanto consideram uma quantidade menor de informações; outros preferem modelar todo o fenômeno através de modelos mais complexos, e portanto com muitos parâmetros (possivelmente difíceis de serem estimados), mas que consideram uma quantidade maior de informações.

Um caso específico é a determinação do pico da cheia de T anos de recorrência, $y(T)$, para projeto de vertedores. A aproximação usual é ajustar uma distribuição de probabilidade $F(\cdot)$ ao conjunto de m máximos $y = \{y_1, y_2, \dots, y_m\}$, obtidos de um registro de vazões diárias e a partir de extrapolação obter $y(T)$, tal que $p(y \geq y(T)) = 1/T$. Entretanto, neste caso grande parte das informações contidas no registro fluviométrico são desprezadas.

Uma forma alternativa para estimar $y(T)$ consiste na utilização de um modelo estocástico para a série de vazões diárias (KELMAN [9]) e baseado neste modelo estimar $y(T)$. Como as vazões diárias, ao contrário do máximo anuais, são variáveis aleatórias dependentes e apresentam sazonalidade, o modelo estocástico utilizado neste caso seria um modelo mais complexo. A desvantagem causada pela complexidade do modelo em termos de dificuldade para estimação dos parâmetros pode ser compensada pelo grande número de informações que contribuirão na estimação do pico da cheia, conduzindo a uma possivelmente melhor estimativa de $y(T)$.

O Capítulo IV faz uso do Bootstrap sob duas formas distintas para avaliação do pico da cheia de 100 anos de retorno (vazão centenária) e seu desvio padrão. Uma delas utiliza apenas os máximos anuais, enquanto a outra observa todo o registro fluviométrico.

Finalmente no Capítulo V são apresentadas as conclusões e algumas recomendações para possíveis trabalhos futuros.

CAPÍTULO II

DESCRIÇÃO DAS METODOLOGIASII.1 INTRODUÇÃO

Neste capítulo apresentaremos três metodologias, a Bayesiana, o Jackknife e o Bootstrap, capazes de considerar a incerteza na estimação de parâmetros de modelos probabilísticos utilizados em Hidrologia. Uma vez que estes métodos permitem considerar a variabilidade amostral na estimação de parâmetros, estes se tornam ferramentas alternativas no cálculo do desvio padrão de estimadores dos parâmetros de modelos probabilísticos.

II.2 METODOLOGIA BAYESIANA

A metodologia Bayesiana é uma das formas mais conhecidas de abordar o problema da variabilidade amostral, na estimação de parâmetros, RAIFFA et alia [10]; BENJAMIN et alia [4].

Nesta metodologia os parâmetros desconhecidos de um modelo probabilístico são considerados variáveis aleatórias (v.a.) as quais se associa uma função densidade de probabilidade construída a partir de um conjunto de informações existentes.

Usualmente o interesse dos hidrólogos por esta metodologia deve-se a possibilidade de serem combinadas informações amostrais (resultados de sorteios) com não-amostrais (também conhecidas como informações priorísticas), na medida que se pode considerar como informações não-amostrais: dados regionais, considerações teóricas e julgamentos subjetivos fruto das experiências anteriores do hidrólogo (DAMÁZIO [11]). Para a combinação destes dois tipos de informações faz-se uso dos conceitos de

verossimilhança, distribuição à priori, distribuição posterior, que serão expostas nos itens subsequentes deste capítulo.

No entanto, mesmo quando não há interesse na consideração de informações não-amostrais (seja porque estas não existem, seja porque se deseja fazer um relato específico de uma experiência), a metodologia bayesiana pode ser aplicada através do uso das chamadas distribuições priorísticas não-informativas, que por sua especial relevância para as situações analisadas nos capítulos seguintes desta tese são comentadas com maiores detalhes posteriormente.

Seja a variável aleatória X cuja a função densidade de probabilidade, $f_X(x)$, é conhecida a menos de m parâmetros representados por $\underline{\theta} = \{\theta_1, \theta_2, \dots, \theta_m\}$. Para refletir o desconhecimento destes parâmetros, $\underline{\theta}$ é considerado uma variável aleatória existindo portanto, uma função densidade de probabilidade para $\underline{\theta}$, $f'(\underline{\theta} | I_r)$, denominada distribuição à priori que considera todas as informações priorísticas (I_r) sobre $\underline{\theta}$.

Suponhamos agora que esteja disponível um vetor \underline{x} constituído de n observações independentes da variável aleatória X , $\underline{x} = \{x_1, x_2, \dots, x_n\}$. A partir do Teorema de Bayes combina-se as informações em \underline{x} sobre $\underline{\theta}$ com $f'(\underline{\theta} | I_r)$, obtendo-se uma nova distribuição para $\underline{\theta}$:

$$f''(\underline{\theta} | I_r, \underline{x}) = \frac{f'(\underline{\theta} | I_r) L(\underline{x} | \underline{\theta})}{\int_{-\infty}^{\infty} L(\underline{x} | \underline{\theta}) \cdot f'(\underline{\theta} | I_r) d\underline{\theta}} \quad (II.1)$$

onde

$L(\underline{x} | \underline{\theta})$ é a função verossimilhança de \underline{x} ,

$f'(\theta | Ir)$ distribuição à priori de θ ,

$f''(\theta | Ir, x)$ distribuição posterior de θ .

e a integral do denominador é a constante de normalização $\phi(x)$, que tem o objetivo de tornar $f''(\theta | Ir, x)$ uma função densidade de probabilidade, isto é:

$$\int_{-\infty}^{\infty} f''(\theta | Ir, x) d\theta = \phi^{-1}(x) \int_{-\infty}^{\infty} f'(\theta | Ir) L(x | \theta) d\theta = 1 \quad (II.2)$$

O valor esperado de θ considerando a distribuição posterior de θ , $f''(\theta | Ir, x)$ é denominado Estimador Bayesiano de θ , e é calculado por:

$$\hat{\theta}_{\text{Bayes}} = E[\theta | x] = \int \theta f''(\theta | Ir, x) d\theta \quad (II.3)$$

A matriz de covariância de θ considerando $f''(\theta | x)$ é denominada Variância Bayesiana de θ e é calculada por

$$S_{\text{Bayes}}^2 = E[(\theta - \hat{\theta}_{\text{Bayes}})^2] = \int (\theta - \hat{\theta}_{\text{Bayes}})^2 f''(\theta | Ir, x) d\theta \quad (II.4)$$

A raiz quadrada de um elemento da diagonal de S_{Bayes}^2 fornece o Desvio Padrão Bayesiano do correspondente parâmetro.

A metodologia bayesiana pode também ser aplicada em processos estocásticos, BOX et alia [12].

II.2.1 Distribuições Conjugadas Naturais

Muitas vezes as informações priorísticas (Ir) não definem uma função densidade de probabilidade à priori para θ . Neste caso é possível utilizar uma família de funções densidade de probabilidades convenientes para definir a distribuição à priori de θ , bastando ajustar os parâmetros destas funções densidade de probabilidades de forma a representar as informações priorísticas. Estas famílias de distribuições são denominadas Distribuições Conjugadas Naturais e quando escolhidas facilitam os cálculos de obtenção da distribuição posterior dos parâmetros θ , as quais passam a ser expressas da mesma forma matemática que a distribuição à priori, distinguindo-se apenas no valor dos parâmetros. Por outro lado, se as informações priorísticas (Ir) indicam evidências de uma particular distribuição à priori, que não a conjugada do processo, esta distribuição deve ser utilizada, não obstante as complicações matemáticas que serão introduzidas na determinação da posterior.

Em geral, nas Distribuições Conjugadas Naturais existe um parâmetro descriptor da quantidade de informações (cujo inverso é uma medida da incerteza existente) para cada parâmetro em θ . Se na distribuição à priori algum destes parâmetros for zero, é porque não existem ou não serão consideradas informações priorísticas para o correspondente parâmetro em θ .

II.2.2 O Caso do Processo Normal

Neste item serão apresentados em detalhes resultados necessários para o cálculo do desvio padrão Bayesiano do estimador de parâmetros de um processo normal independente. A razão para tal, é que no capítulo III são abordados três problemas, nos quais as variáveis aleatórias básicas são independentes e possuem distribuição de

probabilidades log-normal. Consequentemente, o logaritmo destas variáveis são variáveis aleatórias normalmente distribuídas.

Define-se por processo normal independente, o processo gerador de variáveis aleatórias independentes, identicamente distribuídas x_1, x_2, \dots , com função densidade de probabilidade definida por:

$$f_{x|\mu,\sigma}(x_i|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp \left\{ -\frac{1}{2} \left(\frac{x_i - \mu}{\sigma} \right)^2 \right\} \quad (II.5)$$

$$-\infty \leq \mu \leq \infty, \quad \sigma \geq 0$$

sendo μ e σ , os parâmetros desconhecidos.

A função verossimilhança de um vetor $\tilde{x} = \{ x_1, x_2, \dots, x_n \}$ é dada por:

$$L(\mu, \sigma | \tilde{x}) = \prod_{i=1}^n f_{x|\mu,\sigma}(x_i|\mu,\sigma) \quad (II.6)$$

Neste caso temos:

$$L(\mu, \sigma | \tilde{x}) = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} \quad (II.7)$$

$$-\infty \leq \mu \leq \infty, \quad \sigma \geq 0$$

A distribuição conjugada natural à priori dos parâmetros μ e σ é uma distribuição bivariada, produto de uma Normal e uma Gama-2 Invertida, Raiffa¹⁰, assim

$$f'(\mu, \sigma | I_r) = f_N(\mu | \bar{x}', \sigma' / \sqrt{n'}) \cdot f_{IG2}(\sigma | s', v') \quad (II.8)$$

onde n', v', \bar{x}' e s'^2 são parâmetros priorísticos, sendo

$-\infty \leq \bar{x}' \leq \infty$ e $v' \geq 0$, $s'^2 \geq 0$, $n' \geq 0$.

As distribuições marginais priorísticas de μ e σ são obtidas por integração da expressão (II.8). No caso da distribuição marginal de μ , integra-se (II.8) em relação a σ no intervalo $\sigma \geq 0$, resultando na distribuição Student,

$$f_S(\mu | \bar{x}', s', n', v') \propto [v' + (\mu - \bar{x}')^2 n' / s'^2]^{-\frac{(v'+1)}{2}} \quad (II.9a)$$

Integrando (II.8) em relação a μ no intervalo $-\infty \leq \mu \leq \infty$, obtém-se para distribuição marginal de σ uma Gama-2 Invertida,

$$f_{IG2}(\sigma | s', v') \propto \left[\exp \left\{ -\frac{1}{2} \frac{s'^2}{\sigma^2} v' \right\} \right] \cdot \sigma^{-(v'+1)} \quad (II.9b)$$

Aplicando-se o Teorema de Bayes (II.1), obtém-se a distribuição posterior dos parâmetros, dada por:

$$f''_{\mu, \sigma}(\mu, \sigma | I_r, x) = \phi^{-1} f'_{\mu, \sigma}(\mu, \sigma | I_r) \cdot L(\mu, \sigma | x) \quad (II.10)$$

O que nos dá:

$$f''_{\mu, \sigma}(\mu, \sigma | I_r, x) = f''_N(\mu | \bar{x}'', \sigma / \sqrt{n}'') \cdot f''_{IG2}(\sigma | s'', v'') \quad (II.11)$$

Como se esperava a distribuição posterior dos parâmetros é o produto de uma Normal com uma Gama-2 Invertida, cujos parâmetros (ditos posteriores) são:

$$n'' = n' + n \quad (II.12a)$$

$$\bar{x}'' = \frac{n' \bar{x}' + n \bar{x}}{(n' + n)} \quad (II.12b)$$

$$v'' = [v' + \delta(n')] + [v + \delta(n)] - \delta(n'') \quad (II.12c)$$

$$s''^2 = \frac{v' s'^2 + n' \bar{x}'^2 + v s^2 + n \bar{x}^2 - n'' \bar{x}''^2}{v''} \quad (\text{II.12d})$$

$$\delta(n) = \begin{cases} 0 & \text{se } n = 0 \\ 1 & \text{se } n > 0 \end{cases} \quad (\text{II.12e})$$

onde:

n = tamanho da amostra

$$v = n - 1 \quad (\text{II.13a})$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{II.13b})$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{se } n > 1 \quad (\text{II.13c})$$

$$s^2 = 0 \quad \text{se } n = 1 \quad (\text{II.13d})$$

Quando não existem informações priorísticas $n' \approx 0$ e $v' \approx 0$, conseqüentemente os parâmetros posteriores tornam-se iguais as estatísticas amostrais, o que pode ser facilmente concluído observando-se as expressões (II.12).

As distribuições marginais dos parâmetros desconhecidos μ e σ podem ser determinadas integrando-se (II.11). Deste modo a distribuição de μ será uma distribuição Student similar a (II.9a).

$$f_{\Sigma}''(\mu | \bar{x}'', s'', n'') \propto [v'' + (\mu - \bar{x}'')^2 n''/s''^2]^{-\frac{(v'' + 1)}{2}} \quad (\text{II.14})$$

Aplicando-se o valor esperado e a variância em (II.14), obtém-se:

$$\hat{\mu}_{\text{Bayes}} = \bar{x}'' \quad (\text{II.15a})$$

$$\hat{s}_{\text{Bayes}}(\mu) = \left\{ s^{n''} \frac{v^{n''}}{n^{n''}(v^{n''}-2)} \right\}^{1/2} \quad (\text{II.15b})$$

De forma análoga obtém-se a distribuição de σ , $f''_{\text{IG2}}(\sigma | s^{n''}, v^{n''})$, que é uma Gama-2 Invertida de mesma forma que (II.9b).

$$f''_{\text{IG2}}(\sigma | s^{n''}, v^{n''}) \propto \left[\exp \left\{ -\frac{1}{2} \frac{s^{n''2}}{\sigma^2} v^{n''} \right\} \right] \cdot \sigma^{-(v^{n''}+1)} \quad (\text{II.16})$$

O valor esperado e a variância do desvio padrão σ , são:

$$\hat{\sigma}_{\text{Bayes}} = s^{n''} \left(\frac{v^{n''}}{2} \right)^{1/2} \left[\frac{\Gamma \left(\frac{v^{n''}-1}{2} \right)}{\Gamma(v^{n''}/2)} \right] \quad v^{n''} > 1 \quad (\text{II.17a})$$

$$s_{\text{Bayes}}^2(\sigma) = \frac{v^{n''} s^{n''2}}{v^{n''}-2} - \hat{\sigma}_{\text{Bayes}}^2 \quad v^{n''} > 2 \quad (\text{II.17b})$$

A distribuição de σ^2 é proporcional a distribuição qui-quadrado com $(n^{n''}-3)$ graus de liberdade, Raiffa¹⁰,

$$\frac{(n^{n''}-3)}{\sigma^2} s^{n''2} \sim \text{CH}(n^{n''} - 3) \quad (\text{II.18})$$

A metodologia Bayesiana pode ser também usada como uma forma de determinar a distribuição de probabilidades de um quantil x_α , função dos parâmetros μ e σ , a partir da distribuição conjunta destes, $f_{\mu, \sigma}(\mu, \sigma)$, senão vejamos:

Seja o quantil $x_\alpha = \mu + k_\alpha \sigma$, logo $\mu = x_\alpha - k_\alpha \sigma$, substituindo a expressão de μ em $f_{\mu, \sigma}(\mu, \sigma)$, (II.8) e integrando-se em relação a σ , no intervalo $\sigma \geq 0$ obtém-se a densidade de probabilidade do quantil x_α , isto é:

$$f''_{X_\alpha}(x_\alpha) = \int_0^\infty f_{\mu, \sigma}(x_\alpha - k_\alpha \sigma, \sigma) d\sigma \quad (\text{II.19})$$

Esta integração pode ser feita analiticamente ou numericamente. A integração por Monte Carlo neste caso consiste na utilização da distribuição conjunta de μ e σ (II.8) da qual se sorteiam valores destes parâmetros e para cada par (μ, σ) determina-se o quantil x_α correspondente. A repetição deste procedimento inúmeras vezes permite a obtenção de uma aproximação para a distribuição de x_α .

A Metodologia Bayesiana, é adotada por vários pesquisadores para estimação de parâmetros de modelos probabilísticos utilizados em Hidrologia, (STEDINGER et alia [13], [14], VICENS et alia [15], DAMÁZIO [11]) e como foi visto, pode ser também utilizada com forma de determinação do desvio-padrão dos estimadores destes parâmetros. Entretanto esta metodologia apresenta basicamente dois problemas:

- É difícil definir a distribuição à priori dos parâmetros mesmo quando não existe informação priorística. Considere, por exemplo o caso da distribuição normal. A análise do comportamento assintótico da distribuição marginal à priori, definida em (II.9a), indica que quando n' tende para zero a distribuição marginal para μ torna-se progressivamente uniforme sobre o intervalo $-\infty \leq \mu \leq \infty$, mas quando v' tende para zero a razão da densidade marginal de σ (II.9b) em dois pontos σ_1 e σ_2 tende para σ_2/σ_1 (RAIFFA et alia [10]) indicando menor probabilidade para valores altos de σ . A discussão pode ainda ser levada a analisar o fato que quando n' tende a zero a distribuição para $1/\mu$ se aproxima de uma massa na origem abrindo a questão da "arbitrariedade da parametrização". A solução normalmente adotada consiste em se considerar como não informativas as distribuições à priori que podem ser "substancialmente modificadas por um pequeno número de observações". Ou seja, a adoção de $n' = 0$ e $v' = 0$ em (II.12) não corresponde propriamente a usar uma distribuição à priori não informativa. Na realidade neste caso

utiliza-se a distribuição à priori menos informativa possível.

- . Se a distribuição, $f_X(x)$, da variável básica adotada para resolução do problema não for a correta, a aplicação da metodologia conduz a resultados incorretos.

II.3 TÉCNICAS DE REAMOSTRAGEM

As metodologias apresentadas a seguir, denominadas Técnicas de Reamostragem ou Métodos Não-Paramétricos, fazem uso apenas dos dados (registro fluviométrico) para descrever a distribuição de probabilidades dos parâmetros de um modelo, sem necessidade de admitir o conhecimento da função densidade de probabilidade da variável básica (UNNY et alia [16], SANTI [17]), não estando, portanto, sujeitos aos problemas expostos anteriormente.

A utilização destes métodos foi durante muito tempo limitada por exigir o uso intensivo de computadores. Com o surgimento das novas gerações de computadores, verificou-se um aumento na capacidade de armazenamento e velocidade de processamento dos mesmos. Assim cada vez mais tem-se estudado e sugerido a utilização destes métodos (EFRON [18]).

Em geral, estas técnicas se baseiam na produção de "pseudo-amostras" a partir da reamostragem do conjunto de observações. A cada "pseudo-amostra" são calculadas estimativas dos parâmetros. Com as várias estimativas dos parâmetros pode-se definir as distribuições de probabilidade empírica do mesmo, e em particular o desvio padrão. Apesar da existência de vários esquemas de reamostragem (EFRON [19]) estudaremos apenas o Jackknife e Bootstrap.

Uma limitação destes métodos é que não é possível utilizar as informações não-amostrais disponíveis.

II.3.1 O Jackknife

A idéia original do Jackknife foi proposta por Quenouille⁶ com a finalidade de reduzir a tendência do estimador da correlação serial, baseando-se na divisão da amostra em duas "meias-amostras". Em outro trabalho do mesmo autor (QUENOUILLE [20]) esta técnica foi generalizada e aplicada a outros problemas. Posteriormente, TUKEY [7], utilizou a mesma técnica para estimar a variância de uma estatística θ . O uso geral desta técnica foi proposto em TUKEY [21], de modo a reduzir a tendência de estimadores e obter intervalos de confiança aproximados em problemas onde os procedimentos estatísticos clássicos eram difíceis de serem aplicados, ou não existiam. Tukey batizou a nova técnica de Jackknife com intenção de exprimir a idéia de uma ferramenta grosseira, porém eficiente e útil em uma grande variedade de situações.

O Jackknife para determinação do estimador $\hat{\theta}_{\text{Jackk}}$ do parâmetro θ e de seu desvio padrão \hat{S}_{Jackk} consiste em, dado uma amostra $\underline{x} = \{x_1, x_2, \dots, x_n\}$ contendo n sorteios independentes da variável aleatória X , obter-se uma "pseudo-amostra" $\underline{x}_{\text{Jackk}}^i$, omitindo-se a i -ésima observação da amostra,

$$\underline{x}_{\text{Jackk}}^i = \{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n\} \quad (\text{II.20})$$

A estimativa da estatística θ , calculada sobre a "pseudo-amostra", $\underline{x}_{\text{Jackk}}^i$ é então,

$$\hat{\theta}_{-i} = g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \quad (\text{II.21})$$

Onde $g(\cdot)$ é uma função que mapeia a amostra $\underline{x}_{\text{Jackk}}^i$ em $\hat{\theta}_{-i}$.

Desta forma é possível calcular n estimativas de $\hat{\theta}_{-i}$.

A estimativa Jackknife de tendenciosidade determinada por Quenouille (1949), é dada por:

$$\text{BIAS}_{\text{Jackk}} = (n-1) (\hat{\theta}_{(.)} - \hat{\theta}) \quad (\text{II.22})$$

onde $\hat{\theta}$ é a estimativa clássica da estatística θ sobre a amostra x ,

$$\hat{\theta} = g(x_1, x_2, \dots, x_n) \quad (\text{II.23})$$

e $\hat{\theta}_{(.)}$ é o valor esperado dos $\hat{\theta}_{-i}$, isto é:

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{-i} \quad (\text{II.24})$$

O raciocínio usual para $\text{BIAS}_{\text{Jackk}}$ (Efron¹⁹) é o seguinte:

Seja E_n o valor esperado da estatística $\hat{\theta}$, $E_n = E[\hat{\theta}]$, para maioria das estatísticas, tem-se:

$$E_n = \theta + \frac{a_1(F)}{n} + \frac{a_2(F)}{n^2} + \dots \quad (\text{II.25})$$

onde F é a distribuição de probabilidades dos X 's e $a_1(F)$, $a_2(F)$ são funções que não dependem de n .

De modo análogo E_{n-1} é o valor esperado de $\hat{\theta}_{(.)}$,

$$E_{n-1} = \theta + \frac{a_1(F)}{(n-1)} + \frac{a_2(F)}{(n-1)^2} + \dots \quad (\text{II.26})$$

Baseado em (II.25) quando n tende a infinito $E_{\infty} = \theta$.

A figura (II.1) apresenta o gráfico ($E_n \times 1/n$), aproximando a curva da tendenciosidade linearmente temos:

$$\hat{\text{BIAS}}_{\text{Jackk}} = E_n - E_{\infty} = (n-1)[E_{n-1} - E_n] \quad (\text{II.27})$$

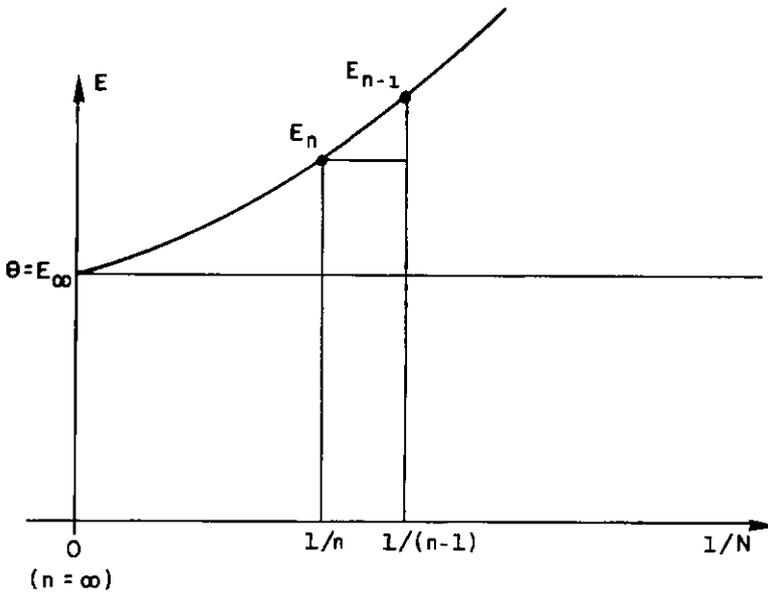


FIGURA II.1-VALOR ESPERADO COMO FUNÇÃO DE $1/n$

que é exatamente a expressão (II.22), quando E_n e E_{n-1} são substituídos pelas suas estimativas não tendenciosas $\hat{\theta}$ e $\hat{\theta}_{(.)}$, respectivamente.

O estimador Jackknife de θ , representado por $\hat{\theta}_{\text{Jackk}}$ é obtido corrigindo-se $\hat{\theta}$ da tendenciosidade, assim:

$$\begin{aligned} \hat{\theta}_{\text{Jackk}} &= \hat{\theta} - \hat{\text{BIAS}}_{\text{Jackk}} \\ &= \hat{\theta} - (n-1)[\hat{\theta}_{(.)} - \hat{\theta}] \\ &= \hat{\theta} + (n-1)\hat{\theta} - (n-1)\hat{\theta}_{(.)} \\ &= n \hat{\theta} - (n-1) \hat{\theta}_{(.)} \end{aligned} \quad (\text{II.28})$$

Utilizando as expressões (II.25) e (II.26) temos que:

$$E[\hat{\theta}_{\text{Jackk}}] = nEn - (n-1) En-1$$

$$= \theta - \frac{a_2(F)}{n(n-1)} + a_3(F) \left(\frac{1}{n^2} + \frac{1}{(n-1)^2} \right) + \dots \quad (\text{II.29})$$

O que nos leva a concluir que enquanto $\hat{\theta}_{\text{Jackk}}$ é tendencioso da ordem de $(1/n^2)$, a tendenciosidade de $\hat{\theta}$ é da ordem de $(1/n)$.

Muitas vezes a correção da tendenciosidade (II.22) pode aumentar o erro médio quadrático devido ao acréscimo na variância, em vez de diminuí-lo através da remoção da tendenciosidade. HINKLEY [22] discute este efeito para o caso do coeficiente de correlação. Entretanto, pode ser ainda interessante computar a $\hat{\text{BIAS}}_{\text{Jackk}}$, mesmo quando a correção da tendenciosidade não for feita, especialmente em conjunto com a variância.

O estimador Jackknife do desvio padrão de estimadores é dado por (EFRON [19]):

$$\hat{S}_{\text{Jackk}} = \left\{ \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{-i} - \hat{\theta}_{(.)})^2 \right\}^{1/2} \quad (\text{II.30})$$

Uma outra forma de se aplicar o Jackknife é dividir a amostra $\underline{x} = \{x_1, \dots, x_n\}$ em w grupos de tamanho h cada, onde $n=wh$, ou seja:

$$\underline{x} = \{x_1, \dots, x_h, x_{h+1}, \dots, x_{2h}, x_{2h+1}, \dots, x_{(n-1)h+1}, \dots, x_{hn}\}$$

$$(II.31)$$

e determinar $\hat{\theta}_{-i}$ como a estimativa do parâmetro θ tendo omitido o i -ésimo grupo de observações.

A maioria das pesquisas sobre o Jackknife tem sido desenvolvidas para o caso especial de $h=1$ e $w=n$, isto porque elimina-se qualquer arbitrariedade na formação dos grupos, e é de consenso entre os pesquisadores que esta é a melhor forma de se usar o Jackknife em qualquer problema, exceto quando a quantidade de dados é muito grande, tornando-o inviável computacionalmente. Entretanto na maioria dos casos, se não em todos, os resultados obtidos para $h=1$ são extendidos para $h > 1$ (MILLER [23]).

II.3.2 O BOOTSTRAP

O Bootstrap desenvolvido por Efron (EFRON [8],[19]) constitui um esquema de reamostragem de maior aplicabilidade que o Jackknife, apesar de exigir um esforço computacional maior.

Esta técnica de reamostragem calcula o parâmetro de interesse ou o conjunto de parâmetros, B vezes, utilizando para tal, B "pseudo-amostras" obtidas por reamostragem com reposição das observações da amostra original. O conjunto de estimativas do parâmetro calculadas desta forma constituem a distribuição empírica de probabilidade do mesmo. Assim como no Jackknife a independência dos dados é pré-requisito.

O algoritmo de Monte Carlo para a determinação da distribuição empírica de probabilidades do parâmetro em estudo, assim como para o cálculo da estimativa e de seu desvio padrão, por Bootstrap, é o a seguir descrito (EFRON [19]).

1. Faz-se uma reamostragem com reposição das observações da amostra x , fornecendo uma "pseudo-amostra"

(ou amostra Bootstrap) $x_{\sim\text{Boot}}^b$

$$\tilde{x}_{\text{Boot}}^b = (x_1^*, x_2^*, \dots, x_n^*) \quad (11.32)$$

2. A partir da "pseudo-amostra" $\tilde{x}_{\text{Boot}}^b$ pode-se calcular a estimativa do parâmetro de interesse,

$$\hat{\theta}_{\text{Boot}}^b = g(x_1^*, x_2^*, \dots, x_n^*) \quad (11.33)$$

Repetições independentes dos passos 1 e 2, fornecem $\hat{\theta}_{\text{Boot}}^1, \hat{\theta}_{\text{Boot}}^2, \dots, \hat{\theta}_{\text{Boot}}^B$, estimativas do parâmetro θ , com as quais é possível determinar a distribuição empírica de probabilidade de θ .

O estimador Bootstrap do parâmetro θ é definido como o valor esperado de $\hat{\theta}_{\text{Boot}}^b$:

$$\hat{\theta}_{\text{Boot}} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{\text{Boot}}^b \quad (11.34)$$

A estimativa Bootstrap da tendenciosidade é dada por:

$$\hat{\text{BIAS}}_{\text{Boot}} = \hat{\theta}_{\text{Boot}} - \hat{\theta} \quad (11.35)$$

E o estimador Bootstrap do desvio padrão de estimadores é (Efron¹⁹):

$$\hat{S}_{\text{Boot}} = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_{\text{Boot}}^b - \hat{\theta}_{\text{Boot}})^2 \right]^{1/2} \quad (11.36)$$

Enquanto no Jackknife o número de pseudo-amostras

x_{\sim}^i Jackk é limitado ao tamanho da amostra x_{\sim} , no Bootstrap o número de possíveis "pseudo-amostras", x_{\sim}^b Boot, é relativamente ilimitado dado que são possíveis $\binom{2n-1}{n}$ reamostras (UNNY [16]). Este resultado provém do estudo de análise combinatória, no qual $C_{m,n}^r$ representa o número possível de combinações com reposição de uma amostra, onde o índice m indica o tamanho da pseudo-amostra (combinação), o índice n o tamanho da amostra e o índice r indica que a reamostragem (combinação) é com reposição. A combinação com reposição equivale a uma combinação sem reposição considerando uma amostra de tamanho $m+n-1$, de modo que o número de possíveis combinações de tamanho m com reposição, de uma amostra de tamanho n é dado por: $C_{m,n}^r = C_{m, n+m-1} = \binom{n+m-1}{m}$. Como no Bootstrap o tamanho da combinação é n , igual ao tamanho da amostra, tem-se $C_{n,n} = C_{n, n+n-1} = \binom{n+n-1}{n} = \binom{2n-1}{n}$. Como ilustração no caso de $n=10$, que corresponde a uma amostra bastante pequena, são possíveis 92378 combinações ("pseudo-amostras").

Apesar do Bootstrap ser computacionalmente mais exigente que o Jackknife em geral os seus resultados são melhores. A tabela II.1 apresenta os resultados de uma comparação (EFRON [19]) entre diversos esquemas de reamostragem para cálculo do desvio padrão da estimativa da correlação entre duas variáveis normais. O experimento consistiu em 200 sorteios de $\{x_1, \dots, x_{14}\}$ onde X_i é um par (U_i, V_i) , da distribuição bivariada normal padrão com $\rho = \text{cov}(U_i, V_i) = 1/2$. Para cada sorteio, obteve-se a estimativa do desvio padrão de ρ através do Bootstrap com $B = 128$ e 512 , através do Jackknife e através da Estatística Clássica, $\hat{S}_{\text{Class}} = (1 - \hat{\rho}^2) / (n-3)^{1/2}$ (neste caso $n=14$). A tabela II.1 apresenta a média, desvio padrão, coeficiente de variação e a raiz quadrada do erro médio quadrático (em relação ao valor real, 0.218), destas estimativas. Pode-se notar que o Bootstrap tem um comportamento melhor que o do Jackknife. Efron¹⁹ apresenta uma comparação mais abrangente entre os dois métodos que favorece o Bootstrap.

Método	Média	Desvio Padrão	Coef. Variação	EMQ ^{1/2}
Bootstrap B = 128	0,206	0,066	0,32	0,067
Bootstrap B = 512	0,206	0,063	0,31	0,064
Jackknife	0,223	0,085	0,38	0,085
Método Clássico	0,217	0,056	0,26	0,056
Valor Verdadeiro	0,218			

Tabela II.1 - Média, desvio padrão, coeficiente de variação e raiz quadrada do erro médio quadrado (EMQ^{1/2}) do estimador do desvio padrão de ρ .

II.4 SUMÁRIO

Neste capítulo foram apresentados três métodos alternativos de cálculo de desvio padrão de estimadores de parâmetros de modelos probabilísticos, o Bayesiano, o Jackknife e o Bootstrap.

O método Bayesiano quando utilizado permite a combinação de informações, amostrais e não amostrais na determinação do desvio padrão de estimadores, para tal faz uso dos conceitos de verossimilhança, distribuição "à priori" e posterior.

O Jackknife e o Bootstrap denominados técnicas de reamostragem (ou Método não Paramétrico), baseiam-se apenas na reamostragem do conjunto de observações, não estando portanto limitados a considerações sobre as distribuições da variável aleatória em estudo e as distribuições à priori.

No capítulo subsequente estes três métodos serão utilizados para cálculo do desvio padrão de estimadores em três problemas hidrológicos clássicos.

CAPÍTULO III

ESTUDOS COMPARATIVOS DOS MÉTODOSIII.1. INTRODUÇÃO

Neste capítulo é abordado um dos problemas mencionados na introdução desta tese, com o qual pesquisadores da área de Hidrologia têm se confrontado ao longo dos tempos e que consiste na incerteza na estimação de parâmetros de modelos probabilísticos, decorrentes da variabilidade amostral.

Uma das formas tradicionais de considerar este problema é o uso de intervalos de confiança, Ashkar³. Nosso interesse entretanto, não está voltado para a determinação do método geral de cálculo de Intervalos de Confiança, que pode ser encontrado na maioria da literatura clássica de Estatística (BENJAMIM et alia [4], YEVJEVICH [24]), mas sim para a estimação do desvio padrão de estimadores, uma vez que o intervalo de confiança é função deste.

Para estimação do desvio padrão de estimadores são utilizados, neste capítulo, as três metodologias descritas no capítulo II: a Bayesiana, o Jackknife e o Bootstrap, como alternativas ao uso da Estatística Clássica, que nem sempre dispõe de fórmulas já derivadas e/ou facilmente encontráveis para sua estimação.

O objetivo deste capítulo é portanto a avaliação das metodologias propostas quando aplicadas a problemas de Hidrologia.

Suponhamos para tal que um hidrólogo precise estimar para uma dada seção de um rio a vazão média, a vazão de cheia decamilenar e o tamanho do reservatório para um nível de regularização de 90%, a partir de um registro de 40 anos de vazões. Se procurará então avaliar estes métodos na:

- . determinação do estimador do desvio padrão da vazão média;
- . determinação do estimador do desvio padrão da vazão decamilenar;
- . determinação do estimador do desvio padrão do tamanho do reservatório para um nível de regularização de 90%.

A apresentação dos estudos nesta ordem, tem como objetivo mostrar a crescente dificuldade na determinação do desvio padrão de estimadores, $S(\hat{\theta})$, pelo Método Clássico quando a complexidade do problema aumenta, o que ocorre em proporções muito menores quando as metodologias propostas são adotadas. Em outras palavras, no caso de parâmetros (θ) usuais, tais como média e variância de processos aleatórios, o hidrólogo encontrará na Estatística Clássica fórmulas já consagradas para o cálculo de $S(\hat{\theta})$. Entretanto para parâmetros mais específicos, como veremos nos itens subsequentes, estas fórmulas precisam ainda ser derivadas. Neste caso o hidrólogo tem como alternativa o uso das chamadas "Técnicas de Reamostragem" (EFRON [19]), através das quais o cálculo de $S(\hat{\theta})$ é extremamente simples e intuitivo, como se verá adiante.

III.2. CARACTERÍSTICAS DOS ESTUDOS REALIZADOS

Os três estudos foram realizados utilizando dois conjuntos de vazões, cada qual constituído de 100 séries, compostas de 40 vazões obtidas de sorteios independentes da distribuição lognormal com média unitária, tendo sido adotado o desvio padrão de 0,1 no primeiro conjunto e 0,8 para o segundo. Estas vazões na realidade não são valores observados num dado rio, mas sim um conjunto de séries sintéticas construídas com a finalidade de realizar um estudo totalmente controlado, onde uma das fontes de incerteza, a escolha do modelo que descreve o comportamento das vazões, é totalmente anulada. Neste caso toda a incerteza introduzida no estudo é proveniente dos erros amostrais de estimação de parâmetros.

Doravante denominaremos estes conjuntos de Cenário A (desvio padrão 0,1) e Cenário B (desvio padrão 0,8) de forma a facilitar a exposição. A seguinte nomenclatura será adotada:

n	número de vazões de uma série (=40)
m	número de séries do cenário (=100)
i	índice da vazão, $i = 1, \dots, n$
j	índice da série, $j = 1, \dots, m$
x_{ij}	vazão i da série j
x_j	série j de vazões
μ_x	valor esperado da população x
σ_x	desvio padrão da população x
\bar{x}_j	média amostral da série j

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (\text{III.1})$$

$s_j(x)$ desvio padrão da série j

$$s_j(x) = \left(\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{(n-1)} \right)^{1/2} \quad (\text{III.2})$$

A avaliação das metodologias Bayesiana, Jackknife e Bootstrap, baseou-se na comparação das variabilidades, em torno do "valor verdadeiro", das estimativas obtidas por estas metodologias com a mesma variabilidade da estimativa obtida com fórmulas da Estatística Clássica.

III.3. ESTUDOS DE CASOS

III.3.1. CASO 1 - Desvio Padrão do Estimador de Vazão Média

A vazão média é talvez o parâmetro mais utilizado em Hidrologia. Usualmente ignora-se a variabilidade amostral e utiliza-se a média amostral, \bar{x} , (da série de vazões) como se fosse o correspondente valor de população, μ_x .

Apesar de \bar{X} ser um bom estimador (não tendencioso e eficiente) da média de população, não é correto supor que a média amostral é o "próprio" valor de população, uma vez que \bar{X} é função da amostra aleatória, sendo portanto uma variável aleatória. Isto é, diferentes amostras podem conduzir a diferentes médias amostrais. O que se pode provar é que sendo \bar{X} um estimador não tendencioso, o valor esperado de \bar{X} ($E[\bar{X}]$) é igual a média populacional (μ_X). Portanto se fossem disponíveis inúmeras séries de vazões, das quais fossem obtidas as médias amostrais, o valor médio destas médias amostrais poderia ser considerado como μ_X . Como esta situação é impossível, o que devemos fazer é estimar μ_X por \bar{x} e construir um intervalo de confiança para μ_X da forma:

$$\bar{x} - z_{(1-\alpha/2)} S(\bar{x}) \leq \mu_X \leq \bar{x} + z_{(1-\alpha/2)} S(\bar{x}) \quad (\text{III.3})$$

o qual deve conter μ_X com uma probabilidade $(1-\alpha)$. O valor de $z_{(1-\alpha/2)}$ é obtido aproximadamente da distribuição normal padrão se $S(\bar{X})$ é conhecido (devido ao teorema do limite central), caso contrário $z_{(1-\alpha/2)}$ é obtido da distribuição t-student com $n-1$ graus de liberdade.

Neste caso surge a necessidade de estimar $S(\bar{X})$ (desvio padrão de \bar{X}) pois o intervalo de confiança será tanto mais preciso, quanto melhor for o estimador de $S(\bar{X})$.

Apresentaremos quatro formas de estimar o desvio padrão de \bar{X} ($S(\bar{X})$), o Método Clássico, o Bayesiano, o Jackknife e o Bootstrap.

MÉTODO CLÁSSICO

Pela Estatística Clássica temos que se os X_i são variáveis aleatórias independentes e identicamente distribuídas com desvio padrão σ_X , então o desvio padrão de \bar{X}

é:

$$\begin{aligned}
 S(\bar{X}) &= [\text{Var}(\bar{X})]^{1/2} \\
 &= \left[\text{Var}\left(\frac{\sum_{i=1}^n x_i}{n}\right) \right]^{1/2} \\
 &= \left[\frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n x_i\right) \right]^{1/2} \\
 &= \frac{\sigma_x}{\sqrt{n}} \qquad \qquad \qquad (\text{III.4})
 \end{aligned}$$

Se o valor de σ_x é conhecido podemos facilmente obter $S(\bar{X})$. Como o estudo apresentado é baseado em cenário artificiais, então dispomos de σ_x e os verdadeiros valores de $S(\bar{X})$ são:

$$\text{CENÁRIO A: } S(\bar{X}) = \frac{\sigma_x}{\sqrt{n}} = \frac{0,1}{\sqrt{40}} = 0,01581$$

$$\text{CENÁRIO B: } S(\bar{X}) = \frac{\sigma_x}{\sqrt{n}} = \frac{0,8}{\sqrt{40}} = 0,12850$$

Entretanto, σ_x é em geral desconhecido e é estimado pelo correspondente valor amostral ($s(x)$).

No caso presente como cada um dos cenários (A e B) é constituído de 100 séries, para cada série é calculado o estimador clássico do desvio padrão de \bar{X} , $\hat{S}_{\text{Class}}(\bar{x}) = s(x)/\sqrt{n}$, resultando num conjunto de 100 $\hat{S}_{\text{Class}}(\bar{x})$, do qual são obtidas a média e a variância dos $\hat{S}_{\text{Class}}(\bar{x})$.

O algoritmo abaixo esquematiza o procedimento para avaliação de $\hat{S}_{\text{Class}}(\bar{x})$.

1. para cada série j de n vazões calcula-se $s_j(x)$ e

$$\hat{S}_{\text{Class}}(\bar{x}) = \frac{s_j(x)}{\sqrt{n}}$$

2. repete-se o item (1) para todas as m séries.
3. após o item (2) ser concluído, calcula-se a média e a variância de $\hat{S}_{\text{Class}}(\bar{x})$.

A figura III.1 apresenta o esquema para a estimação do desvio padrão de \bar{X} pelo método clássico.

MÉTODO BAYESIANO

Como mencionado no capítulo II, a metodologia Bayesiana considera os parâmetros desconhecidos do modelo probabilístico como variáveis aleatórias, as quais tem sua distribuição de probabilidades construída a partir de informações amostrais e não-amostrais.

Desta forma se o modelo probabilístico que descreve o comportamento das vazões (X) é conhecido, é possível determinar a distribuição de probabilidade de seus parâmetros.

No caso em estudo, como as vazões são variáveis aleatórias independentes, identicamente distribuídas, com distribuição log-normal, uma forma de obter $\hat{S}(\bar{x})$ consiste em considerar os logaritmos das vazões $Y = \log X$, que também são variáveis aleatórias independentes e identicamente distribuídas, sendo modeladas pela distribuição normal. Neste caso μ_y e σ_y , média e desvio padrão dos logaritmos de x , respectivamente, são os parâmetros do modelo normal, que segundo a metodologia Bayesiana possuem distribuição posterior conjunta definida por (II.11). Como μ_x pode ser expressa como uma função de μ_y e σ_y (BENJAMIN et alia [4]), isto é:

$$\mu_x = \exp \left\{ \mu_y + \frac{1}{2} \sigma_y^2 \right\} \quad (\text{III.7a})$$

$$\sigma_y = \left\{ \log \left(\frac{\sigma_x^2}{\mu_x^2} + 1 \right) \right\}^{1/2} \quad (\text{III.7b})$$

é possível determinar a distribuição de μ_x a partir da distribuição conjunta de μ_y , σ_y (II.11) da forma:

$$f_{\mu_x}''(\mu_x) = \int_0^{\infty} \frac{1}{\mu_x} f_{\mu_y, \sigma_y}'' \left(\log \mu_x - \frac{1}{2} \sigma_y^2, \sigma_y \right) d\sigma_y \quad (\text{III.8})$$

A partir de (III.8) podemos obter o valor esperado e o desvio padrão de μ_x :

$$E[\mu_x] = \int_{-\infty}^{\infty} u f''_{\mu_x}(u) du \quad (\text{III.9})$$

$$\sigma(\mu_x) = \left[\int_{-\infty}^{\infty} (u - E(\mu_x))^2 f''_{\mu_x}(u) du \right]^{1/2} \quad (\text{III.10})$$

Outro procedimento seria obter a distribuição de μ_x por Monte Carlo, ou seja, a partir de (II.11) é possível obter por sorteio um grande número de pares (μ_y, σ_y) e através de (III.7a) transformá-los num grande número de sorteios μ_x .

Esta série de μ_x pode ser suficientemente grande para definir a distribuição de μ_x e seus parâmetros $E(\mu_x)$ e $\sigma(\mu_x)$. Para efeito de determinação de intervalos de confiança pode-se considerar $\hat{S}_{\text{Bayes}}(\bar{X}) = \sigma(\mu_x)$.

No caso em estudo optou-se pela determinação do desvio padrão de \bar{X} pela segunda alternativa pela evidente facilidade de sua aplicação bastando para isto algoritmos de geração de variáveis aleatórias qui-quadrado e normais (NAYLOR et alia [25]). Note que a medida que aumenta a complexidade do problema obter expressões para um determinado parâmetro pode se tornar uma tarefa ainda mais complexa do que as integrações em (III.8), (III.9) e (III.10).

O algoritmo abaixo esquematiza a forma de avaliação de $\hat{S}_{\text{Bayes}}(\bar{x})$:

1. obter série de y_{ij} a partir de uma série j de n vazões (x_{ij}) .
 - 1.1. construir f''_{μ_y, σ_y} (equação (II.12) com $n' = 0$).
 - 1.2. obter L duplas (μ_y, σ_y) sorteando σ_y^2 da equação (II.18) e μ_y da normal condicionada que aparece na equação (II.11). Usou-se $L = 1000$.

- 1.3. a partir da série de L valores de μ_y e σ_y obter séries de μ_x . através da equação (III.7).
- 1.4. determinar $\hat{S}_{\text{Bayes}}(\bar{x})$ pelo desvio padrão dos μ_x 's.
2. repetir item (1) para todas as m séries.
3. após o item (2), obter a média e a variância de $\hat{S}_{\text{Bayes}}(\bar{x})$.

A figura III.2 apresenta o esquema desta metodologia.

ESTIMADOR DO DESVIO PADRÃO DE \bar{X} MÉTODO BAYESIANO

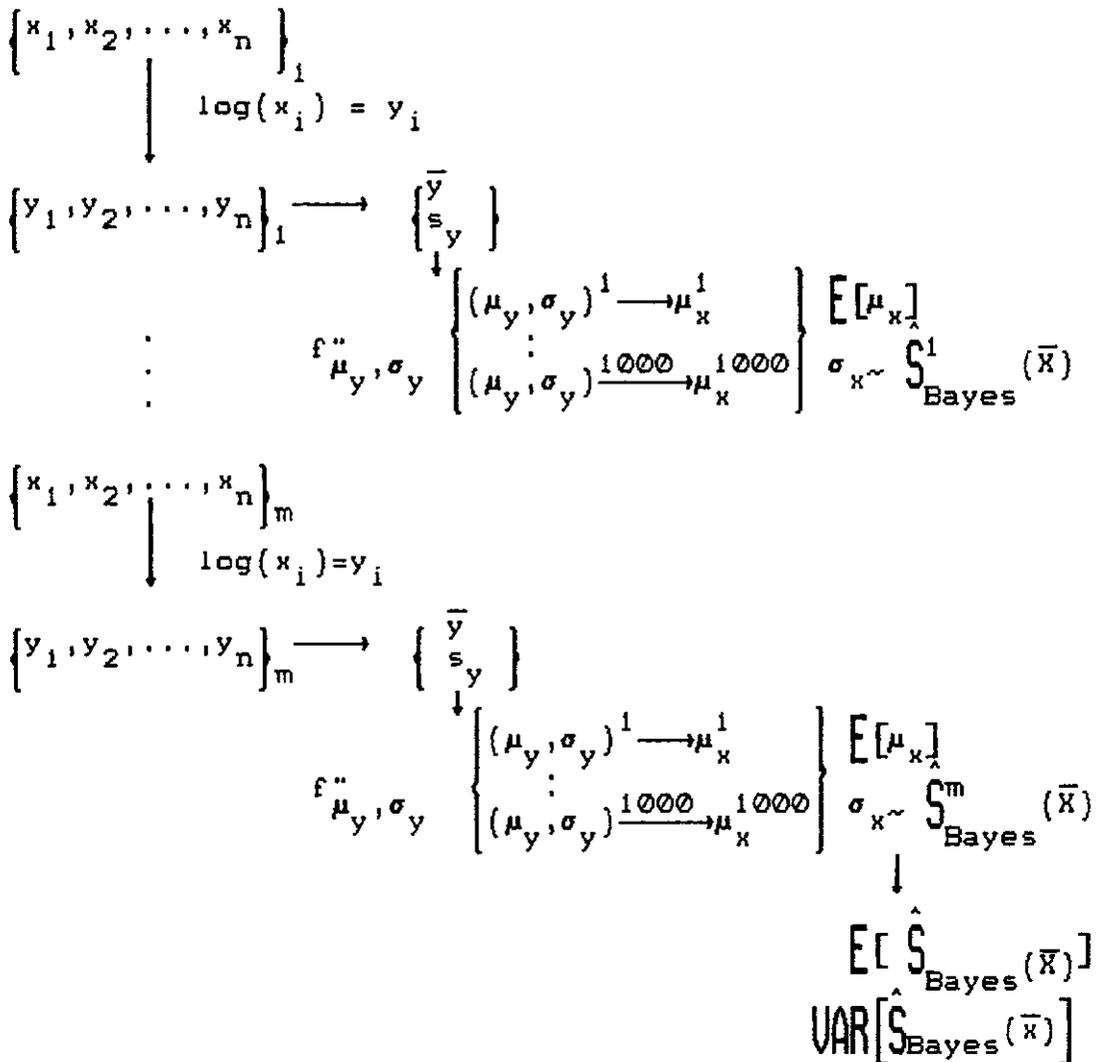


Figura III.2 - Esquema do Método Bayesiano para o Cálculo do Desvio Padrão de \bar{X} .

JACKKNIFE

O cálculo do estimador Jackknife do desvio padrão conforme exposto no item II.3.1 para o caso em que $\hat{\theta} = \bar{X}$ é dado por Efron¹⁹:

$$\hat{S}_{\text{Jackk}}(\bar{X}) = \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 \right\}^{1/2} \quad (\text{III.11})$$

pois,

$$\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{e} \quad \hat{\theta}_{-i} = \bar{x}_{-i}$$

então:

$$\hat{\theta}_{-i} = \frac{n \hat{\theta} - x_i}{n-1} \quad \text{e}$$

$$\hat{\theta}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{-i} = \frac{1}{n} \sum_{i=1}^n \frac{n \hat{\theta} - x_i}{n-1} = \hat{\theta}$$

$$\hat{\theta}_{-i} - \hat{\theta}_{(.)} = \frac{n \hat{\theta} - x_i}{n-1} - \hat{\theta} = \frac{\bar{x} - x_i}{n-1} \quad (\text{III.12})$$

Substituindo a equação (III.12) na equação (II.30) obtemos a equação (III.11). Ou seja, neste caso o Estimador Jackknife do desvio padrão é igual ao Estimador Clássico.

BOOTSTRAP

O cálculo do Estimador Bootstrap do desvio padrão conforme exposto no item (II.3.2), para o caso em que $\hat{\theta} = \bar{X}$ consiste em fazer na equação (II.32),

$$g(x_1^*, x_2^*, \dots, x_n^*) = \frac{1}{n} \sum_{i=1}^n x_i^* \quad (\text{III.13})$$

Consideremos x_i^* como variáveis aleatórias independentes e igualmente distribuídas com distribuição igual a distribuição empírica de x_i , $g(x_1^*, x_2^*, \dots, x_n^*)$ é uma variável aleatória cuja variância é dada por:

$$\text{Var}[g(X_1^*, X_2^*, \dots, X_n^*)] = n \frac{1}{n^2} \text{Var}(X_i^*) \quad (\text{III.14})$$

Se interpretarmos o estimador Bootstrap do desvio padrão como o desvio padrão de $g(x_1^*, x_2^*, \dots, x_n^*)$, tem-se que:

$$\lim_{B \rightarrow \infty} \hat{S}_{\text{Boot}}(\bar{x}) = \left\{ \frac{\text{Var}[X_i^*]}{n} \right\}^{1/2} \quad (\text{III.15})$$

Substituindo-se na equação (III.15) a fórmula do $\text{Var}[X_i^*]$ tem-se:

$$\lim_{B \rightarrow \infty} \hat{S}_{\text{Boot}}(\bar{x}) = \frac{s(x)}{\sqrt{n}} \cdot \sqrt{\frac{(n-1)}{n}} \quad (\text{III.16})$$

Ou seja, neste caso os resultados do método Bootstrap com B muito grande podem ser obtidos aplicando-se o fator $\sqrt{(n-1)/n}$ aos resultados do Método Clássico.

RESULTADOS

As tabelas III.1 e III.2 sumarizam os resultados dos estudos realizados com os cenários A e B respectivamente.

Como pode ser observado, o Método Clássico tende a subestimar o valor verdadeiro. Esta tendência é mais acentuada no Bootstrap, visto que a estimativa do Bootstrap para o desvio padrão de estimadores da média (vazão média) é obtida aplicando-se ao estimador do Método Clássico o fator

de correção $\sqrt{(n-1)/n}$, equação (III.16). Tendência inversa ocorre para a metodologia Bayesiana; por este método há uma tendência a superestimar o valor verdadeiro.

A variabilidade medida pelo coeficiente de variação dos métodos Clássico e Bootstrap são iguais, devido a relação entre eles. O Método Bayesiano apresenta uma variabilidade equivalente no cenário A e levemente superior no cenário B.

Quanto ao erro médio quadratico pode-se observar que o Método Clássico e o Bootstrap apresentam valores bastante próximos e ligeiramente menores que o Bayesiano.

Nas tabelas III.1 e III.2 apresentou-se ainda, na linha correspondente ao Bootstrap valores obtidos usando-se a equação (II.35) com $B=1000$. Pode-se notar que os valores obtidos são bastante próximos aos correspondentes no caso limite de $B \rightarrow \infty$.

MÉTODO	MÉDIA	DESVIO PADRÃO	COEF VARIAÇÃO	EMQ ^{1/2}
Clássico e Jackknife	15,76x10 ⁻³	1,75x10 ⁻³	0,111	1,75x10 ⁻³
Bayesiano	16,34x10 ⁻³	1,85x10 ⁻³	0,113	1,92x10 ⁻³
Bootstrap B → ∞	15,56x10 ⁻³	1,73x10 ⁻³	0,111	1,75x10 ⁻³
Bootstrap B = 1000	15,64x10 ⁻³	1,81x10 ⁻³	0,115	1,82x10 ⁻³
Valor Verdadeiro	15,81x10 ⁻³			

Tabela III.1 - Sumário dos estudos realizados para o estimador do desvio padrão da vazão média com o Cenário A. EMQ^{1/2} = raiz quadrada do erro médio quadrático. Média, desvio padrão, coeficiente de variação e EMQ^{1/2} do estimador de desvio padrão de \bar{X} .

MÉTODO	MÉDIA	DESVIO	COEF	EMQ ^{1/2}
		PADRÃO	VARIAÇÃO	
Clássico e Jackknife	0,1205	0,0325	0,269	0,0330
Bayesiano	0,1428	0,0329	0,231	0,0367
Bootstrap B → ∞	0,1190	0,0321	0,269	0,0330
Bootstrap B = 1000	0,1198	0,0326	0,272	0,0333
Valor Verdadeiro	0,1265			

Tabela III.2 - Sumário dos estudos realizados para o estimador do desvio padrão de Vazão Média com Cenário B. EMQ^{1/2} = raiz quadrada do médio quadrático. Média, desvio padrão coeficiente de variação e EMQ^{1/2} do estimador do desvio padrão de \bar{X} .

III.3.2 CASO 2 - Desvio Padrão do Estimador da Vazão Decamilenar

A ocorrência de cheias tem causado prejuízos em diversas sociedades que vão de perdas das benfeitorias localizadas nas áreas inundadas até perdas de vidas humanas. A previsão de cheias pode sem dúvida reduzir estes prejuízos.

No dimensionamento de obras hidráulicas ou de seus órgãos, isto é, na determinação da forma do vertedor de barragens, da altura de diques de proteção marginal, do diâmetro de coletores, ou de obras de artes em estradas e aeroportos e ainda na delimitação de áreas susceptíveis a inundações, é muitas vezes suficiente, pelo menos em primeira aproximação, a determinação do pico da cheia correspondente a um intervalo de recorrência de T anos, $x(T)$, tal que a probabilidade de ocorrer vazões maiores ou iguais a ela é dada por: $P(X \geq x(T)) = T^{-1}$; em vez da determinação completa da forma da hidrógrafa da cheia.

Para a maioria das estruturas hidráulicas de grande responsabilidade, T é em geral adotado como 10.000 anos e $x(10.000)$ é denominada pico da cheia decamilenar, ou simplesmente vazão decamilenar.

A aproximação usual para o cálculo de $x(T)$ é o ajuste de uma distribuição de probabilidade $F(\cdot)$ ao conjunto de máximos anuais observados $\{x_1, x_2, \dots, x_n\}$ e através desta obter o valor de $x(T)$ por extrapolação.

Contudo o valor obtido desta forma para $x(T)$ é apenas uma estimativa que é afetada pela variabilidade amostral. Como descrito no item III.3.1, uma forma de considerar esta variabilidade amostral na estimativa de $x(T)$ é construir um intervalo de confiança para $x(T)$. Para tal precisamos determinar o desvio padrão de $\hat{x}(T)$. As quatro metodologias: Método Clássico, Bayesiano, Jackknife e

Bootstrap foram utilizadas para estimar o desvio padrão de $\hat{x}(T)$.

MÉTODO CLÁSSICO

A determinação do desvio padrão de um quantil $x(T)$, pelo Método Clássico, quando adota-se a distribuição de probabilidades dos x 's log-normal dois parâmetros e estima-se pelo método dos momentos é dado por KITE [26], como:

$$\sigma(x_T) = \delta \frac{\sigma_x}{\sqrt{n}} \quad (\text{III.17})$$

onde:

n = número de máximos anuais, (x), $n=40$

σ_x = desvio padrão dos máximos anuais

$$\delta = [1 + (Cv^3 + 3Cv)F + (Cv^8 + 6Cv^6 + 15Cv^4 + 16Cv^2 + 2)F^2/4]^{1/2} \quad (\text{III.18})$$

Cv = coeficiente de variação dos máximos anuais,

$$Cv = [\exp\{\sigma_y^2\} - 1]^{1/2} \quad (\text{III.19})$$

σ_y é o desvio padrão dos logaritmos dos máximos anuais,

$$\sigma_y = \left\{ \log\left(\frac{\sigma_x^2}{\mu_x^2} + 1\right) \right\}^{1/2} \quad (\text{III.20})$$

F fator de frequência,

$$F = \frac{\exp\{[\log(1+Cv^2)]^{1/2} z - [\log(1+Cv^2)]/2\} - 1}{Cv} \quad (\text{III.21})$$

$z =$ é o valor da distribuição normal padronizada associada a probabilidade $p=1-T^{-1}$, (T=período de retorno).

Se os valores de σ_x e Cv são conhecidos pode-se facilmente obter $\sigma(x_T)$. A tabela III.3 apresenta os valores dos parâmetros necessários a determinação de $\sigma(x_T)$ e o próprio $\sigma(x_T)$ para os cenários A e B.

CENÁRIO	μ_x	σ_x	Cv	x_T	F	δ	$\sigma(x_T)$
A	1,0	0,1	0,1	1,379	4,420	3,590	0,057
B	1,0	0,8	0,8	10,688	12,110	27,880	3,520

Tabela III.3 - Valores de $\sigma(x_T)$ e dos parâmetros necessários para o cálculo dos mesmos, considerando os cenários A e B.

A obtenção do "valor verdadeiro" de $\sigma(x_T)$ é possível neste caso porque se conhece o valor de população de μ_x e σ_x , média e desvio padrão dos máximos anuais. Geralmente entretanto, estes valores são desconhecidos e são estimados pelos correspondentes valores amostrais \bar{x} e s_x , obtendo assim o Estimador Clássico do desvio padrão de $\hat{x}(T)$, $\hat{S}_{Class}(\hat{x}_T)$.

Este procedimento foi adotado para o cálculo de $\hat{S}_{Class}(\hat{x}_T)$ para as 100 séries de máximos anuais dos cenários A e B.

O algoritmo que esquematiza este procedimento é o mesmo apresentado para o Método Clássico no item III.3.1, apenas substituindo o item 1 pelo exposto a seguir:

1. Para cada série j de máximos anuais

1.1- calcula-se \bar{x}_j , $s_j(x)$ e Cv_j

1.2- calcula-se $F_j = \delta_j$

1.3- calcula-se $S_{\text{Class}}^j(\hat{X}_T) = \delta_j s_j(x) / \sqrt{n}$

A figura III.3 apresenta o esquema de Método Clássico para o cálculo do desvio padrão de \hat{X}_T .

ESTIMADOR DO DESVIO PADRÃO DE \hat{X}_T MÉTODO CLÁSSICO

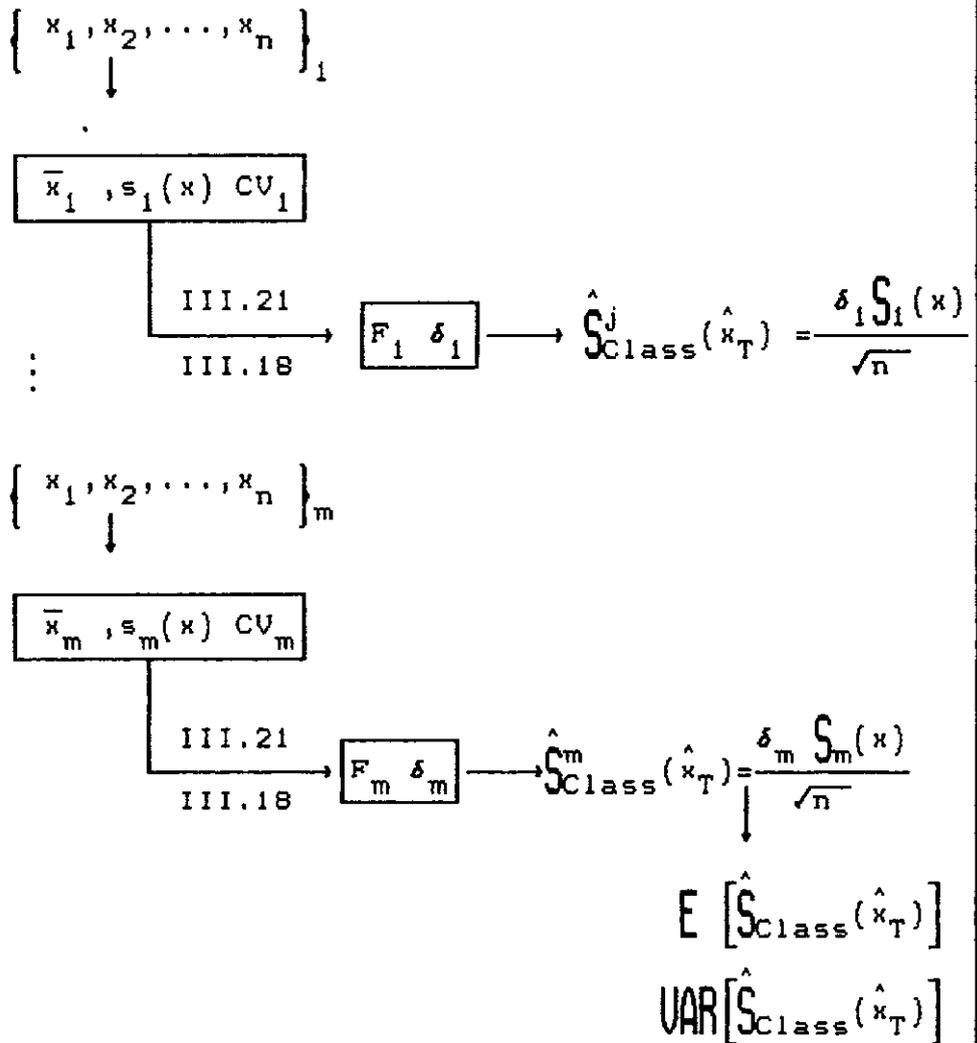


Figura III.3 - Esquema do Método Clássico para Estimação do Desvio Padrão de \hat{X}_T (Estimador da Vazão com T anos de Recorrência).

MÉTODO BAYESIANO

Assim como no item III.3.1 conhecendo-se f_{μ_y, σ_y}'' , distribuição posterior da média e do desvio padrão dos logaritmos dos máximos anuais (x), equação (II.11) é possível determinar a distribuição do quantil X_T . Senão, vejamos:

Seja $Y_T = \log X_T$, isto é, Y_T é o quantil da distribuição normal. Sabe-se que:

$$Y_T = \mu_y + z_T \sigma_y \quad (\text{III.22})$$

onde:

z_T é o valor obtido da distribuição normal padrão relativo a uma probabilidade T^{-1} .

A distribuição de Y_T é calculado por:

$$f_{Y_T}''(y_T) = \int_0^{\infty} f_{\mu_y, \sigma_y}''(y_T - z_T \sigma_y, \sigma_y) d\sigma_y \quad (\text{III.23})$$

E a distribuição de X_T é dada por:

$$f_{X_T}''(X_T) = \frac{1}{X_T} f_{Y_T}''(y_T) \quad (\text{III.24})$$

Com a distribuição de X_T , equação (III.24), pode-se determinar o valor esperado e o desvio padrão de X_T :

$$E[X_T] = \int_{-\infty}^{\infty} x_T f_{X_T}''(x_T) dx_T \quad (\text{III.25})$$

$$\sigma_{X_T} = \left[\int_{-\infty}^{\infty} (x_T - E[X_T])^2 f_{X_T}''(x_T) dx_T \right]^{1/2} \quad (\text{III.26})$$

Para efeito de determinação do intervalo de confiança pode-se considerar $\hat{S}_{\text{Bayes}}(\hat{X}_T) = \sigma_{X_T}$.

Neste caso também é possível determinar a distribuição de X_T por Monte Carlo, o algoritmo exposto no Método Bayesiano do item III.3.1 para avaliação de $\hat{S}_{\text{Bayes}}(\bar{X})$ pode ser adaptado, alterando-se apenas o item 1, como segue:

1- obter série de y_{ij} a partir de uma série de n máximos anuais (x_{ij}) ;

1.1- construir f_{μ_y, σ_y}'' (equação (II.12) com $n'=0$);

1.2- obter L pares (μ_y, σ_y) sorteando σ_y^2 de (II.18) e μ_y da normal condicionada que aparece em (II.11), $L=1000$;

1.3- a partir da série de L valores de μ_y e σ_y obter série de x_T através de $x_T = \exp(\mu_y + z_a \sigma_y)$;

1.4- determinar $\hat{S}_{\text{Bayes}}(\hat{x}_T)$ pelo desvio padrão dos X_T 's.

A figura III.4 apresenta o esquema do Método Bayesiano para a estimação do desvio padrão de \hat{X}_T .

ESTIMADOR DO DESVIO PADRÃO DE \hat{X}_T MÉTODO BAYESIANO

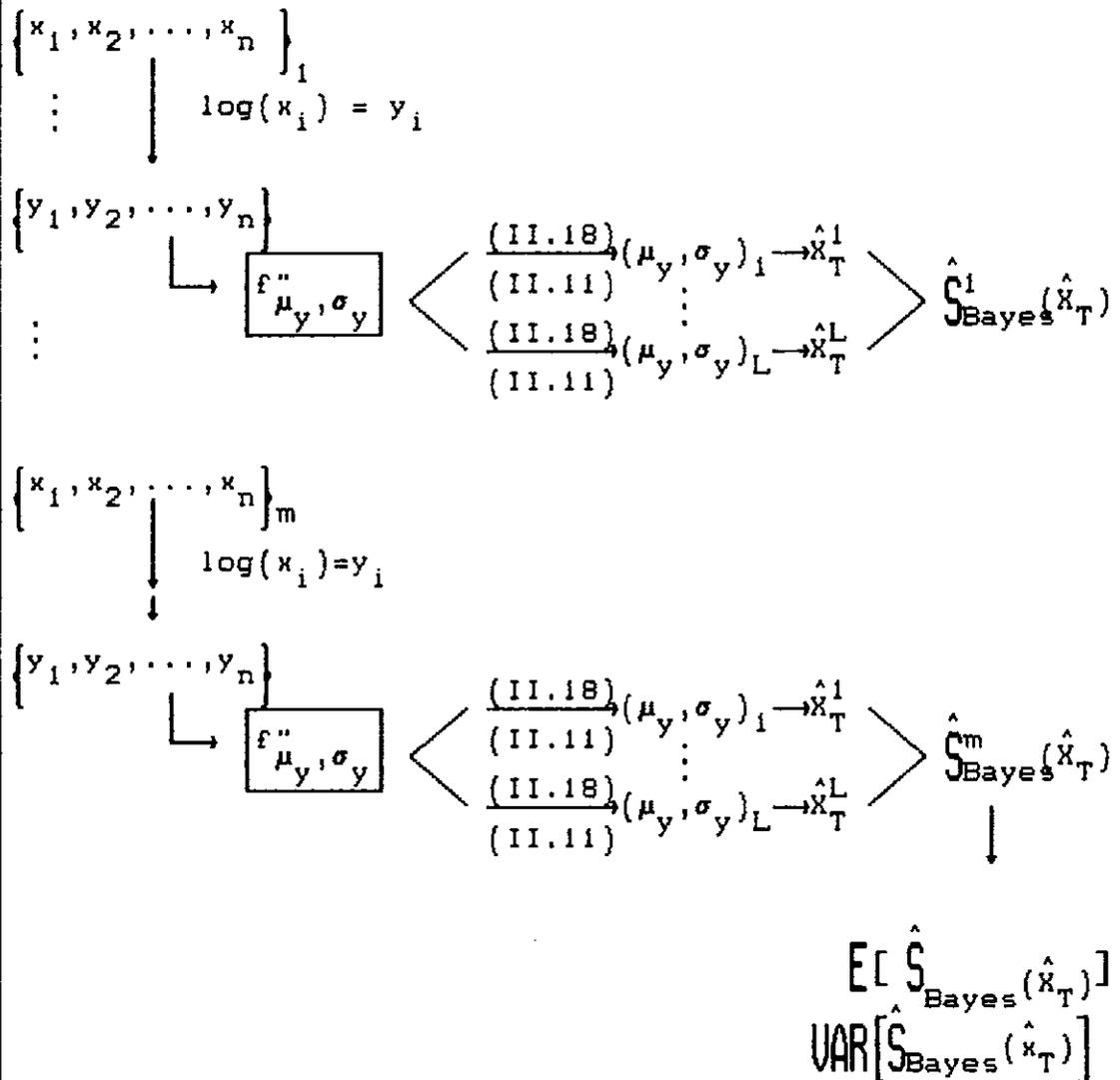


Figura III.4 - Esquema do Método Bayesiano para o Estimação do Desvio Padrão de \hat{X}_T (Estimador da Vazão de T anos de Recorrência)

JACKKNIFE

O cálculo do Estimador Jackknife do desvio padrão do parâmetro $\hat{\theta}$, quando $\hat{\theta} = \hat{X}_T$, $\hat{S}_{Jackk}(\hat{X}_T)$ não é um caso particular como $\hat{\theta} = \bar{X}$, item III.3.1, em que $\hat{S}_{Jackk}(\bar{X}) = \hat{S}_{Class}(\bar{X})$. Neste caso para determinar $\hat{S}_{Jackk}(\hat{X}_T)$ deve-se recorrer ao procedimento descrito no item II.3.1 que pode ser esquematizado pelo seguinte algoritmo:

- 1- obter série de y_{ij} a partir de uma série de n máximos anuais (x_{ij}) ;
 - 1.1- a partir da série de y_{ij} obter uma pseudo-amostra y_{Jackk}^i , omitindo-se o i -ésimo ponto da série de y_{ij} ;
 - 1.2- determina-se \bar{y}^i e s_y^i , média e desvio padrão da pseudo-amostra y_{Jackk}^i ;
 - 1.3- calcula-se $\hat{x}_T^i = \exp(\bar{y}^i + z_\alpha s_y^i)$;
 - 1.4- repete-se os itens (1.1) a (1.3) n vezes;
 - 1.5- determina-se $\hat{S}_{Jackk}^j(\hat{X}_T)$, equação (II.30), a partir dos n \hat{X}_T^i obtidos;
- 2- repete-se o item (1) para as m séries do cenário;
- 3- determina-se a média e a variância dos $\hat{S}_{Jackk}^j(\hat{X}_T)$.

A figura III.5 apresenta o esquema do Jackknife para a estimação do desvio padrão de \hat{X}_T .

ESTIMADOR DO DESVIO PADRÃO DE \hat{X}_T JACKKNIFE

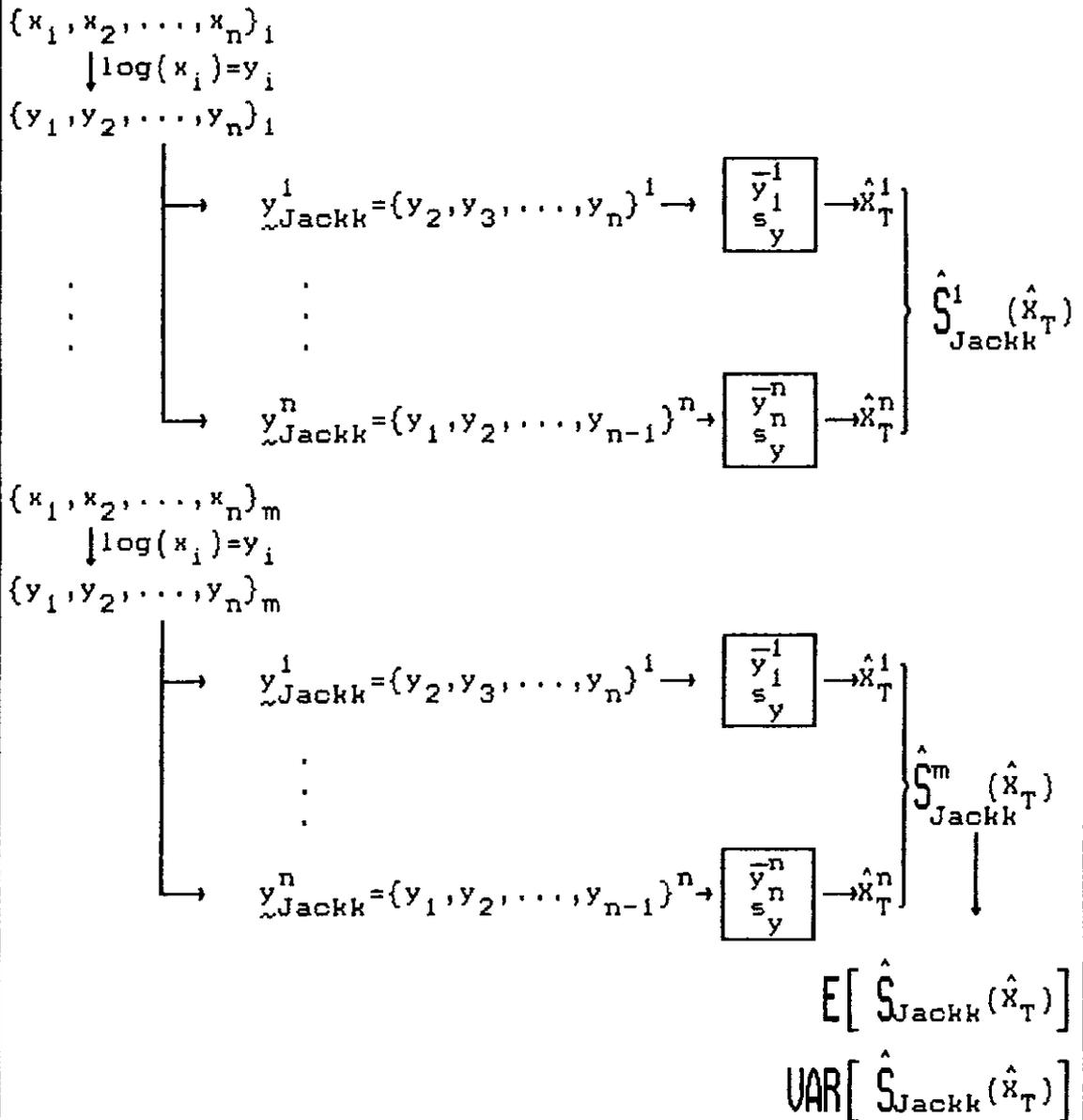


Figura III.5 - Esquema do Jackknife para a Estimação do Desvio Padrão de \hat{X}_T (Estimador da Vazão de T Anos de Recorrência).

BOOTSTRAP

O procedimento adotado para avaliação de $\hat{S}_{\text{Boot}}(\hat{X}_T)$ segue o algoritmo de Monte Carlo descrito no item II.3.2 com algumas alterações que o tornam específico para o caso de $\hat{\theta} = \hat{X}_T$. O algoritmo a seguir descreve este procedimento.

- 1- obter série de y_{ij} a partir de série de n máximos anuais (x_{ij});
 - 1.1- a partir da série de y_{ij} obter uma pseudo-amostra Bootstrap $y_{\sim\text{Boot}}^b$;
 - 1.2- determina-se \bar{y}^b e s_y^b , média e desvio padrão de pseudo-amostra $y_{\sim\text{Boot}}^b$;
 - 1.3- calcula-se $\hat{x}_T^b = \exp(\bar{y}^b + z_\alpha s_y^b)$;
 - 1.4- repete-se os itens (1.1) à (1.3) B vezes ($B=1000$);
 - 1.5- determina-se $\hat{S}_{\text{Boot}}(\hat{X}_T)$ equação (II.35) a partir dos B \hat{x}_T^b ;
- 2- repete-se o item (1) para as m séries do cenário;
- 3- determina-se a média e a variância dos $\hat{S}_{\text{Boot}}(\hat{X}_T)$.

A figura III.6 apresenta o esquema para estimação do desvio padrão de \hat{X}_T pelo Bootstrap.

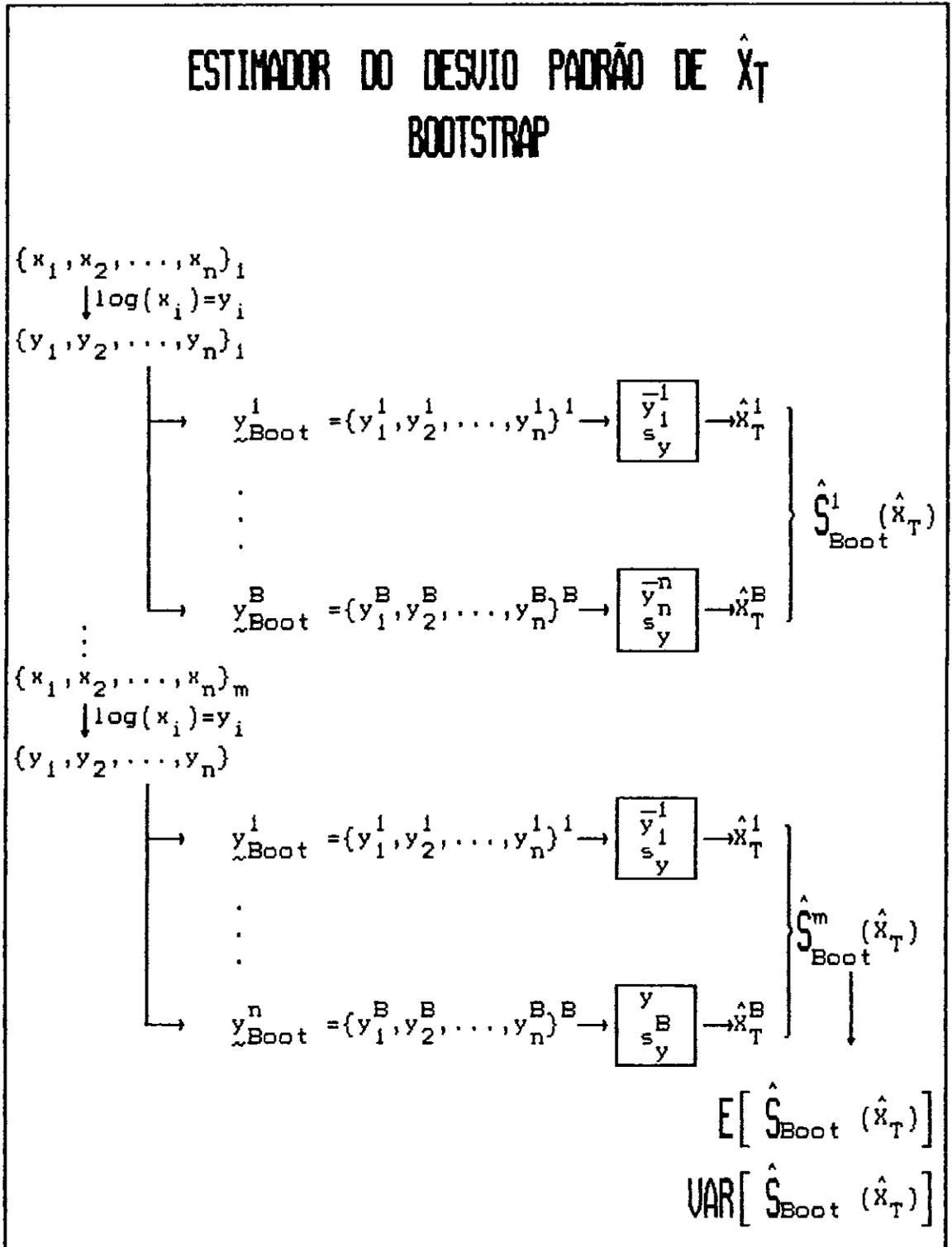


Figura III.6 - Esquema do Bootstrap para a Estimação do Desvio Padrão de \hat{X}_T (Estimador da Vazão de T Anos de Recorrência).

RESULTADO

As tabelas III.4 e III.5 sumarizam os resultados dos estudos realizados no CASO 2 com os cenários A e B respectivamente.

Analizando as tabelas (III.4) e (III.5), que sumarizam os resultados do estudo realizados para vazão decamilenar, pode-se observar que o Método Clássico apresentou uma forte tendência a subestimação em relação ao "valor verdadeiro", enquanto o Jackknife apresentou comportamento distinto para os dois cenários considerados: para o cenário A a tendência é de superestimação em relação ao valor verdadeiro e no cenário B a tendência é de subestimação, apesar disto em ambos os cenários os valores obtidos foram bastante razoáveis quando comparados com o "valor verdadeiro". No Bootstrap ambos os cenários apresentaram valores muito próximos dos "valores verdadeiros", com leve tendência a superestimação. A Metodologia Bayesiana como no caso da Vazão Média apresentou considerável tendência a superestimação em relação ao "valor verdadeiro".

Quanto a variabilidade, os métodos Bootstrap e Jackknife apresentaram valores similares em ambos os cenários, enquanto que o Método Clássico para o cenário A apresentou a menor variabilidade entre as metodologias e para o cenário B a maior variabilidade. A metodologia Bayesiana foi a que apresentou a menor variabilidade no cenário A e variabilidade muito próxima as Técnicas de Reamostragem no cenário B.

Quanto ao erro médio quadrado as Técnicas de Reamostragem (Bootstrap e Jackknife) apresentaram os menores valores.

MÉTODO	MÉDIA	DESVIO PADRÃO	COEF VARIAÇÃO	EMQ ^{1/2}
Clássico	0,036	$4,96 \times 10^{-3}$	0,014	$2,16 \times 10^{-2}$
Bayesiano	0,072	$1,08 \times 10^{-2}$	0,015	$1,85 \times 10^{-2}$
Jackknife	0,065	$1,40 \times 10^{-2}$	0,022	$1,61 \times 10^{-2}$
Bootstrap	0,062	$1,33 \times 10^{-2}$	0,021	$1,42 \times 10^{-2}$
Valor Verdadeiro	0,057			

Tabela III.4 - Sumário dos estudos realizados para o estimador do desvio padrão da vazão decamilenar com cenário A. EMQ^{1/2}=raiz quadrada do erro médio quadrado. Média, desvio padrão, coeficiente de variação e EMQ^{1/2} do estimador do desvio padrão de \hat{X}_T .

MÉTODO	MÉDIA	DESVIO	COEF	EMQ ^{1/2}
		PADRÃO	VARIAÇÃO	
Clássico	2,33	1,22	0,52	2,64
Bayesiano	5,29	2,32	0,44	2,92
Jackknife	3,42	1,46	0,43	1,46
Bootstrap	3,54	1,63	0,46	1,63
Valor Verdadeiro	3,52			

Tabela III.5 - Sumário dos estudos realizados para o estimador do desvio padrão da vazão decamilenar com cenário B, $EMQ^{1/2}$ = raiz quadrada do erro médio quadrado. Média, desvio padrão, coeficiente de variação e $EMQ^{1/2}$ do estimador do desvio padrão de \hat{X}_T .

III.3.2 CASO 3 - Desvio Padrão do Estimador do Tamanho do Reservatório para um Nível de Regularização de 90%

Não há dúvida que a situação ideal para a utilização dos cursos d'água seria que os mesmos mantivessem uma vazão natural constante, não sendo influenciados pelos períodos secos ou chuvosos. Este fato, obviamente não ocorre. A criação de reservatórios nos cursos d'água, permite a regularização das vazões.

Quando o problema a contornar são as cheias projeta-se um volume para o reservatório, denominado "volume de espera", com o objetivo de encaixar o excesso d'água provocado pelas cheias, garantindo-se uma defluência menor que um dado limite, chamada "vazão de restrição" (KELMAN [27], STEDINGER et alia [28]). Quando o problema é a estiagem projeta-se um volume para o reservatório, denominado "volume útil", o qual deve ter a capacidade de garantir uma defluência maior ou igual a um certo limite, chamada "vazão garantida" (DAMÁZIO et alia [29]). O cálculo do Volume de Espera e do Volume Útil são problemas matematicamente simétricos, distinguindo-se pelo intervalo de discretização do tempo, enquanto a cheia é um fenômeno rápido, a estiagem é um fenômeno lento.

Neste item nos dedicaremos ao problema exclusivamente da estiagem, ou seja, garantir uma vazão defluente durante todo decorrer da vida útil do reservatório.

Até as últimas décadas o dimensionamento de reservatórios utilizava a Abordagem Empírica, que é baseada na aplicação da curva de massa (RIPPL [30], 1887) à sequência de vazões observada. Esta forma de dimensionamento traz inúmeras desvantagens (KELMAN [31]) devido a utilização de uma única sequência de vazões. Deste modo dimensiona-se um reservatório de tamanho tal que garantirá a vazão defluente pré-fixada, se uma sequência de vazões como a observada no passado voltar a ocorrer no futuro, o que é altamente improvável.

Uma alternativa ao método empírico é analisar o dimensionamento de reservatório do ponto de vista probabilístico.

Seja $\tilde{x} = \{x_1, x_2, \dots, x_n\}$ uma série de vazões anuais afluentes a um reservatório. Então, para um reservatório inicialmente cheio, a capacidade requerida (K_n) para defluir $\alpha\%$ da vazão média anual do histórico (\bar{x}), é dada por:

$$K_n = \max\{q_i\} \quad (\text{III.27})$$

$$i=1, \dots, n$$

sendo:

$$q_0 = 0$$

$$q_i = \max\{0; q_{i-1} + \frac{\alpha}{100} \bar{x} - x_i\} \quad i = 1, \dots, n \quad (\text{III.28})$$

Como distintas séries de vazões anuais \tilde{x} estão associadas por (III.27) e (III.28) a distintas capacidades, K_n deve ser encarada como uma variável aleatória a qual se associa uma distribuição de probabilidades. Vamos supor que o projetista interessa-se pelo valor esperado e pelo quantil a 95% desta distribuição. GOMIDE [32], 1975, derivou para o caso de afluências independentes e regularização total a distribuição assintótica de K_n . Apresentou ainda um método aplicável a qualquer distribuição das afluências (supostas independentes) e de regularização total. Em geral o método de Gomide exige a solução numérica de integrais ou o uso de distribuições discretas para aproximar distribuições contínuas. Evidentemente, outra alternativa é o uso do Método de Monte Carlo.

As tabelas III.6 à III.8 apresentam para os níveis de regularização de 70%, 90% e 95% o valor esperado ($m_{k_{40}}$), o desvio padrão ($S_{k_{40}}$), e o quantil de 95% ($k_{40}(95\%)$) da distribuição de k_{40} para afluências log-normais

NÍVEL DE REGULARIZAÇÃO = 70%

CV_x	m_{k40}	S_{k40}	$k_{40}(95\%)$	$S(\hat{m}_{k40})$	$S(\hat{k}_{40}(95\%))$
1.00	1.8152	0.7518	3.2923	0.6696	1.2414
0.95	1.7798	0.7330	3.1858	0.6454	1.1954
0.90	1.7401	0.7111	3.1328	0.6196	1.1452
0.85	1.6956	0.6869	3.0542	0.5917	1.0903
0.80	1.6444	0.6607	2.9377	0.5623	1.0351
0.75	1.5879	0.6311	2.8300	0.5314	0.9766
0.70	1.5243	0.5988	2.7041	0.4988	0.9157
0.65	1.4533	0.5660	2.5848	0.4645	0.8540
0.60	1.3747	0.5316	2.4174	0.4288	0.7875
0.55	1.2898	0.4967	2.2256	0.3917	0.7156
0.50	1.1958	0.4600	2.0477	0.3539	0.6369
0.45	1.0920	0.4223	1.8479	0.3160	0.5519
0.40	0.9794	0.3829	1.6510	0.2787	0.4670
0.35	0.8556	0.3422	1.4683	0.2438	0.3892
0.30	0.7134	0.3104	1.2639	0.2127	0.3230
0.25	0.5399	0.2966	1.0686	0.1835	0.2705
0.20	0.3174	0.2775	0.8145	0.1423	0.2360
0.15	0.0869	0.1744	0.4574	0.0706	0.1877
0.10	0.0200	0.0155	0.0702	0.0132	0.0693
0.05	0.0000	0.0000	0.0000	0.0000	0.0000

Tabela III.6 - Valores da média (m_{k40}), desvio padrão (S_{k40}), quantil à 95% ($k_{40}(95\%)$) da distribuição do tamanho do reservatório (kn) e de $S(\hat{m}_{k40})$ (desvio padrão da média de kn), $S(\hat{k}_{40}(95\%))$ (desvio padrão do quantil à 95% de kn), para $n=40$, nível de regularização 70%, afluência log-normal, independente, em função do coeficiente de variação das afluências. Todos os valores da tabela estão padronizadas pelo desvio padrão das afluências.

NÍVEL DE REGULARIZAÇÃO = 90%

CV_x	m_{k40}	S_{k40}	$k_{40}(95\%)$	$S(\hat{m}_{k40})$	$S(\hat{k}_{40}(95\%))$
1.00	4.1880	1.8197	7.7936	1.2773	2.2878
0.95	4.1773	1.8214	7.7062	1.2467	2.2463
0.90	4.1595	1.8205	7.6849	1.2134	2.2022
0.85	4.1344	1.8161	7.6989	1.1771	2.1545
0.80	4.1011	1.8068	7.6892	1.1379	2.1001
0.75	4.0577	1.7925	7.5944	1.0955	2.0345
0.70	4.0041	1.7712	7.4072	1.0500	1.9606
0.65	3.9376	1.7416	7.2629	1.0013	1.8686
0.60	3.8563	1.7026	7.1205	0.9499	1.7689
0.55	3.7570	1.6533	6.9713	0.8958	1.6633
0.50	3.6367	1.5932	6.7219	0.8395	1.5623
0.45	3.4933	1.5233	6.4914	0.7809	1.4705
0.40	3.3207	1.4378	6.1031	0.7208	1.3740
0.35	3.1142	1.3321	5.6484	0.6594	1.2539
0.30	2.8623	1.2028	5.1308	0.5962	1.1073
0.25	2.5572	1.0483	4.5609	0.5301	0.9696
0.20	2.1780	0.8729	3.8225	0.4578	0.8238
0.15	1.7112	0.6939	2.9951	0.3737	0.6376
0.10	1.1227	0.5051	2.0990	0.2862	0.4682
0.05	0.3046	0.3026	0.8224	0.0969	0.2191

Tabela III.7 - Valores da média (m_{k40}), desvio padrão (S_{k40}), quantil à 95% ($k_{40}(95\%)$) da distribuição do tamanho do reservatório (kn) e de $S(\hat{m}_{k40})$ (desvio padrão da média de kn), $S(\hat{k}_{40}(95\%))$ (desvio padrão do quantil à 95% de kn), para $n=40$, nível de regularização 90%, afluência log-normal, independente, em função do coeficiente de variação das afluências. Todos os valores da tabela estão padronizadas pelo desvio padrão das afluências.

NÍVEL DE REGULARIZAÇÃO = 95%

CV_x	m_{k40}	S_{k40}	$k_{40}(95\%)$	$S(\hat{m}_{k40})$	$S(\hat{k}_{40}(95\%))$
1.00	5.0696	2.1812	9.3444	1.4440	2.5380
0.95	5.0884	2.1985	9.4143	1.4121	2.4992
0.90	5.1030	2.2141	9.4705	1.3776	2.4564
0.85	5.1126	2.2277	9.5178	1.3406	2.4082
0.80	5.1161	2.2388	9.5482	1.3008	2.3523
0.75	5.1117	2.2465	9.4789	1.2582	2.2865
0.70	5.0976	2.2504	9.4792	1.2129	2.2100
0.65	5.0724	2.2497	9.4875	1.1648	2.1236
0.60	5.0347	2.2432	9.4609	1.1141	2.0348
0.55	4.9820	2.2284	9.3701	1.0609	1.9465
0.50	4.9108	2.2040	9.3416	1.0057	1.8662
0.45	4.8164	2.1680	9.1398	0.9493	1.7864
0.40	4.6924	2.1148	8.8923	0.8923	1.6956
0.35	4.5313	2.0381	8.5979	0.8351	1.5923
0.30	4.3161	1.9316	8.1637	0.7768	1.4936
0.25	4.0338	1.7936	7.5984	0.7157	1.4008
0.20	3.6548	1.6017	6.6984	0.6497	1.2497
0.15	3.1359	1.3346	5.7221	0.5738	1.0688
0.10	2.3803	0.9700	4.2293	0.4701	0.8411
0.05	1.2310	0.5634	2.3327	0.2780	0.5089

Tabela III.8 - Valores da média (m_{k40}), desvio padrão (S_{k40}), quantil à 95% ($k_{40}(95\%)$) da distribuição do tamanho do reservatório (kn) e de $S(\hat{m}_{k40})$ (desvio padrão da média de kn), $S(\hat{k}_{40}(95\%))$ (desvio padrão do quantil à 95% de kn), para $n=40$, nível de regularização 95%, afluência log-normal, independente, em função do coeficiente de variação das afluências. Todos os valores da tabela estão padronizadas pelo desvio padrão das afluências.

independentes. As figuras (III.7), (III.8) e (III.9) apresentam curvas de m_{k40} , S_{k40} e $k_{40}(95\%)$ para os níveis de regularização de 70%, 90% e 95%, em função do coeficiente de variação das aflúências.

Através da tabela III.7 é possível determinar, para um determinado nível de regularização tanto m_{k40} quanto $k_{40}(95\%)$ se conhecidos média e desvio padrão das aflúências. Assim para o cenário A e um nível de regularização de 90% tem-se: $m_{k40}=0,11$ e $k_{40}(95\%)=0,21$. Contudo, na prática, o projetista não dispõe dos valores populacionais de média e desvio padrão das aflúências, mas sim de estimativas amostrais. Interessa portanto considerar a variabilidade amostral de \hat{m}_{k40} e de $\hat{k}_{40}(95\%)$. A seguir passamos a investigar a aplicação das quatro metodologias, utilizadas nos itens III.3.1 e III.3.2, na determinação do desvio padrão de \hat{m}_{k40} e de $\hat{k}_{40}(95\%)$.

MÉTODO CLÁSSICO

A Análise Estocástica de Reservatórios (Gomide³²) ainda não se desenvolveu o bastante para fornecer fórmulas de desvio padrão de estimadores. No entanto, é fácil obter pelo Método de Monte Carlo este desvio padrão para casos específicos. As duas últimas colunas das tabelas (III.6), (III.7) e (III.8) fornecem o desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ para aflúências log-normais independentes e níveis de regularização de 70%, 90% e 95%. As figuras (III.10) e (III.11) apresentam as respectivas curvas, e as figuras (III.12), (III.13) e (III.14) que correspondem aos níveis de regularização 70%, 90% e 95% respectivamente apresentam as curvas de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ com seus intervalos de um desvio padrão. Se os valores de μ_x e σ_x são conhecidos pode-se obter facilmente $S(\hat{m}_{k40})$ e $S(\hat{k}_{40}(95\%))$. Para o cenário A obtem-se $S(\hat{m}_{k40})=0,0286$ e $S(\hat{k}_{40}(95\%))=0,0468$, e para o cenário B $S(\hat{m}_{k40})=0,9100$ e $S(\hat{k}_{40}(95\%))=1,6800$. Novamente, na prática o projetista obterá uma estimativa para $S(\hat{m}_{k40})$ e $S(\hat{k}_{40}(95\%))$

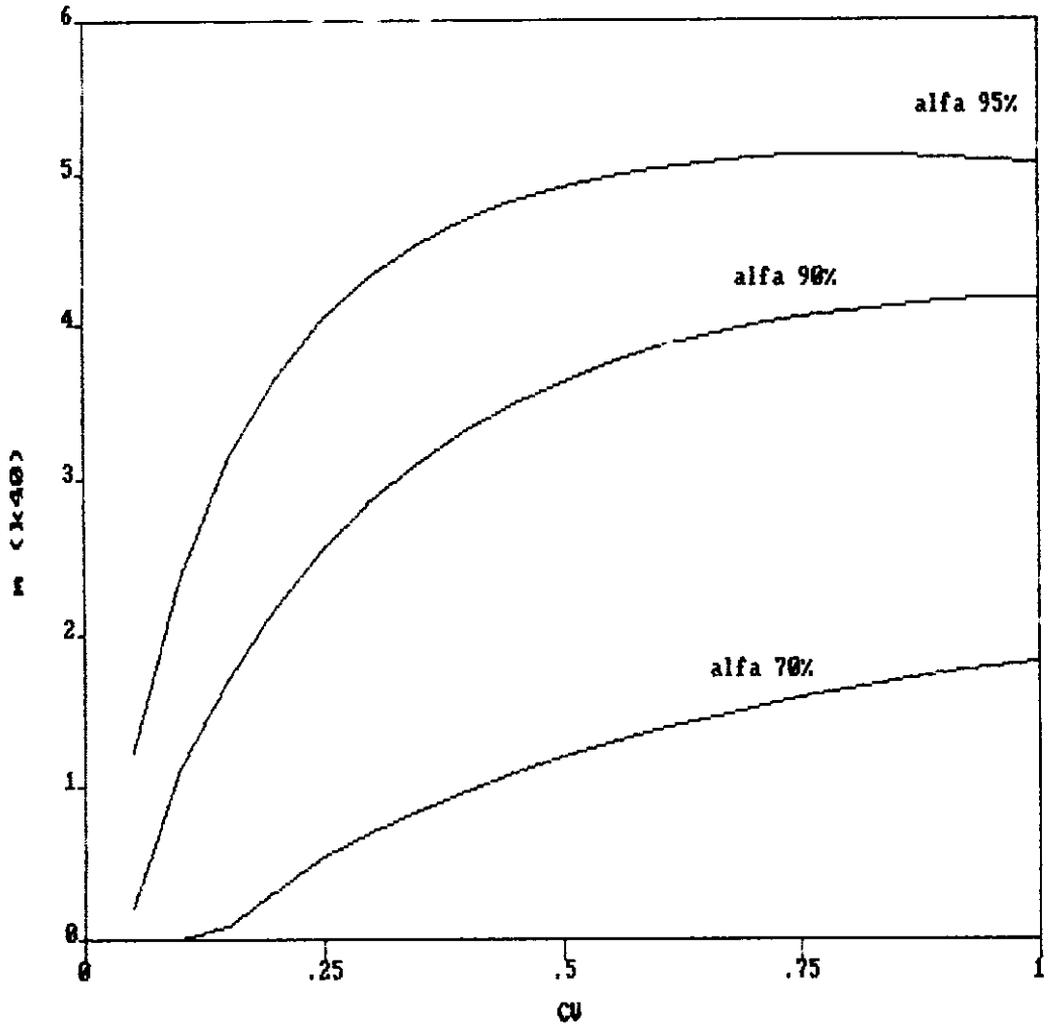


Figura III.7 - Curvas do valor esperado do tamanho do reservatório em função do coeficiente de variação das afluências, ($m_{k40} \times CV$), considerando registro de 40 afluências independentes, lognormais e níveis de regularização de 70%, 90% e 95%. Os valores de m_{k40} estão em unidades de desvio padrão das afluências.

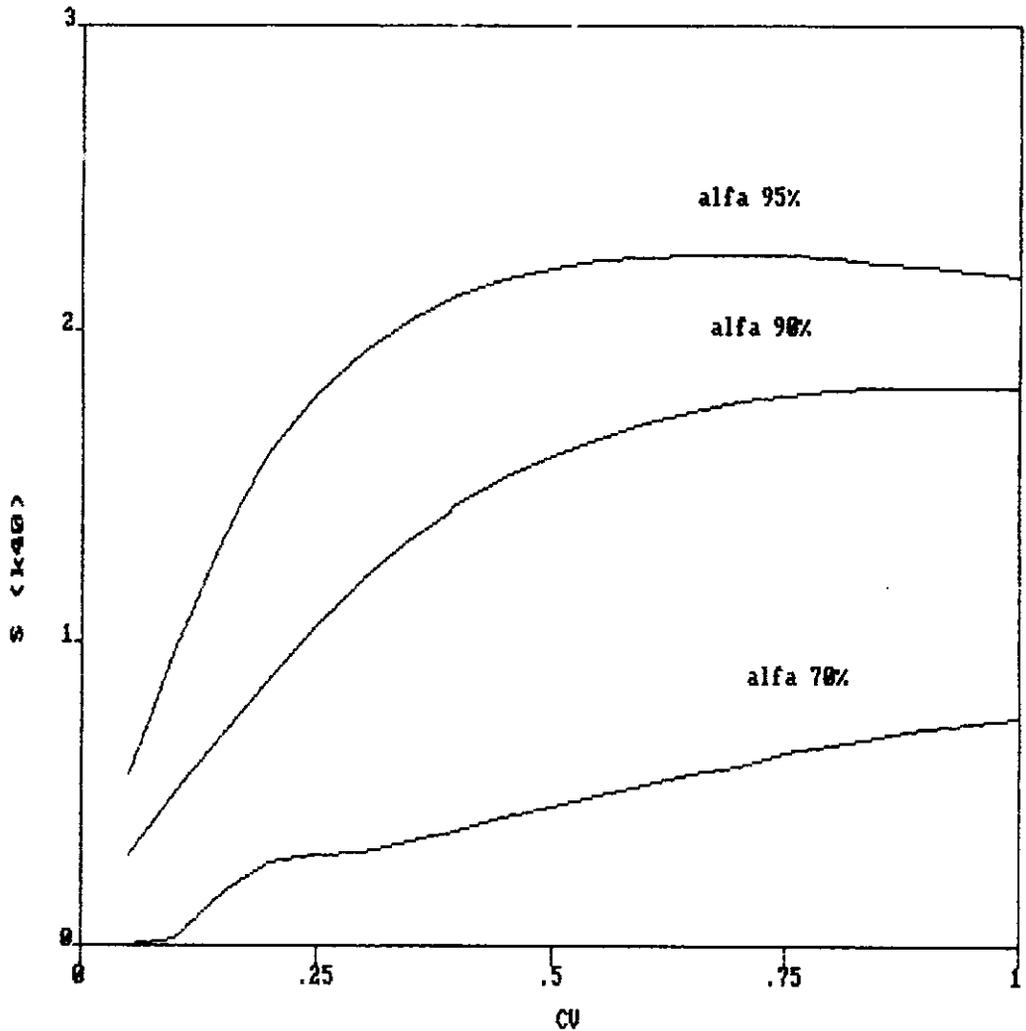


Figura III.8 - Curvas do desvio padrão do tamanho do reservatório em função do coeficiente de variação das afluências, $(S_{k40} \times CV)$, considerando registro de 40 afluências independente log-normal e níveis de regularização de 70%, 90% e 95%. Os valores de S_{k40} estão em unidades de desvio padrão das afluências.

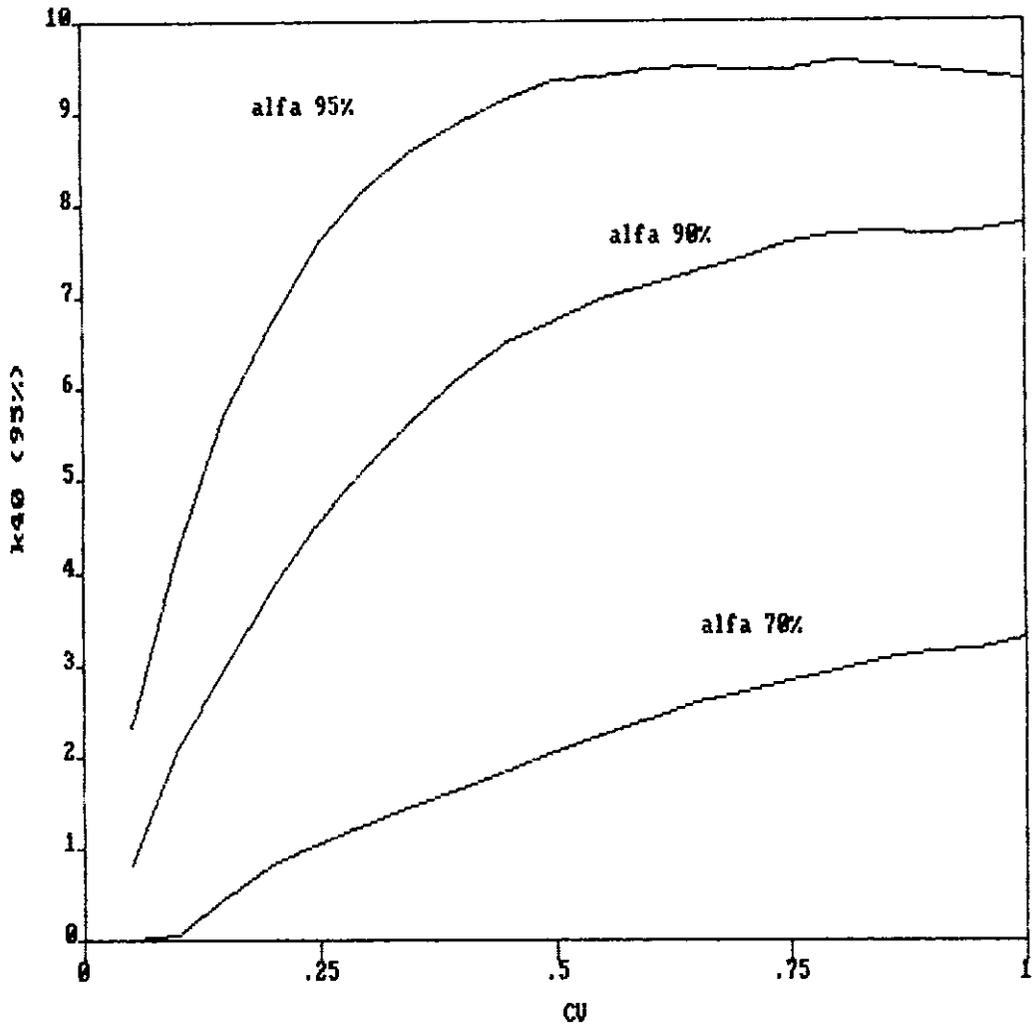


Figura III.9 - Curvas do quantil à 95% da distribuição do tamanho do reservatório em função do coeficiente de variação das afluências, $(k_{40}(95\%) \times CV)$, considerando registro de 40 afluências independentes, lognormais e os níveis de regularização de 70%, 90% e 95%. Os valores de $k_{40}(95\%)$ estão em unidades de desvio padrão das afluências.

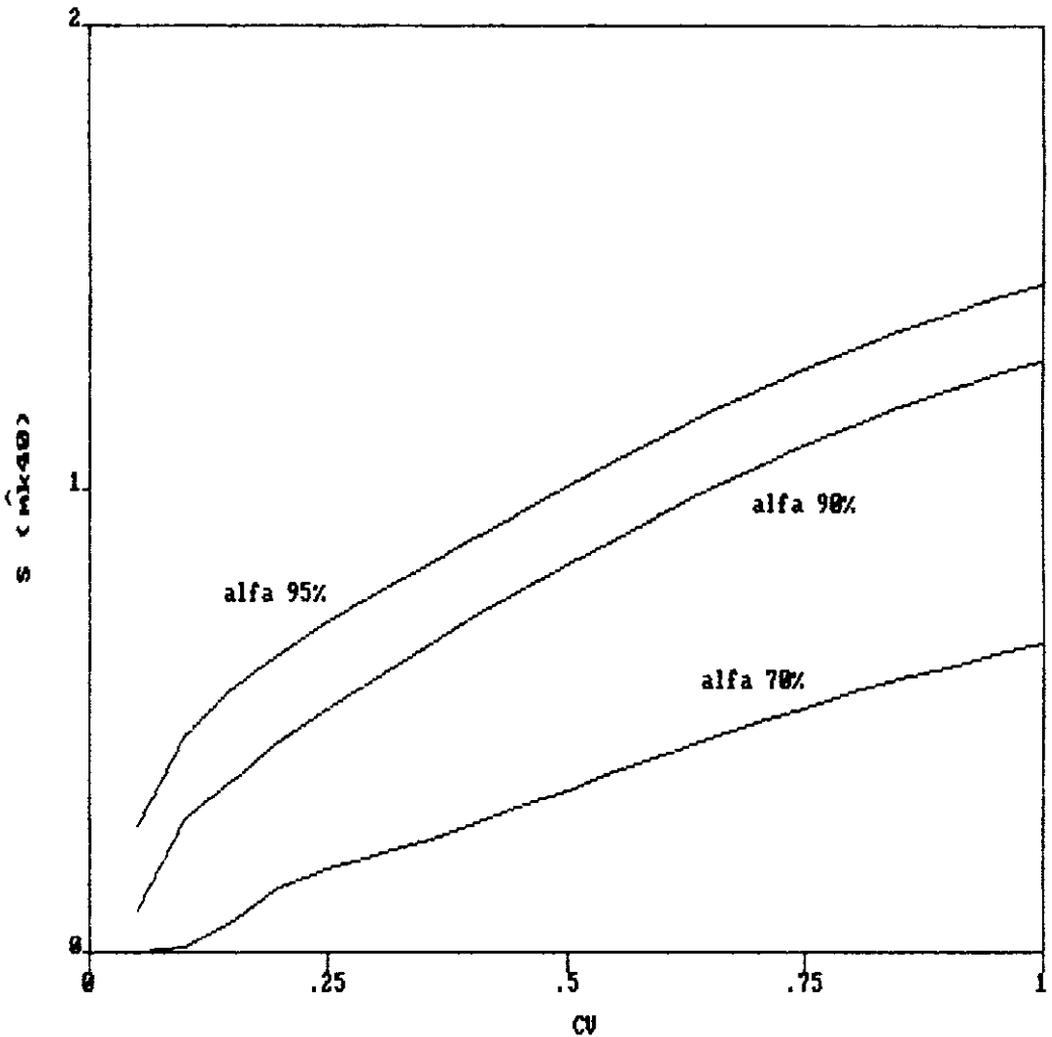


Figura III.10 - Curvas do desvio padrão do valor esperado do tamanho do reservatório em função do coeficiente de variação das afluições, $(S(\hat{m}_{k40}) \times CV)$, considerando o registro de 40 afluições independentes, lognormas e níveis de regularização de 70%, 90% e 95%. Os valores de $S(\hat{m}_{k40})$ estão em unidades de desvio padrão das afluições.

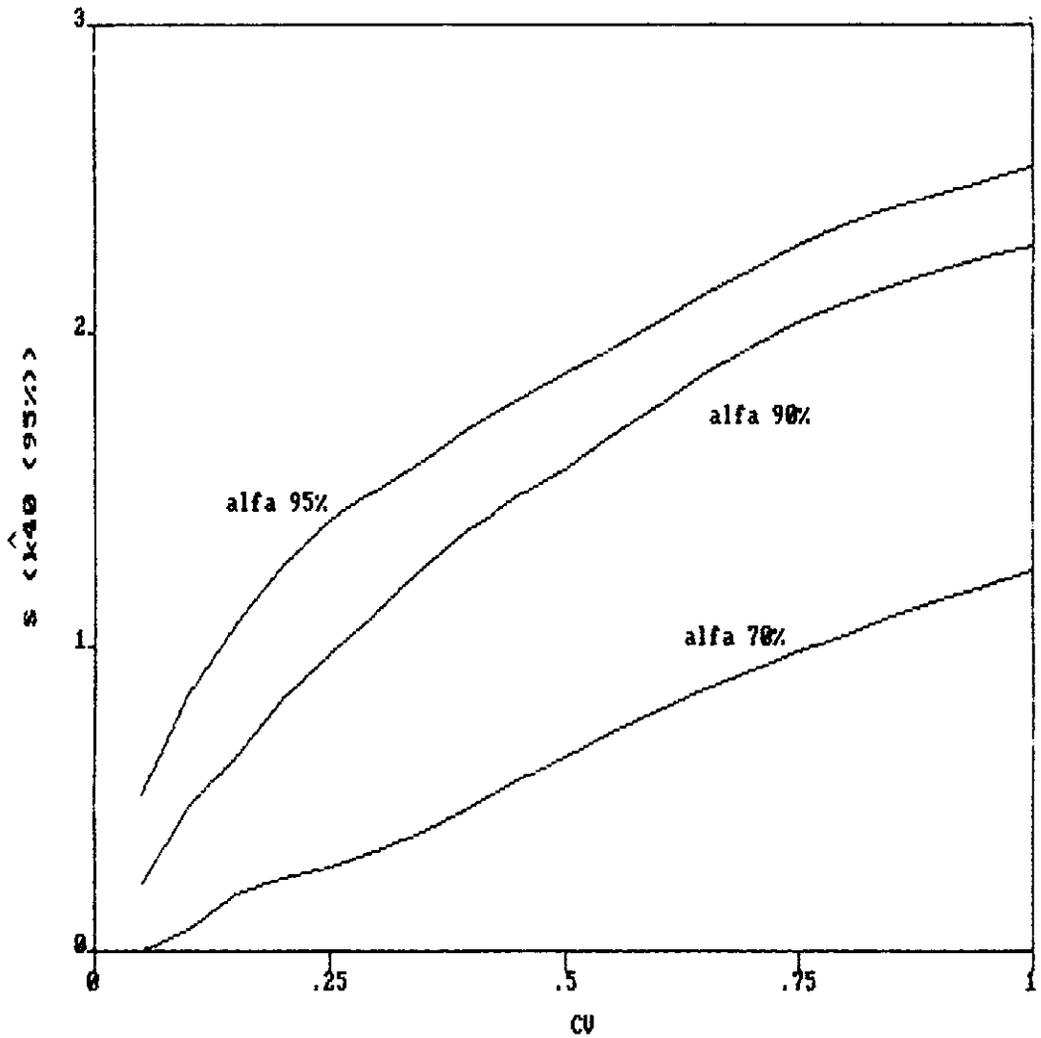


Figura III.11 - Curvas do desvio padrão do quantil 95% do tamanho do reservatório em função do coeficiente de variação das afluições, $(S(\hat{k}_{40}(95\%)) \times CV)$, considerando o registro de 40 afluições independentes, lognormas e níveis de regularização de 70%, 90% e 95%. Os valores de $S(\hat{k}_{40}(95\%))$ estão em unidades de desvio padrão das afluições

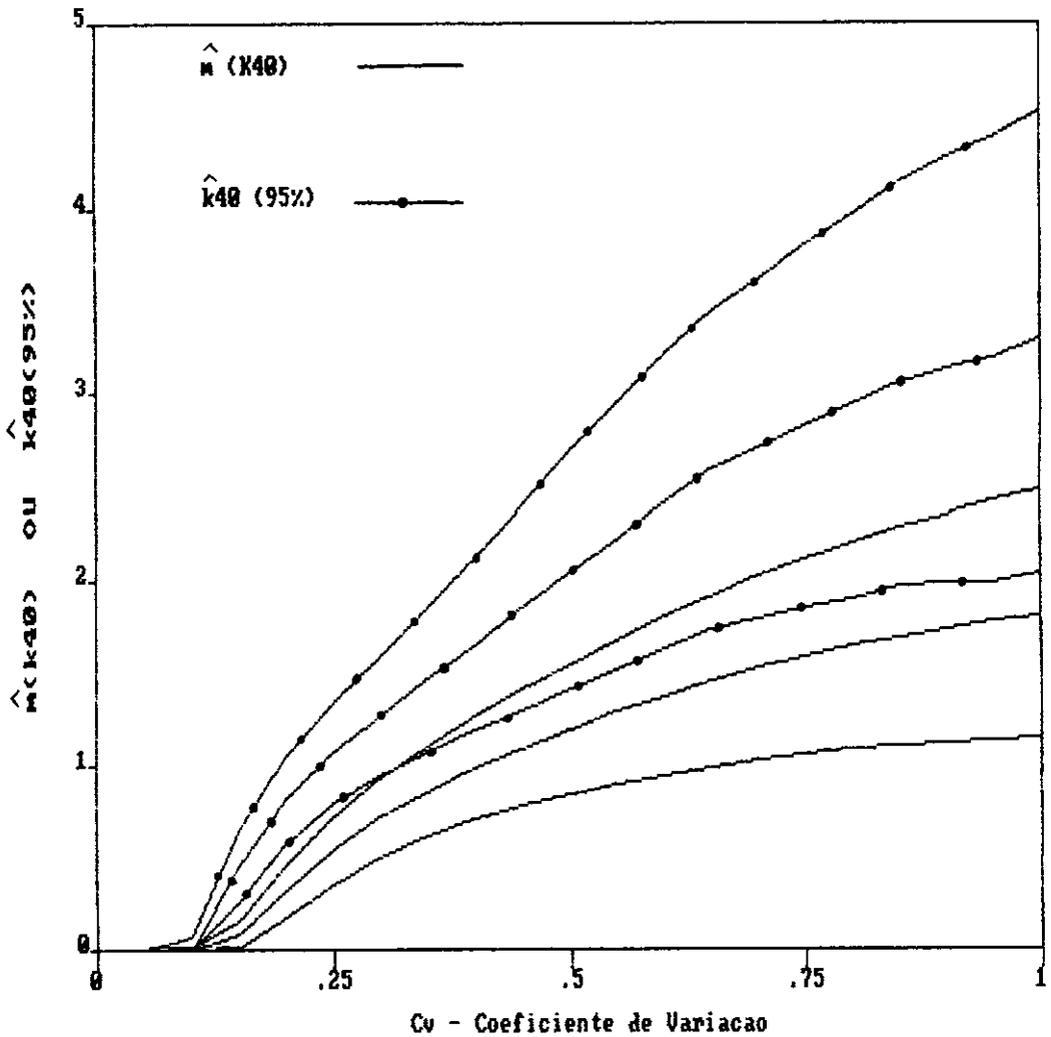


Figura III.12 - Curvas de $\hat{m}_{k=40}$ e $\hat{k}_{40}(95\%)$ com o intervalo de confiança de um desvio padrão em função do coeficiente de variação das afluências, considerando registro de 40 afluências independentes, lognormais e nível de regularização de 70%. Os valores de $\hat{m}_{k=40}$ e $\hat{k}_{40}(95\%)$, estão em unidades de desvio padrão das afluências.

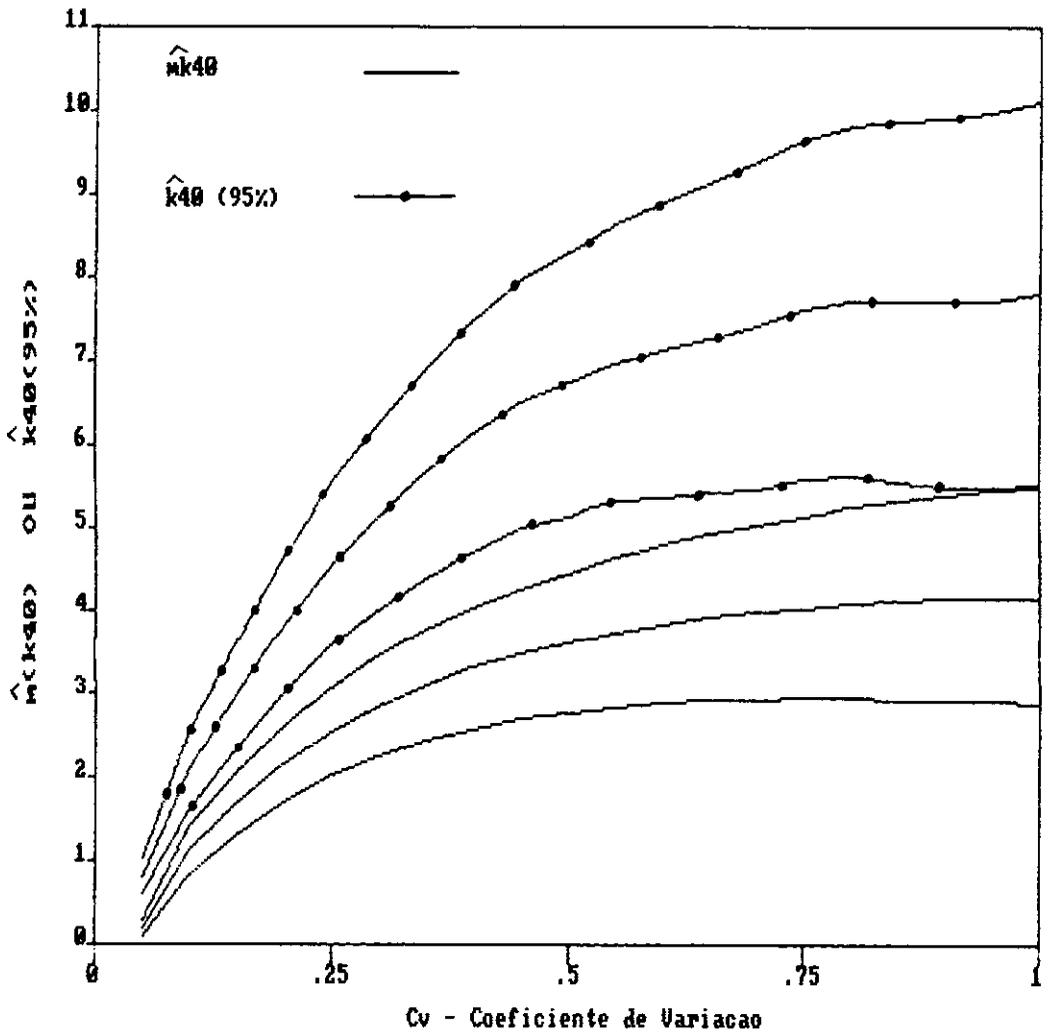


Figura III.13 - Curvas de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ com o intervalo de confiança de um desvio padrão em função do coeficiente de variação das afluências, considerando registro de 40 afluências independentes, lognormais e nível de regularização de 90%. Os valores de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ estão em unidades de desvio padrão das afluências.

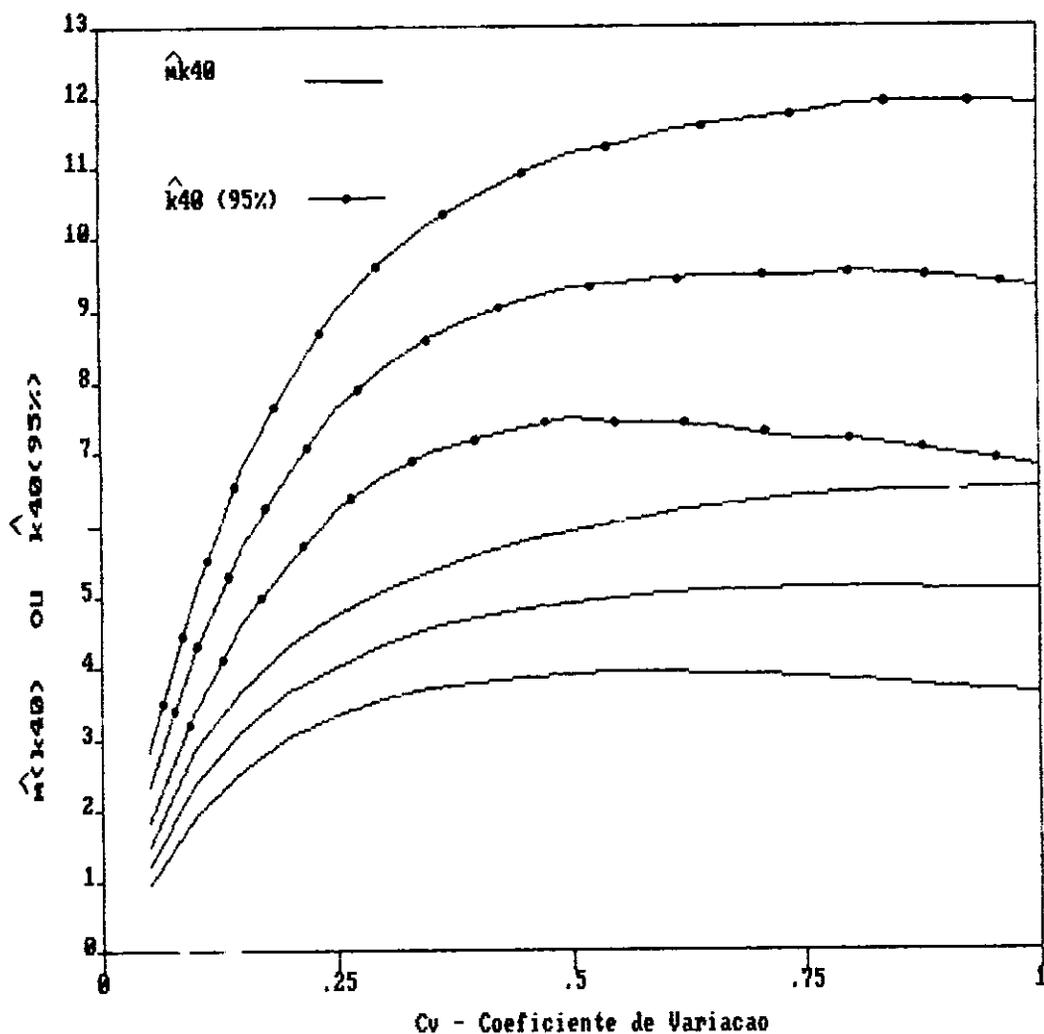


Figura III.14 - Curvas de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ com o intervalo de confiança de um desvio padrão em função do coeficiente de variação das afluências, considerando registro de 40 afluências independentes, lognormais e nível de regularização de 95%. Os valores de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ estão em unidades de desvio padrão das afluências.

usando s_x e \bar{x} . Para cada uma das 100 séries de afluições dos cenários A e B foram obtidos $\hat{S}_{\text{Class}}(\hat{m}_{k40})$ e $\hat{S}_{\text{Class}}(\hat{k}_{40}(95\%))$ segundo uma alteração do item 1 do algoritmo em III.3.1 (Método Clássico).

1- Para cada série j de afluições (x_{ij}) ;

1.1- calcula-se s_{x_j} , \bar{x}_j e CV_{x_j} ;

1.2- obtem-se na tabela III.7 $\hat{S}_{\text{Class}}(\hat{m}_{k40})$ e $\hat{S}_{\text{Class}}(\hat{k}_{40}(95\%))$.

A figura III.15 apresenta o esquema de estimação do desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ pelo método clássico.

ESTIMADOR DO DESVIO PADRÃO DE \hat{m}_{k40} E $\hat{k}_{40}(95\%)$ MÉTODO CLÁSSICO

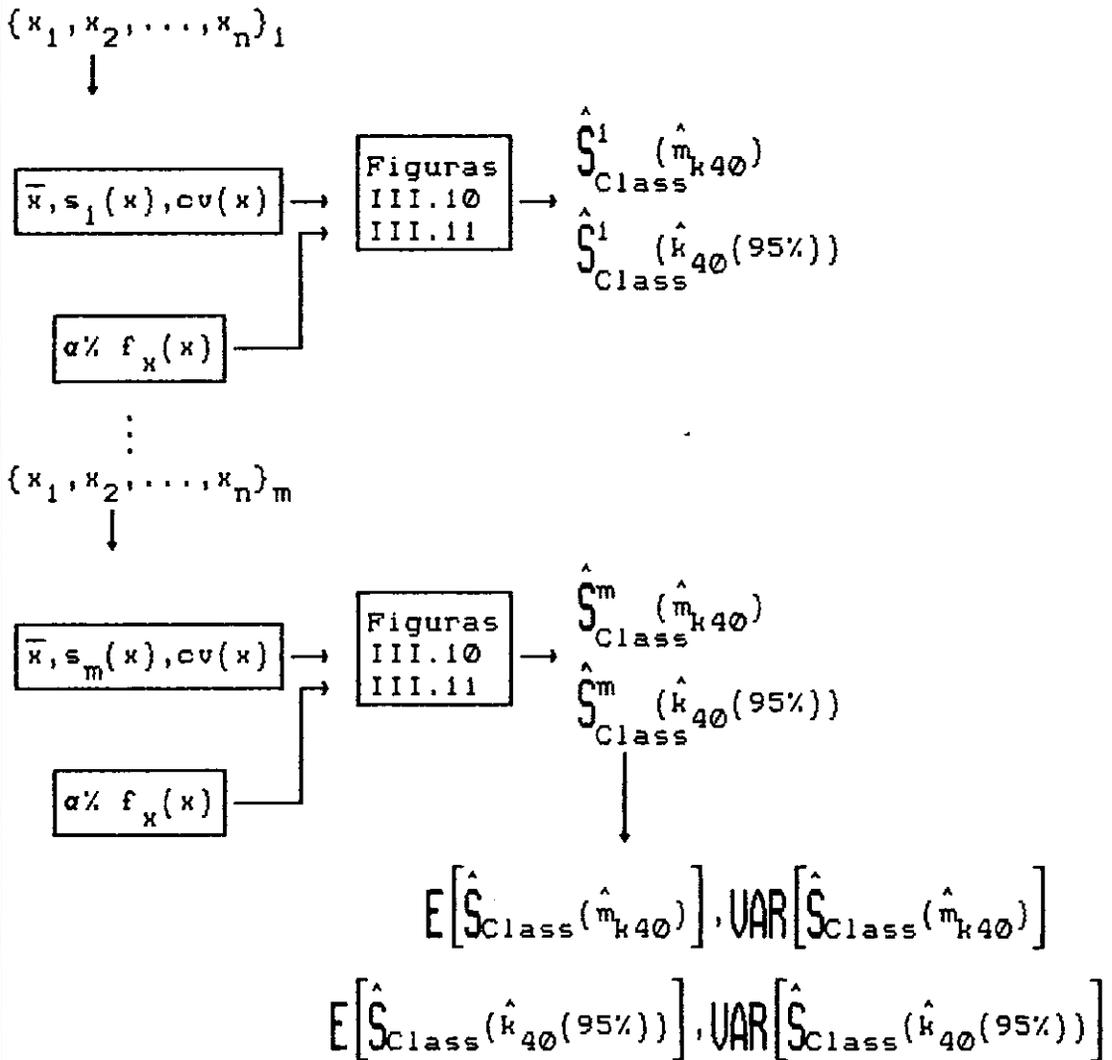


Figura III.15- Esquema para Estimação do Desvio Padrão de \hat{m}_{k40} e de $\hat{k}_{40}(95\%)$ (média e quantil a 95% da Distribuição do Tamanho do Reservatório capaz de Regularizar $\alpha\%$ da Vazão Média) pelo Método Clássico.

MÉTODO BAYESIANO

Neste caso também é possível determinar $\hat{S}_{\text{Bayes}}(\hat{m}_{k40})$ e $\hat{S}_{\text{Bayes}}(\hat{k}_{40}(95\%))$ para cada série j pelo Método Bayesiano utilizando o Método de Monte Carlo. O algoritmo inicialmente considerado (item III.3.1, Método Bayesiano) pode ser adaptado para este caso alterando-se o item 1, como segue:

- 1- A partir de uma série j de alfuências anuais (x_{ij}) obter uma série de y_{ij} , logaritmo dos x_{ij} ;
 - 1.1- construir f''_{μ_y, σ_y} (equação (II.12) com $n'=0$);
 - 1.2- obter L pares (μ_y, σ_y) sorteando σ_y^2 da equação (II.18) e μ_y da normal condicionada que aparece em (II.11); $L=1000$;
 - 1.3- para cada par (μ_y, σ_y) obter por (III.7) e (III.20) μ_x, σ_x e na tabela (III.7) \hat{m}_{k40} e $\hat{k}_{40}(95\%)$;
 - 1.4- a partir da série de L valores de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ determinar $\hat{S}_{\text{Bayes}}(\hat{m}_{k40})$ e $\hat{S}_{\text{Bayes}}(\hat{k}_{40}(95\%))$.

A figura III.16 apresenta o esquema de estimação do desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ pelo método Bayesiano.

ESTIMADOR DO DESVIO PADRÃO DE \hat{m}_{k40} E \hat{k}_{40} (95%) MÉTODO BAYESIANO

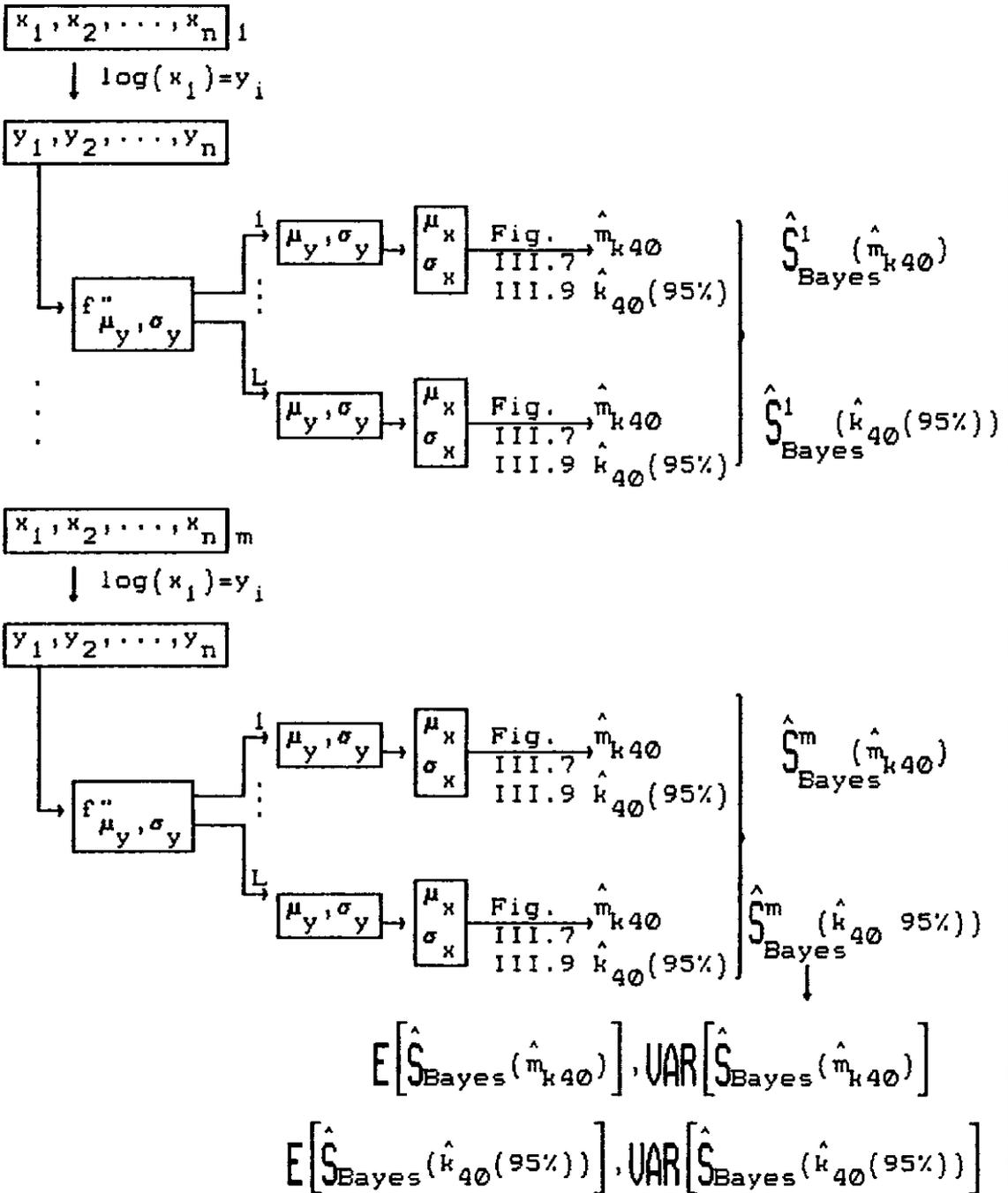


Figura III.16- Esquema para Estimação do Desvio Padrão de \hat{m}_{k40} e de \hat{k}_{40} (95%) (média e quantil a 95% da Distribuição do Tamanho do Reservatório Capaz de Regularizar $\alpha\%$ da Vazão Média) pelo Bayesiano.

MÉTODO JACKKNIFE

O procedimento adotado para a avaliação do desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ segue o algoritmo do item III.3.2 referente a este método, alterando-se apenas o item 1, como exposto a seguir:

- 1- para uma série de afluições anuais x_{ij} ;
 - 1.1- obter uma pseudo-amostra $x_{\sim\text{Jackk}}^i$, omitindo-se o i -ésimo elemento da série de x_{ij} ;
 - 1.2- com a pseudo-amostra obtida calcular \bar{x}^i e $s^i(x)$. Usar a tabela III.7 para obter \hat{m}_{k40} e $\hat{k}_{40}(95\%)$;
 - 1.3- repete-se os itens (1.1) e (1.2) n vezes;
 - 1.4- determina-se $\hat{S}_{\text{Jackk}}^j(\hat{m}_{k40})$ e $\hat{S}_{\text{Jackk}}^j(\hat{k}_{40}(95\%))$ equação (II.30) a partir dos n \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ obtidos.

A figura III.17 apresenta o esquema de estimação do desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ pelo Jackknife.

ESTIMADOR DO DESVIO PADRÃO DE \hat{m}_{k40} E $\hat{k}_{40}(95\%)$ (95%)

METODO JACKKNIFE

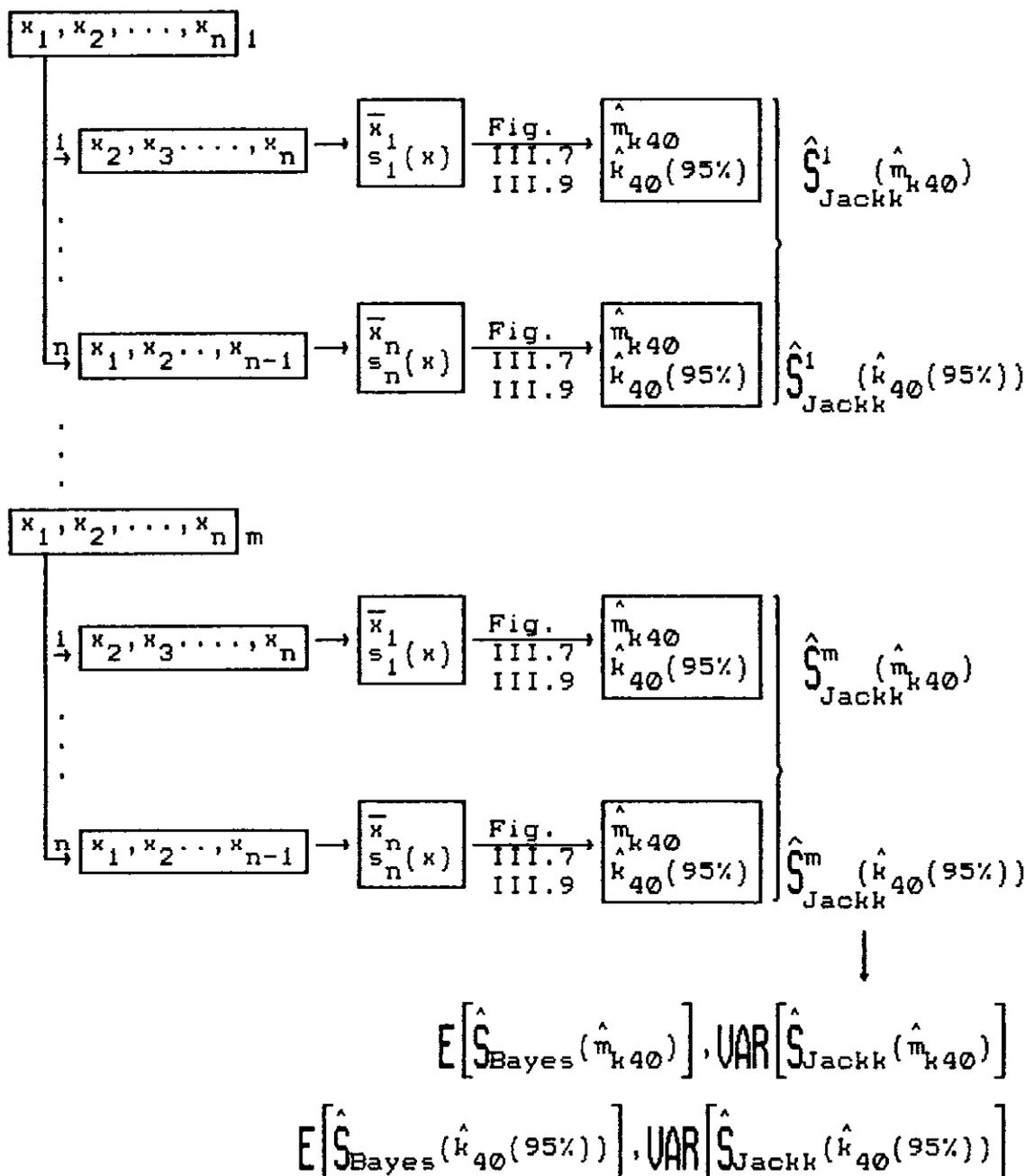


Figura III.17- Esquema para Estimação do Desvio Padrão de \hat{m}_{k40} e de $\hat{k}_{40}(95\%)$ (média e quantil a 95% da Distribuição do Tamanho do Reservatório Capaz de Regularizar $\alpha\%$ da Vazão Média) pelo Jackknife.

MÉTODO BOOTSTRAP

Para a determinação de $\hat{S}_{\text{Boot}}^b(\hat{m}_{k40})$ e $\hat{S}_{\text{Boot}}^b(\hat{k}_{40}(95\%))$, desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ respectivamente, foi adotado o mesmo algoritmo de Monte Carlo do item III.3.2, alterando-se apenas o item 1, que é apresentado a seguir:

- 1- para uma série de afluições anuais x_{ij} ;
 - 1.1- obtem-se uma pseudo-amostra $x_{\sim\text{boot}}^b$;
 - 1.2- com $x_{\sim\text{Boot}}^b$ determina-se \bar{x}^b e s^b_x . Usar a tabela (III.7) para obter \hat{m}_{k40}^b e $\hat{k}_{40}^b(95\%)$;
 - 1.3- repetir os itens (1.1) e (1.2) B vezes ($B=1000$);
 - 1.4- determina-se $\hat{S}_{\text{Boot}}^b(\hat{m}_{k40})$ e $\hat{S}_{\text{Boot}}^b(\hat{k}_{40}(95\%))$ equação (II.35), a partir dos B valores de \hat{m}_{k40}^b e $\hat{k}_{40}^b(95\%)$ obtidos respectivamente.

A figura III.18 apresenta o esquema de estimação do desvio padrão de \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ pelo Bootstrap.

ESTIMADOR DO DESVIO PADRÃO DE \hat{m}_{k40} E \hat{k}_{40} (95%)

METODO BOOTSTRAP

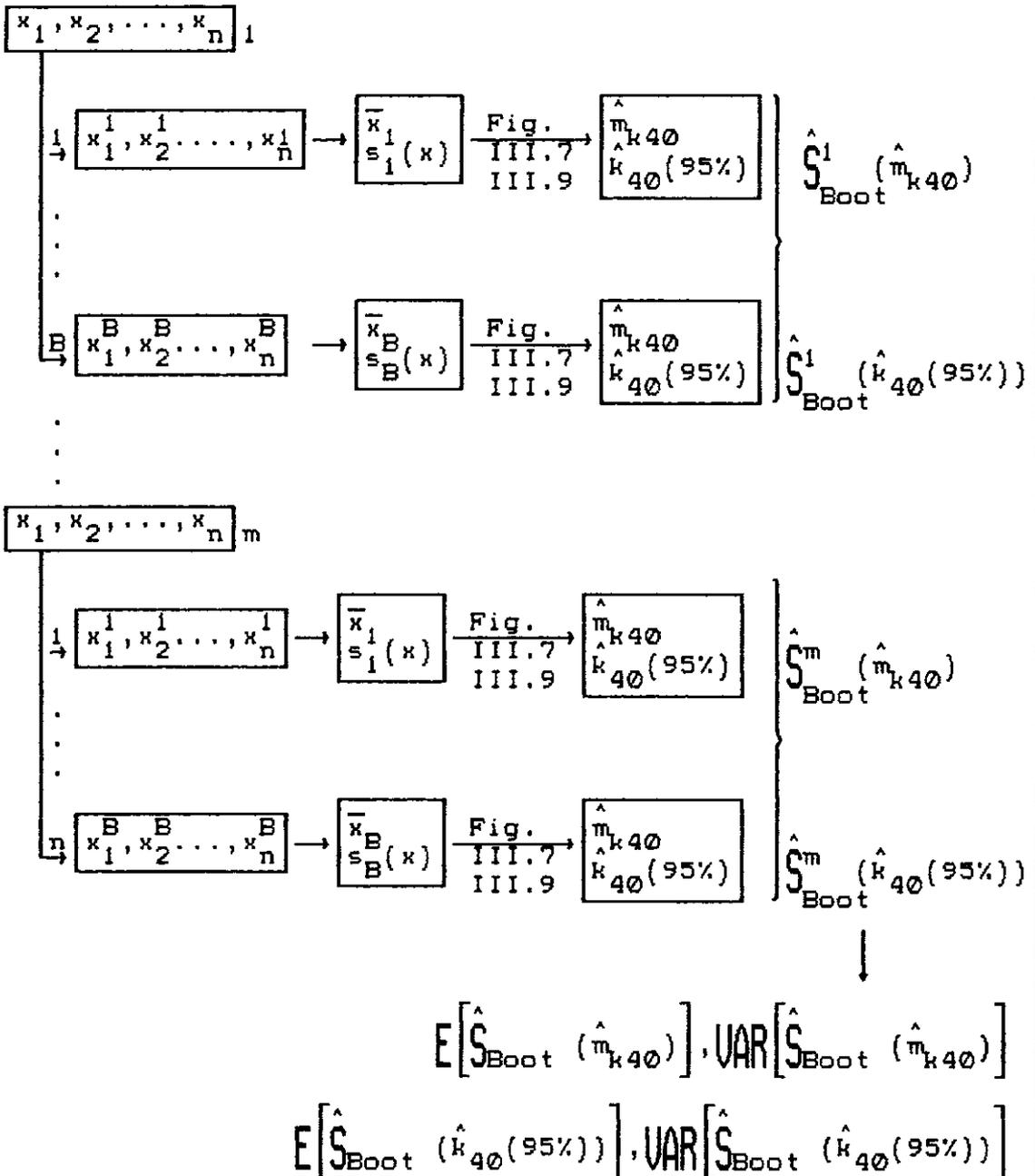


Figura III.18- Esquema para Estimação do Desvio Padrão de \hat{m}_{k40} e de $\hat{k}_{40}(95\%)$ (média e quantil a 95% da Distribuição do Tamanho do Reservatório Capaz de Regularizar $\alpha\%$ da Vazão Média) pelo Bootstrap.

RESULTADOS

Os resultados dos estudos realizados neste item (CASO 3) são apresentados nas tabelas (III.9) à (III.12) para os cenários A e B. Nas tabelas (III.9) e (III.10) os resultados referem-se ao estimador do desvio padrão de \hat{m}_{k40} ($\hat{S}(\hat{m}_{k40})$) sendo \hat{m}_{k40} a média do tamanho do reservatório necessário para um nível de regularização de 90% da vazão média e 40 anos. Nas tabelas (III.11) e (III.12) os resultados referem-se ao estimador do desvio padrão do $\hat{k}_{40}(95\%)$ ($\hat{S}(\hat{k}_{40}(95\%))$), onde $\hat{k}_{40}(95\%)$ é o quantil a 95% do tamanho do reservatório necessário para um nível de regularização de 90% da vazão média e 40 anos.

A observação das tabelas (III.9) a (III.12) mostra-nos que os resultados obtidos para \hat{m}_{k40} e $\hat{k}_{40}(95\%)$ apresentam o mesmo comportamento.

Analisando primeiramente o cenário A temos que o Método Clássico e o Bootstrap subestimam levemente o valor verdadeiro. O Jackknife por outro lado apresenta comportamento inverso, isto é, superestima o valor verdadeiro. De uma forma geral estes três métodos apresentam valores bem próximos do verdadeiro. O mesmo não ocorre com o Método Bayesiano que apresenta uma forte tendência a subestimação.

Quanto a variabilidade as quatro metodologias apresentam valores em torno de 0,20, os menores valores são apresentados pela Bayesiana e pela Clássica.

Quanto ao erro médio quadrático, o Método Clássico apresenta o menor valor, seguido do Bootstrap.

Na análise do Cenário B, tem-se que todas as quatro metodologias subestimam o valor verdadeiro, inclusive o Jackknife que no Cenário A apresentou uma superestimação, sendo que a tendência é mais forte no Bayesiano.

MÉTODO	MÉDIA (z)	DESVIO PADRÃO (z)	COEF VARIACÃO (z)	EMQ ^{1/2} (z)
Clássico	0,0279	0,0060	0,2151	0,0060
Bayesiano	0,0117	0,0023	0,1966	0,0171
Jackknife	0,0295	0,0068	0,2305	0,0069
Bootstrap	0,0276	0,0063	0,2283	0,0064
Valor Verdadeiro	0,0286			

Tabela III.9 - Sumariza os resultados dos estudos realizados para o estimador do desvio padrão de m_{k40} , nível de regularização 90% e cenário A.

$$z = \hat{S}(\hat{m}_{k40})$$

MÉTODO	MÉDIA (z)	DESVIO PADRÃO (z)	COEF VARIACÃO (z)	EMQ ^{1/2} (z)
Clássico	0,854	0,371	0,371	0,322
Bayesiano	0,433	0,131	0,303	0,495
Jackknife	0,745	0,488	0,655	0,515
Bootstrap	0,639	0,364	0,570	0,454
Valor Verdadeiro	0,910			

Tabela III.10- Sumariza os resultados dos estudos realizados para o estimador do desvio padrão de \hat{m}_{k40} , nível de regularização 90% e cenário B.

$$z = \hat{S}(\hat{m}_{k40})$$

MÉTODO	MÉDIA (z)	DESVIO PADRÃO (z)	COEF VARIACÃO (z)	EMQ ^{1/2} (z)
Clássico	0,0463	0,0095	0,2051	0,0095
Bayesiano	0,0328	0,0054	0,1638	0,0150
Jackknife	0,0482	0,0113	0,2344	0,0114
Bootstrap	0,0453	0,0106	0,2340	0,0107
Valor Verdadeiro	0,0468			

Tabela III.11- Sumariza os resultados dos estudos realizados para o estimador do desvio padrão de $\hat{k}_{40}(95\%)$, nível de regularização 90% e cenário A.

$$z = \hat{S}(\hat{k}_{40}(95\%))$$

MÉTODO	MÉDIA (z)	DESVIO PADRÃO (z)	COEF VARIACÃO (z)	EMQ ^{1/2} (z)
Clássico	1,566	0,557	0,356	0,569
Bayesiano	1,023	0,326	0,319	0,733
Jackknife	1,387	0,894	0,645	0,941
Bootstrap	1,183	0,660	0,558	0,826
Valor Verdadeiro	1,680			

Tabela III.12- Sumariza os resultados dos estudos realizados para o estimador do desvio padrão de $\hat{k}_{40}(95\%)$, nível de regularização 90% e cenário B.

$$z = \hat{S}(\hat{k}_{40}(95\%))$$

Quanto a variabilidade o comportamento das quatro metodologias é o mesmo apresentado para o cenário A. O mesmo ocorrendo com o erro médio quadrado.

III.4. CONCLUSÕES

Da análise dos resultados obtidos nos três casos estudados pode-se observar que o Método Clássico apresentou ótimos resultados, exceto para o caso da vazão decamilenar. Entretanto, os resultados obtidos pelas Técnicas de Reamostragem (Bootstrap e Jackknife) são muito satisfatórios, visto que o erro médio quadrático destes é bastante pequeno, sendo que entre as duas Técnicas de Reamostragem, o Bootstrap mostrou-se em geral superior ao Jackknife. Estes resultados confirmam a idéia inicial, exemplo da tabela (II.1), de que estas Técnicas de Reamostragem são métodos alternativos, bastante viáveis, em situações onde o Método Clássico não é disponível ou sua aplicação é muito complexa, visto que o Bootstrap e o Jackknife são técnicas bastante simples de aplicação.

Quanto a Metodologia Bayesiana, que apresentou os resultados menos satisfatórios, pode-se justificar este comportamento pelo fato que a mesma não tem como objetivo a estimação de desvio padrão de estimadores de parâmetros, mas sim determinar a distribuição de probabilidade dos parâmetros de modelos probabilísticos, por exemplo equações (II.14) (II.16), e a partir desta distribuição determinar sua média e desvio padrão, os quais foram utilizados como os estimadores de desvio padrão de estimadores de parâmetros.

Selecionando o Bootstrap como uma boa técnica alternativa, no capítulo seguinte veremos uma aplicação do Bootstrap como forma de investigar uma questão bastante discutida entre os hidrólogos, que constitui na divergência entre o uso de modelos de vazões diárias e modelos de máximos anuais em estudos de cheias.

III.5. SUMÁRIO

Neste capítulo foram apresentados três exemplos de aplicação das Técnicas de Reamostragem e dos Métodos Bayesianos como alternativa aos Métodos Estatísticos Clássicos para o cálculo de desvio padrão de estimadores de parâmetros hidrológicos, como forma de considerar a incerteza paramétrica de modelos probabilísticos.

A aplicação das Técnicas de Reamostragem mostrou-se bastante simples e abrangente e os resultados obtidos foram bastante satisfatórios quando comparados com a Estatística Clássica.

Em virtude do Bootstrap ter se mostrado superior ao Jackknife, no capítulo subsequente o problema de quantidade de informação considerada na modelagem probabilística será abordada através do Bootstrap.

CAPÍTULO IV

COMPARAÇÃO DE METODOLOGIAS PARA MODELAGEM DE CHEIASIV.1- INTRODUÇÃO

Este capítulo aborda duas metodologias para modelagem de cheias. A diferença fundamental entre elas é que uma utiliza vazões máximas anuais e a outra utiliza vazões diárias.

Como foi exposto na introdução desta tese, Capítulo I, a utilização dos máximos anuais leva-nos a trabalhar com variáveis independentes, que são relativamente simples de serem modeladas. Por outro lado, grande parte das informações disponíveis no registro fluviométrico é desprezada. Qualquer tentativa de utilização total do registro fluviométrico, conduz a um estudo mais complexo, que considera a dependência das vazões diárias. Neste caso, é necessário a adoção de modelos estocásticos.

Trabalhar com vazões diárias, implica na utilização de um maior número de informações, talvez conduzindo a resultados melhores que a utilização apenas de máximos anuais.

Alguns trabalhos já foram desenvolvidos em torno deste assunto. KELMAN [27], KELMAN et alia [33] (1983), num estudo comparativo entre modelo estocástico e análise de frequência de cheias, utilizou como "cenário" 100.000 períodos úmidos, gerados por um processo estocástico denominado XADREZ (KELMAN [27]) para barragem de Furnas, Rio Grande, conseguindo desta maneira estimar a distribuição dos máximos anuais, para um tempo de recorrência de até 500 anos. Vinte segmentos de 10 anos foram utilizados por um outro modelo estocástico, DIANA (KELMAN [34]), para produzir 5000 períodos úmidos sintéticos para cada um dos 20 segmentos. As amostras de máximos anuais de cada segmento foram ajustados diferentes distribuições de probabilidades (normal, lognormal-2, parâmetros, lognormal-3

parâmetros, Pearson III, log-Pearson III, exponencial, gumbel, gama-2, generalizada de extremos).

A comparação das distribuições de máximos anuais obtidas tanto pelo DIANA, quanto pelo ajuste das distribuições acima citadas, com a "Distribuição Cenário" (100.000 anos), levou as seguintes conclusões:

O Modelo DIANA, que utiliza a totalidade do registro fluviométrico, conduziu a um conjunto de distribuições de probabilidades dos máximos anuais menos dispersas que as obtidas pelo ajuste das distribuições. No entanto, as distribuições obtidas pelo Modelo DIANA tenderam a subestimação.

Em outro trabalho, KELMAN [9] (1985), analisou este mesmo problema, só que desta vez o modelo estocástico das vazões diárias era conhecido e as três alternativas consideradas foram: utilização dos máximos anuais ajustados à Gumbel, utilização dos máximos anuais ajustados Exponencial e a modelagem de séries temporais, em que apenas os parâmetros foram considerados desconhecidos (e não o processo estocástico em si). Para este estudo a Gumbel mostrou-se a pior das três alternativas. A melhor das alternativas foi a Exponencial, a menos da tendenciosidade negativa na estimação de vazões com tempos de recorrência na faixa de 1000 a 10.000 anos. A utilização de vazões diárias se encontra entre as duas outras alternativas. E no caso de $T=1000$ ou 10.000 anos, pode ser a mais indicada, por não ter apresentado "tendenciosidade negativa".

As experiências controladas que serão estudadas neste capítulo, têm por finalidade verificar as vantagens e desvantagens de se utilizar máximos anuais ou vazões diárias baseado numa nova abordagem, o Bootstrap, técnica de reamostragem descrita no Capítulo II. A adoção desta técnica deve-se a facilidade com que se pode fazer estudos com vazões diárias, e também a possibilidade que as técnicas de reamostragem introduzem no cálculo de intervalos de confiança

da variável em estudo (vazão), assim como determinação da variabilidade desta, apresentadas no Capítulo III desta tese.

IV.2- METODOLOGIA 1: BOOTSTRAP NOS MÁXIMOS ANUAIS

A metodologia apresentada neste item baseia-se na utilização dos máximos anuais.

Consideremos uma série de vazões, constituída de M segmentos de N vazões diárias:

$$\begin{aligned}
 x_{ij} &= \text{vazão diária} && i = 1 \text{ à } M \text{ (segmento)} \\
 & && j = 1 \text{ à } N \text{ (dia)} \\
 \tilde{x}_1 &= \{x_{1,1}, x_{1,2}, \dots, x_{1,N}\} \\
 &\vdots \\
 \tilde{x}_i &= \{x_{i,1}, x_{i,2}, \dots, x_{i,N}\} && \text{(IV.1)} \\
 &\vdots \\
 \tilde{x}_M &= \{x_{M,1}, x_{M,2}, \dots, x_{M,N}\}
 \end{aligned}$$

De cada segmento é retirada a vazão máxima, obtendo-se uma série de M vazões máximas:

$$\begin{aligned}
 y_1 &= \max_{j=1, N} \{x_{1,j}\} \\
 &\vdots \\
 y_i &= \max_{j=1, N} \{x_{i,j}\} && \text{(IV.2)} \\
 &\vdots \\
 y_M &= \max_{j=1, N} \{x_{M,j}\}
 \end{aligned}$$

Logo:

$$\tilde{y} = \{y_1, y_2, \dots, y_M\} \quad \text{(IV.3)}$$

Como os máximos anuais são variáveis aleatórias independentes, podemos aplicar o Bootstrap diretamente aos M máximos anuais (\tilde{y}). Adotando $B = 1000$ ($B = n^\circ$ de reamostragens no Bootstrap), são obtidas 1000 "novas" séries de máximos anuais.

A cada "nova" série de máximos anuais ajusta-se uma distribuição, a qual supõe-se que seja a distribuição de probabilidade da população a partir do qual a série $\{y_1, y_2, \dots, y_M\}$ tenha sido amostrada.

É possível portanto, obter uma estimativa da vazão de tempo de recorrência T (\hat{y}_T), através da extrapolação da distribuição dos y 's.

Desta forma cada "nova" série de máximos anuais produz uma estimativa \hat{y}_T . Como são obtidas 1000 "novas" séries, são produzidas 1000 estimativas \hat{y}_T , que formam a distribuição empírica de probabilidades de \hat{y}_T .

Este procedimento descrito para uma série de vazões diárias é feito para 100 séries como esta.

Podemos observar que nesta metodologia, nenhuma "nova" série de máximos anuais obtida apresenta observações diferentes das observadas na série de máximos original. O que pode ocorrer é a repetição de valores, já que o resorteio no Bootstrap é com repetição.

O Fluxograma da Figura IV.1 descreve esta metodologia.

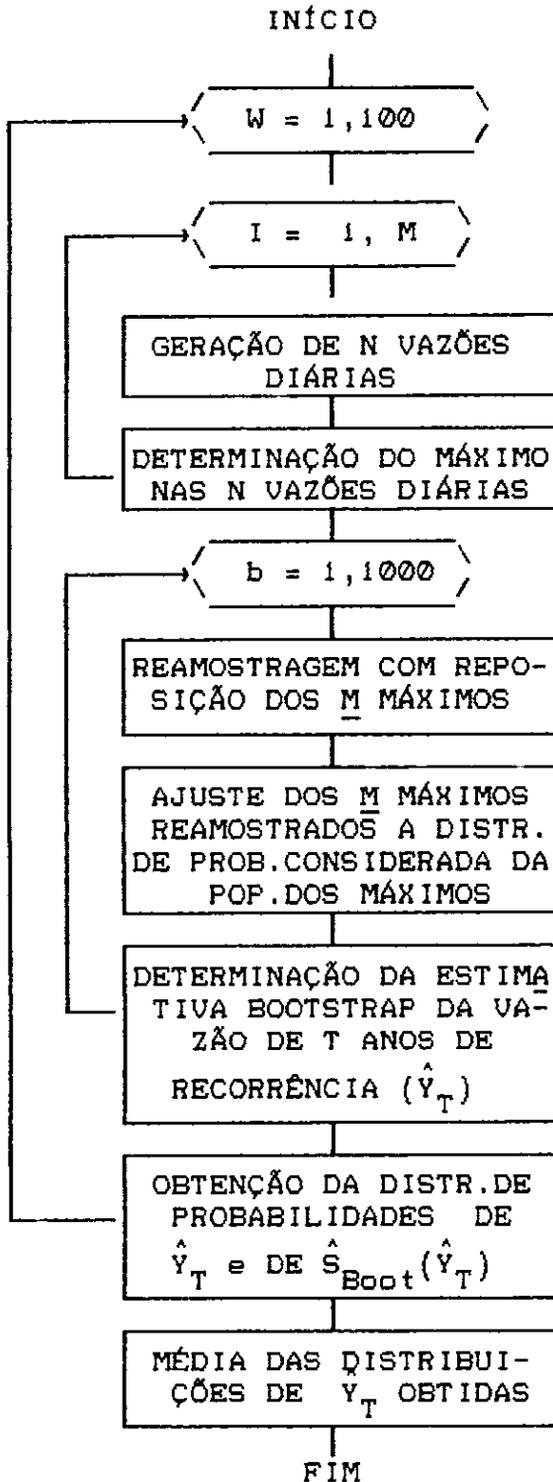


FIGURA IV.1

IV.3- METODOLOGIA 2: BOOTSTRAP NAS VAZÕES DIÁRIAS

A metodologia a seguir apresentada, utiliza as vazões diárias para modelagem de cheias, e consiste basicamente em aplicar o Bootstrap à estas vazões, em lugar de aplicar o Bootstrap aos máximos anuais como a metodologia 1.

No caso de aplicar-se o Bootstrap às vazões diárias alguns cuidados devem ser tomados. Em geral as vazões diárias são variáveis aleatórias dependentes, modeladas por um processo estocástico. Neste caso o Bootstrap não pode ser diretamente aplicado a estas vazões. Para contornar este problema, o Bootstrap é aplicado aos ruídos do processo estocástico, que são variáveis aleatórias independentes e identicamente distribuídas, UNNY et alia [16] (1985), como será visto a seguir.

Consideremos uma série de vazões diárias constituída de M segmentos de N vazões.

$$\begin{aligned}
 x_{ij} &= \text{vazão diária} & i &= 1 \text{ à } M \text{ (segmento)} \\
 & & j &= 1 \text{ à } N \text{ (dia)} \\
 \tilde{x}_1 &= \{x_{1,1}, x_{1,2}, \dots, x_{1,N}\} \\
 &\vdots \\
 \tilde{x}_i &= \{x_{i,1}, x_{i,2}, \dots, x_{i,N}\} & & \text{(IV.4)} \\
 &\vdots \\
 \tilde{x}_M &= \{x_{M,1}, x_{M,2}, \dots, x_{M,N}\}
 \end{aligned}$$

Através da série formada pelas M.N vazões diárias são estimados os parâmetros $\tilde{\theta} = \{\theta_1, \theta_2, \dots, \theta_L\}$ e $\tilde{\phi} = \{\phi_1, \phi_2, \dots, \phi_R\}$ do processo estocástico:

$$x_{ij} = \sum_{\ell=1}^L \theta_{\ell} x_{ij-\ell} + \sum_{k=0}^K \phi_k a_{ij-k} \quad \begin{array}{l} i = 1 \text{ à } M \\ j = 1 \text{ à } N \\ k < N \\ L < N \end{array} \quad (\text{IV.5})$$

Com os parâmetros estimados do modelo e os segmentos $\{x_{i,1}, \dots, x_{i,N}\}$ de vazões diárias estimam-se os segmentos de "ruídos".

$$\begin{array}{l} \tilde{a}_1 = \{a_{1,K+1}, \dots, a_{1,N}\} \\ \vdots \\ \tilde{a}_i = \{a_{i,K+1}, \dots, a_{i,N}\} \\ \vdots \\ \tilde{a}_M = \{a_{M,K+1}, \dots, a_{M,N}\} \end{array} \quad (\text{IV.6})$$

onde: K = ordem do modelo.

Aplica-se então, o Bootstrap a série de $M(N-K)$ ruídos $(a_{i,j-k})$, sendo $B=1000$ (B número de reamostragens do Bootstrap).

Em cada reamostragem é obtida uma "nova" série de $M(N-K)$ ruídos, que são separados em M segmentos, os quais juntamente com os parâmetros estimados darão origem a M "novos" segmentos de N vazões diárias:

$$\begin{array}{l} \tilde{x}_1^b = \{x_{1,1}^b, \dots, x_{1,N}^b\} \\ \vdots \\ \tilde{x}_M^b = \{x_{M,1}^b, \dots, x_{M,N}^b\} \end{array} \quad \begin{array}{l} b = 1 \text{ à } B \\ \\ \end{array} \quad (\text{IV.7})$$

De cada segmento $\{x_{i,1}^b, \dots, x_{i,N}^b\}$ é obtida a vazão máxima y_i .

$$\begin{aligned}
 y_1^b &= \max_{j=1, N} \{x_{1, j}^b\} \\
 &\vdots \\
 y_M^b &= \max_{j=1, N} \{x_{M, j}^b\}
 \end{aligned}
 \tag{IV.8}$$

Obtém-se assim 1000 "novas" série de máximos anuais

$$\tilde{y}^b = (y_1^b, \dots, y_M^b).$$

Como na metodologia 1, a cada "nova" série de máximo anuais ajusta-se uma distribuição de probabilidades a partir da qual a série de máximos foi amostrada.

Por extrapolação desta distribuição é possível obter a estimativa da vazão de T anos de recorrência \hat{y}_T .

Este procedimento é aplicado a 100 séries de vazões diárias.

No caso particular em que as vazões diárias são independentes (coeficiente de correlação, ρ , igual a zero) esta metodologia se torna mais simples, pois o Bootstrap pode ser aplicado diretamente as vazões diárias.

Nesta metodologia, ao contrário da Metodologia 1, novos valores de máximos anuais podem aparecer. No caso particular de $\rho = 0$ os "novos" valores de y pertencem ao conjunto dos x 's (registro de vazões diárias). No caso geral $\rho \neq 0$, as "novas" observações de y (máximos anuais) podem não pertencer ao registro de vazões diárias. Isto é um dos fatores que nos levam a cogitar a hipótese de que a consideração das vazões diárias pode conter maior quantidade de informações.

O Fluxograma da Figura IV.2 descreve esta metodologia.

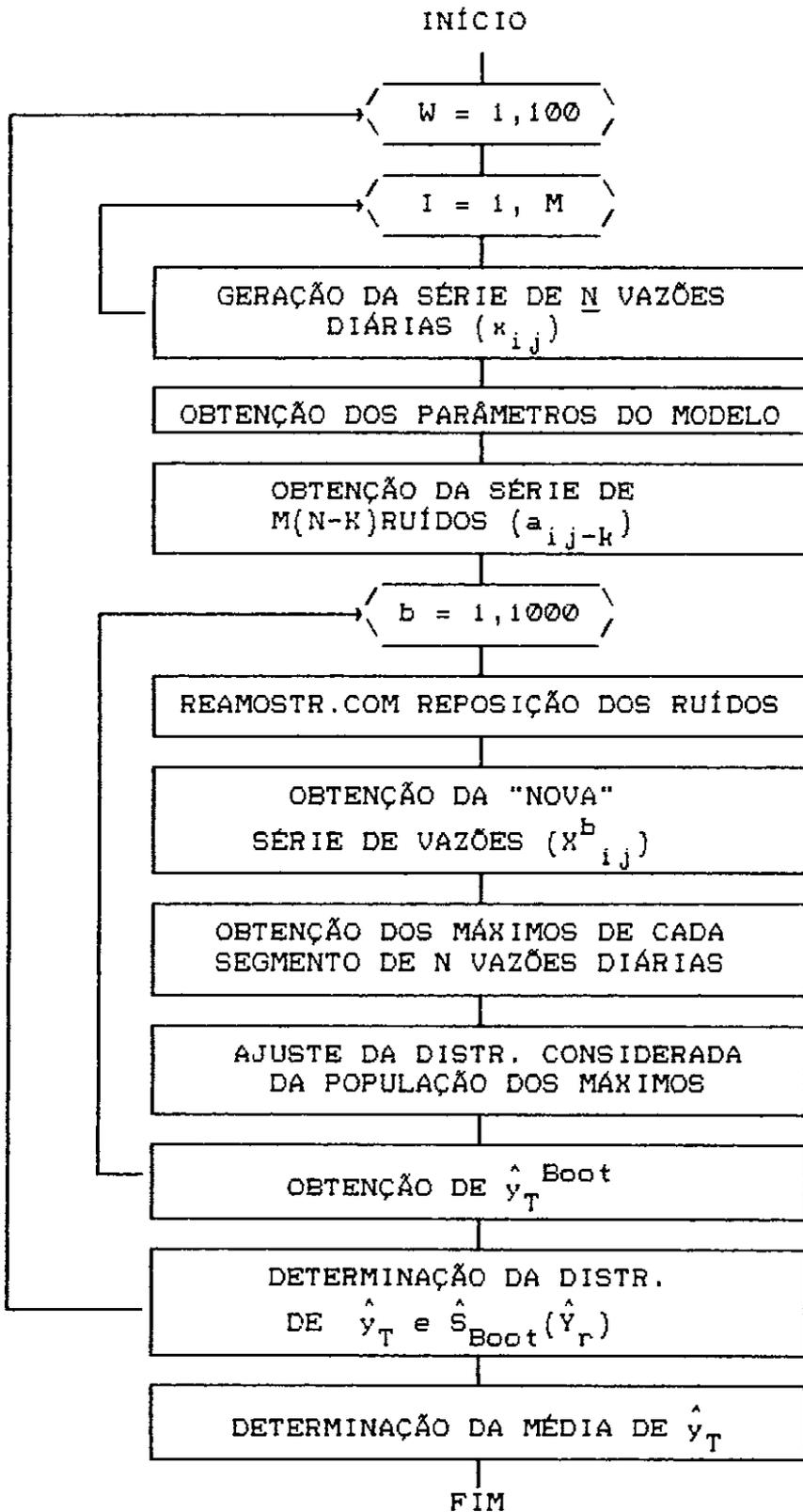


FIGURA IV.2

IV.4- ESTUDOS REALIZADOS

Os estudos realizados para comparação das metodologias, utilizaram como "cenário" 100 séries, cada uma constituída de 10 (M=10) segmentos de 30 (N=30) vazões diárias.

Adota-se um modelo estocástico, autoregressivo de ordem um (AR(1)) para descrever o comportamento das vazões diárias. É importante ressaltar que não podemos esperar que este simples processo estocástico realmente represente as vazões diárias, mas ele é útil para lançar luz no tópico específico deste estudo.

Assim:

$$x(i) = \rho(1) x(i-1) + \sqrt{1-\rho^2(1)} a(i) \quad (\text{IV.9})$$

onde:

$x(i)$ = vazão diária padronizada ($\mu_x = 0$, $\sigma_x = 1$) no tempo i .

$\rho(1)$ = coeficiente de correlação LAG 1, isto é, correlação entre $x(i)$ e $x(i+1)$

$a(i)$ = ruído normal (0,1)

Os coeficientes de correlação adotados neste estudo foram:

$\rho = 0,95$ vazões altamente correlacionadas

$\rho = 0,50$ vazões mediamente correlacionadas

$\rho = 0,00$ vazões independentes

No caso de $\rho = 0$ a expressão (IV.9) reduz-se a:

$$x(i) = a(i) \quad (\text{IV.10})$$

E conseqüentemente, as vazões diárias passam a ser variáveis aleatórias independentes, identicamente distribuídas com distribuição $N(0,1)$.

No caso de $\rho \neq 0$ a primeira vazão de cada um dos segmentos de 30 vazões diárias é independente da última vazão do segmento anterior como se cada segmento correspondesse a um determinado mês de anos diferentes.

A distribuição dos máximos anuais é desconhecida. Como aproximação desta distribuição foi utilizada a distribuição "GUMBEL", uma vez que, um dos principais resultados da "Teoria dos Extremos" é que se as variáveis aleatórias X_i são independentes, com distribuição tipo exponencial, então o máximo definido como $y = \max \{x_1, \dots, x_n\}$ tem distribuição de probabilidades, tal que: (GUMBEL [35], 1958).

$$\lim_{n \rightarrow \infty} F(x) = \exp[-\exp(-\psi(x-\mu))] \quad (\text{IV.11})$$

Para $N = 30$ esta conclusão não é correta, e a extrapolação para valores cujo tempo de recorrência é grande conduz a erros. KELMAN [9] (1985), verificou que neste caso há uma superestimação do valor de y_T , quando a distribuição dos X_i é normal, e uma subestimação quando a distribuição é lognormal. Além disto, no caso da metodologia 2, quando $\rho \neq 0$, as vazões diárias não são independentes.

Apesar disto, a distribuição GUMBEL é largamente utilizada. E foi a adotada neste estudo.

Assim à série de $M = 10$ máximos anuais foi ajustada a distribuição GUMBEL, cujos parâmetros são: (BENJAMIN et alia [4]).

$$\alpha = \frac{\pi}{\sqrt{6} \sigma_y} \quad \mu = m_y - \frac{0,577}{\alpha} \quad (\text{IV.12})$$

onde:

m_y - média dos máximos anuais

σ_y - desvio padrão dos máximos anuais

A extrapolação da distribuição para obter \hat{y}_T é obtida por:

$$\hat{y}_T = \mu + (z/\alpha) \quad (\text{IV.13})$$

sendo $z = 4.60$, no caso de $T = 100$ anos.

O quantil objeto de estudo foi o correspondente ao tempo de recorrência de 100 anos (y_{100}), isto é, a vazão centenária.

A vazão centenária de população (y_{100}) para o caso especial em que $\rho = 0$ e conseqüentemente as vazões diárias são $N(0,1)$, é facilmente obtido, por:

$F_X(x)$ - distribuição das vazões diárias (normal)

$F_Y(y)$ - distribuição dos máximos anuais

$$F_Y(y) = P(Y \leq y) \quad (\text{IV.14})$$

Logo:

$$P(Y \leq y_T) = P(X_1 \leq y_T \cap X_2 \leq y_T \cap \dots \cap X_N \leq y_T) \quad (\text{IV.15})$$

$$F_Y(y) = F_X^{(1)}(y_T) \cdot F_X^{(2)}(y_T) \cdot \dots \cdot F_X^{(N)}(y_T) \quad (\text{IV.16})$$

Como os X são independentes e identicamente distribuídos:

$$F_Y(y) = [F_X(y_T)]^N$$

Para $N = 30$ e $T = 100$ anos:

$$F_Y(y_{100}) = [F_X(y_{100})]^{30}$$

Logo:

$$F_X(y_{100}) = [F_Y(y_{100})]^{1/30}$$

$$F_X(y_{100}) = (1 - 1/T)^{1/30} = (0,99)^{1/30}$$

Assim:

$y_{100} \approx 3,40$ (obtido da tabela normal a partir da probabilidade $F_X(y_{100})$).

No caso geral, em que $\rho \neq 0$, y_{100} não pode ser obtido desta forma. Os valores foram obtidos por simulação, para tal foram gerados 100.000 segmento de 30 vazões diárias. A tabela IV.1 apresenta os valores de y_{100} para os três casos:

	COEF. DE CORRELAÇÃO		
	0.0	0.5	0.95
y_{100}	3.40	3.39	3.07

Tabela IV.1 - Valores da Vazão Centenária para coeficiente de Variação 0,00; 0,50 e 0,95.

IV.5- ANÁLISE DOS RESULTADOS

Primeiramente são avaliadas as metodologias 1 e 2 como método de estimação do desvio padrão e da tendenciosidade de estimadores de parâmetros, no caso em estudo o parâmetro considerado é a vazão centenária (y_{100}). A tabela IV.2 apresenta estes resultados:

P	y_{100}	$\hat{\sigma}y_{100}$	$\bar{\sigma}(\hat{y}_{100})$		BIAS(\hat{y}_{100})	$\overline{\text{BIAS}}$	
			Boot			Boot	
			1	2		1	2
0,00	3,40	0,40	0,37	0,39	0,01	-0,05	-0,02
0,50	3,39	0,57	0,40	0,47	0,02	-0,05	0,08
0,95	3,07	0,76	0,52	0,64	0,08	-0,06	0,10

Tabela IV.2 - Resultados da comparação das metodologias 1 e 2 como métodos de estimação do desvio padrão e da tendenciosidade.

- 1- Metodologia 1: Bootstrap nos máximos anuais
 2- Metodologia 2: Bootstrap nas vazões diárias

onde:

ρ é o coeficiente de correlação das vazões diárias.

y_{100} o valor verdadeiro da vazão centenária considerado as três populações ($\rho=0,00$; $0,50$ e $0,95$).

$\hat{\sigma}_{y_{100}}$ desvio padrão da estimativa de y_{100} clássico, definido por:

$$\hat{\sigma}_{y_{100}} = \left\{ \frac{1}{100} \sum_{j=1}^{100} (\hat{y}_j(100) - \bar{y}(100))^2 \right\}^{1/2} \quad (\text{IV.17})$$

onde:

$$\bar{y}(100) = \frac{1}{100} \sum_{j=1}^{100} \hat{y}_j(100) \quad (\text{IV.18})$$

$\bar{\sigma}_{\text{Boot}}(\hat{y}_{100})$ é o valor esperado do desvio padrão Bootstrap da estimativa da vazão centenária definido por:

$$\bar{\sigma}_{\text{Boot}}(\hat{y}_{100}) = \frac{1}{100} \sum_{j=1}^{100} \sigma_{\text{Boot}}^j \quad (\text{IV.19})$$

onde:

σ_{Boot}^j é obtido da equação (II.36),

sendo $\hat{\theta}_{\text{Boot}}^b = \hat{y}^b(100)$ e

$$\hat{\theta}_{\text{Boot}}^j = \hat{y}_{\text{Boot}}^j(100) = \frac{1}{B} \sum_{b=1}^B \hat{y}^b(100).$$

$\text{BIAS}[\hat{y}_{(100)}]$ é a tendenciosidade de \hat{y}_{100} em relação ao valor verdadeiro, definida por:

$$\text{BIAS}[\hat{y}_{(100)}] = \frac{1}{100} \sum_{j=1}^{100} [\hat{y}_j(100) - y(100)] \quad (\text{IV.20})$$

$\overline{\text{BIAS}}_{\text{Boot}}$ é o valor esperado da tendenciosidade Bootstrap da vazão centenária, definido por:

$$\overline{\text{BIAS}}_{\text{Boot}} = \frac{1}{100} \sum_{j=1}^{100} [y_{\text{Boot}}^j(100) - \hat{y}_j(100)] \quad (\text{IV.21})$$

Analisando os resultados apresentados na tabela IV.2 pode-se ver que a aplicação do Bootstrap às vazões diárias apresentou leve superioridade em relação a aplicação do Bootstrap aos máximos anuais, o que é possível constatar observando os resultados da tabela IV.2 e a figuras IV.3 e IV.4, que corresponde a distribuição de frequência do desvio padrão Bootstrap $\hat{\sigma}_{\text{Boot}}^j$ obtido pelas duas metodologias para o cenário ($p = 0,0$). É possível observar que no segundo caso (vazões diárias) a distribuição é mais concentrada e a frequência associada a $\hat{\sigma}_{100}^j$ é maior do que quando aplicado ao máximo anuais.

Quanto a tendenciosidade, os resultados obtidos pelo Bootstrap não foram tão satisfatórios. No caso da aplicação do Bootstrap aos máximos anuais (metodologia 1) os valores obtidos foram da mesma ordem de grandeza que $\text{BIAS}[\hat{y}_{100}]$ porém negativos, já na metodologia 2 valor negativo só ocorreu para o caso de $p = 0,00$, nos outros dois cenários ($p=0,50$ e $p=0,95$) os valores obtidos foram apenas levemente superiores aos valores de $\text{BIAS}[\hat{y}_{100}]$. Pode-se dizer que quanto a tendenciosidade a metodologia 2 é mais eficiente.

A segunda forma de analisar as metodologias 1 e 2 é considerá-la como método de estimação da vazão centenária y_{100} . Isto é aplicar o Bootstrap como forma de estimação do parâmetro e não de seu desvio padrão. A tabela IV.3

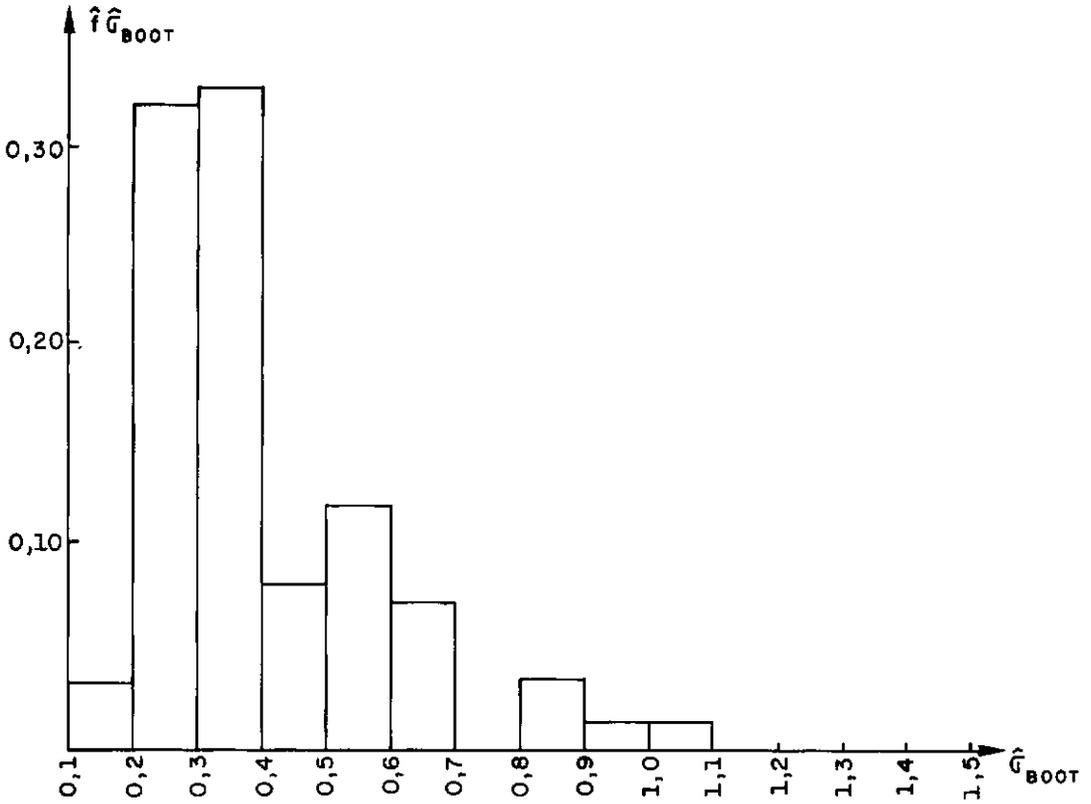


FIGURA IV.3 - DISTRIBUIÇÃO DE FREQUÊNCIA DE \hat{G}_{BOOT}
 BOOTSTRAP APLICADO DIRETAMENTE AOS MÁXIMOS ANUAIS
 COEF. DE CORRELAÇÃO DAS VAZÕES DIÁRIAS IGUAL A ZERO

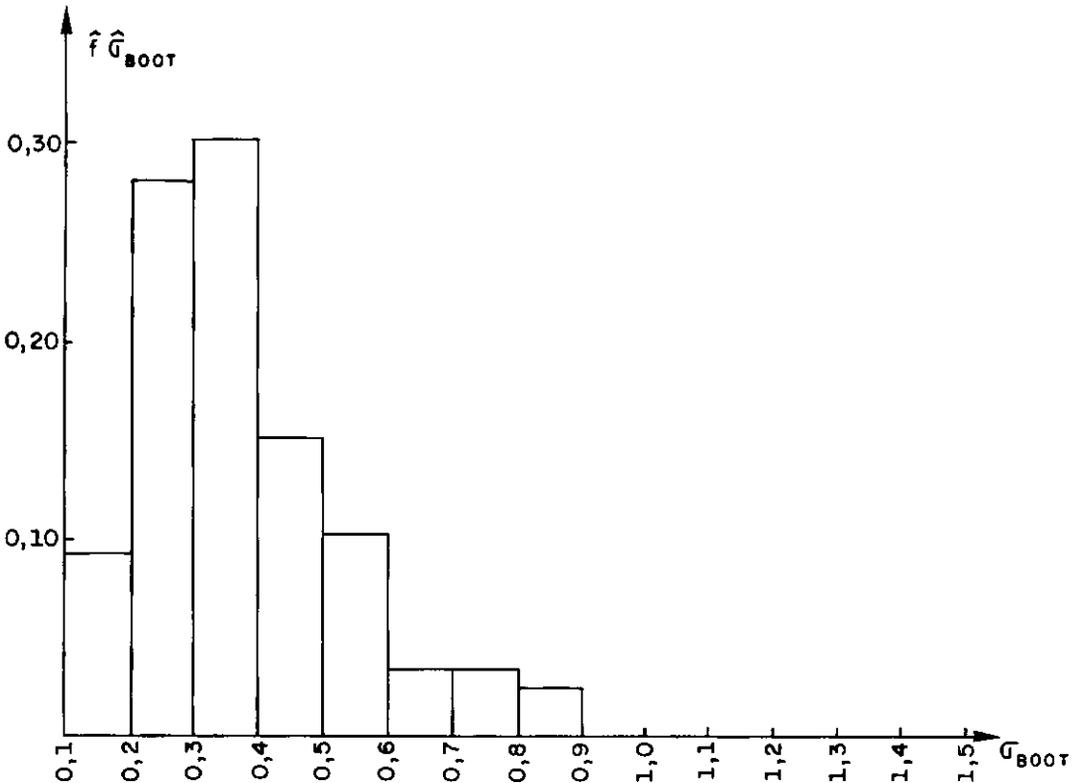


FIGURA IV.4 - DISTRIBUIÇÃO DE FREQUÊNCIA DE \hat{G}_{BOOT}
 BOOTSTRAP APLICADO AS VAZÕES DIÁRIAS
 COEF. DE CORRELAÇÃO DAS VAZÕES DIÁRIAS IGUAL A ZERO

apresenta os resultados desta análise.

P	Y ₁₀₀	VAR [\hat{Y}_{100}]	BIAS [\hat{Y}_{100}]	EMQ [\hat{Y}_{100}]	BIAS _{Boot}		EMQ _{Boot}	
					[\hat{Y}_{100}]		[\hat{Y}_{100}]	
					1	2	1	2
0,00	3,40	0,23	0,01	0,24	0,05	0,06	0,20	0,20
0,50	3,39	0,32	0,01	0,34	0,09	0,16	0,36	0,21
0,95	3,07	0,57	0,08	0,66	0,23	0,38	0,67	0,39

Tabela IV.3 - Resultados da aplicação das metodologias 1 e 2 como método de estimação de \hat{y}_{100} .

1- Metodologia 1: Aplicação do Bootstrap aos máximos anuais.

2- Metodologia 2: Aplicação do Bootstrap as vazões diárias.

onde:

$VAR[\hat{y}_{100}]$ é a variância da estimativa clássica de y_{100} .

$BIAS[\hat{y}_{100}]$ é a tendenciosidade de \hat{y}_{100} em relação ao valor verdadeiro, definida por

$$BIAS[\hat{y}(100)] = \frac{1}{100} \sum_{j=1}^{100} [\hat{y}_j(100) - y(100)]. \quad (IV.22)$$

$EMQ[\hat{y}(100)]$ é o erro médio quadrático da estimativa clássica da vazão centenária, $\hat{y}(100)$ em relação ao valor verdadeiro da vazão centenária, $y(100)$, definido por:

$$EMQ[\hat{y}(100)] = \frac{1}{100} \sum_{j=1}^{100} [\hat{y}_j(100) - y(100)]^2 \quad (IV.23)$$

BIAS[$y_{\text{Boot}}(100)$] é a tendenciosidade da estimativa Bootstrap da vazão centenária, $y_{\text{Boot}}(100)$ em relação a $y(100)$, definida por:

$$BIAS[\hat{y}_{\text{Boot}}(100)] = \frac{1}{100} \sum_{j=1}^{100} [y_{\text{Boot}}^j(100) - y(100)] \quad (IV.24)$$

onde:

$$y_{\text{Boot}}^j(100) = \frac{1}{B} \sum_{b=1}^B y^b(100) \quad j=1, \dots, 100 \text{ e}$$

$y^b(100)$ é a vazão centenária obtida em cada reamostragem do Bootstrap.

$EMQ[\hat{y}_{\text{Boot}}(100)]$ é o erro médio quadrado da estimativa BOOTSTRAP da vazão centenária, $y_{\text{Boot}}^j(100)$, em relação a $y(100)$, definida por:

$$EMQ[\hat{y}_{\text{Boot}}(100)] = \frac{1}{100} \sum_{j=1}^{100} [y_{\text{Boot}}^j(100) - y(100)]^2 \quad (IV.25)$$

A comparação dos resultados apresentados na tabela IV.3 conduz as seguintes conclusões:

1- Apesar do Bootstrap ter sido derivado para ser um método de estimação de desvio padrão, variância e tendenciosidade de estimadores de parâmetros, como aplicado nos estudos anteriores (capítulo III e tabela IV.2), e não como método de estimação de parâmetros, os resultados obtidos quando o Bootstrap foi adotado como método de estimação da vazão centenária foram bastante

satisfatórios. Analisando o "estimador Bootstrap" quanto ao erro médio quadrático, $EMQ[y_{Boot(100)}]$, verifica-se que o Bootstrap quando aplicado às vazões diárias apresentou valores consideravelmente menores do que quando aplicado aos máximos anuais e do que o método clássico. O Bootstrap quando aplicado aos máximos anuais apresentou erro médio quadrado equivalente ao método clássico.

2- A análise do "estimador Bootstrap" da vazão centenária mostrou que o Bootstrap é um estimador mais tendencioso que o estimador clássico, porém de menor variância.

IV.6 SUMÁRIO

Neste capítulo analisou-se o problema da quantidade de informação a considerar na modelagem de fenômenos hidrológicos, especificamente modelagem do pico da cheia centenária. Para tal adotou-se o Bootstrap, técnica de reamostragem descrita no Capítulo II, sob duas formas distintas: Bootstrap aplicado aos máximos anuais e Bootstrap aplicado as vazões diárias.

Além disto analisou-se o problema de aplicar o Bootstrap quando as variáveis aleatórias não são independentes (caso das vazões diárias). Neste caso as vazões são modeladas por processos estocásticos e o Bootstrap aplicado ao "ruídos" independentes do processo.

Os estudos realizados indicam que a aplicação do Bootstrap as vazões diárias produz resultados levemente superiores.

CAPÍTULO V

CONCLUSÕES

O trabalho desenvolvido nesta tese teve como objetivos:

- 1 - A introdução das Técnicas de Reamostragem (especificamente o Bootstrap e o Jackknife) no conjunto de ferramentas estatísticas disponíveis para a consideração da incerteza paramétrica, através da estimação do desvio padrão de estimadores de parâmetros para a construção de intervalos de confiança destes parâmetros.

- 2 - A avaliação do Bootstrap na determinação do pico da cheia de T anos de recorrência e de seu desvio padrão, aplicado de duas formas distintas (utilização de vazões diárias e utilização de máximos anuais) com a intenção de avaliar a questão da quantidade de informação a utilizar na modelagem de um fenômeno hidrológico.

Quanto a consideração da incerteza paramétrica podemos verificar que a utilização das Técnicas de Reamostragem para cálculo de desvio padrão de estimadores conduz a procedimentos extremamente intuitivos e simples, inclusive quando os procedimentos da estatística clássica não são disponíveis e necessitam ser derivados, o que pode ser observado no Caso 3 do Capítulo III, onde o interesse era a determinação do desvio padrão da média e do quantil a 95% da distribuição do tamanho do reservatório capaz de regularizar 90% da vazão média. Neste caso fez-se necessário derivar as curvas que determinam estes valores (Figuras III.10 e III.11). Estas curvas entretanto são específicas para os níveis de regularização de 70%, 90% e 95%, afluências log-normais e um período de 40 anos de vazões, enquanto que as Técnicas de Reamostragem dispensam esta derivação.

Cabe mencionar que nos casos em que a teoria estatística clássica dispõe de fórmulas ou procedimentos para o cálculo do desvio padrão de estimadores de parâmetros, as Técnicas de Reamostragem não apresentam grandes vantagens, (caso 1, Capítulo III). Mesmo assim os resultados obtidos são muito bons, próximos aos da estatística clássica, (Tabelas III.1 e III.2).

A análise dos resultados apresentados no Capítulo III permite concluir que as Técnicas de Reamostragem experimentadas (Bootstrap e Jackknife) são razoavelmente precisas, podendo ser usadas na construção de intervalos de confiança para parâmetros hidrológicos. A tendência destas técnicas à subestimação do tamanho do intervalo é apenas levemente pior que à mesma tendência das fórmulas clássicas. Dentre as duas, a técnica do Bootstrap se revelou levemente superior. Os Métodos Bayesianos não se mostraram competitivos, o que pode ser justificado pelo fato de não terem sido desenvolvidos, para esta aplicação.

Ainda sobre a consideração da incerteza dos parâmetros, ressaltamos que neste trabalho fixamos nosso interesse na comparação de técnicas de cálculo de desvio padrão para posterior construção do intervalo de confiança pela forma clássica. Entretanto, EFRON [19], apresenta uma aproximação para cálculo de intervalos de confiança para parâmetros (ex: mediana), considerando pequenas amostras e tendo como método Técnicas de Reamostragem, que valem a pena serem investigadas em trabalhos futuros.

Quanto a quantidade de informação a considerar na modelagem de fenômenos hidrológicos, podemos verificar que no Bootstrap a consideração de maior quantidade de informações também conduz a procedimentos mais complexos.

De uma forma geral a utilização das vazões diárias produziu resultados levemente melhores do que a consideração exclusiva dos máximos anuais.

BIBLIOGRAFIA

- [1] FULLER, W.E., "Flood Flows", Trans. ASCE, Vol, 77, pp. 564-517, 1914.
- [2] USWRC, Guideline for Determining Flood Frequency Bulletin 17A of the Hydrology Committee, 1977.
- [3] ASHKAR, F., BOBÉE, B., FORTIER, L., "Confidence Intervals for Design Flood Events under Different Statistical Flood Models", International Symposium on Flood Frequency and Risk Analysis, Baton Rouge, Louisiana, 1986.
- [4] BENJAMIN, J.R., CORNELL, G.A., Probability Statistics and Decision for Civil Engineers, McGraw-Hill Book Company, 1970.
- [5] DONALD, R.D., KISIEL, C.C., DUCKESTEIN, L., "Bayesian Decision Theory Applied to Design in Hydrology", Water Resources Research, Vol. 8, N° 1, 1972.
- [6] QUENOUILLE, M.M., "Approximate Tests of Correlation in Time Series", J.R. Statist. Soc., B11, pp. 68-84, 1949.
- [7] TUKEY, J.W., "Bias and Confidence in not Quite Large Sample", Ann. Math. Statist, Vol. 29, 1958.
- [8] EFRON, B., "Bootstrap Method: Another Look at Jackknife", Ann. Statist., Vol, 7, pp. 1-26, 1979.

- [9] HELMAN, J., "Statistical Approach to Flood", Simposia on Statistics in Honours of Professor V.W. JOSHI's 70th Birthday, University of Western Ontario, Canadá, maio 1985.
- [10] RAIFFA, M., SCHLAIFER, R., Applied Statistical Decision Theory, the M.I.T. Press, Massachusetts Institute of Technology, 1977.
- [11] DAMÁZIO, J.M., "Aplicação de Modelos Regionais e Conceituais em Estudos Hidroenergéticos. Uma Abordagem Bayesiana", Tese de Mestrado, COPPE, UFRJ. 1980.
- [12] BOX, G.E.P., TIAO, G.C., Bayesian Inference in Statistical Analysis, Addison-Wesley Publishing Company, 1973.
- [13] STEDINGER, R.J., TAYLOR, M.R., "Synthetic Stream Flow Generation 2. Effect of Parameter Uncertainty" Water Resources Research, Vol. 12, N° 4, pp. 919-924, agosto 1982.
- [14] STEDINGER, R.J., PEI, D., COHN, T.A., "A Condensed Desaggregation Mode. for Incorporation Parameter Uncertainty into Monthly Reservoirs Simulation", Water Resources Research, Vol, 21, N° 5, pp. 665-675, maio 1985.
- [15] VICENS, J.G., RODRIGUES-ITURBE, I. SCHAARKE, J.C.Jr., "Bayesian Generation of Synthetic Streamflow", Water Resources Research, Vol, 11, N° 6, pp. 827-838

- [16] UNNY, T.E., COVER, K.A., "Application of Computer Intensive Statistics to Parameters Uncertainty in Streamflow Synthesis", Symposia on Statistics in Honours of Professor V.W. Joshi's 70th Birthday, University of Western Ontario, Canada, maio 1985.
- [17] SANGIT. C., "Bootstrapping ARMA Model: Some Simulations", IEEE Transactions on Systems, Man and Cybernetics. Vol. SMC-16, N° 2, março/abril 1986.
- [18] EFRON, B., "Computers and the Theory of Statistics: Thinking the Unthinkable", SIAM, Rev. 21, pp. 460-480, 1979.
- [19] EFRON, B., The Jackknife, The Bootstrap and Others Resampling Plans, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1982.
- [20] QUENOUILLE, M., "Notes on Bias in Estimation". Biometrika, Vol, 43, pp. 353-360, 1956.
- [21] TUCKEY, J.W., "Data Analysis and Behavioral Sciences", Unpublished Manuscripte.
- [22] HINKLEY, D.V., "Improving the Jackknife With Special Reference to Correlation Estimation", Biometrika, Vol. 65, pp. 13-21, 1978.
- [23] MILLER, R.G. Jr., "The Jackknife - A Review", Biometrika, Vol. 61, pp. 1-15, 1979.
- [24] YEVJEVICH, V., Probability and Statistics in Hydrology, Water Resources Publications, Fort Collins, Colorado, U.S.A., 1972.

- [25] NAYLOR, T.H., BALINTFY, J.L., BURDICK, D.S., CHU, K., Técnicas de Simulação em Computação, Ed. Vozes em Colaboração com Ed. da Universidade de São Paulo, 1971.
- [26] KITE, G.W., Frequency and Risk Analysis in Hydrology, Water Resources Publication, 1977.
- [27] KELMAN, J., "Cheias e Aproveitamentos Hidrelétricos", Tese para Concursos de Professor Titular, Departamento de Hidrologia e Recursos Hídricos, UFRJ, 1983.
- [28] STEDINGER, J.R. LOUCKS, D.P., HAITH, D.A., Water Resources, Systems, Planning and Analysis, Printice-Hall, Inc. Englewood Clypps, New Jersey, 1981.
- [29] DAMÁZIO, J.M., KELMAN, J., "Vazão Garantida por um Reservatório", III Simpósio Brasileiro de Hidrologia, Brasília, 1979.
- [30] RIPPL, W., "The Capacity of Storage Reservoirs for Water Supply", Proc. Instm. Civil Enges. Vol. 17, pp. 270-278, 1883.
- [31] KELMAN, J., "Modelos Estocásticos no Gerenciamento dos Recursos Hídricos, in: Modelos para Gerenciamento de Recursos Hídricos, Nobel/ABRH, 1987.
- [32] GOMIDE, F.L.S., "The Application of Surplus, Deficit and Range in Hydrology", Hydrology Papers, Vol 1, N° 10, Colorado University, Fort Collins, Colorado 1985.

- [33] KELMAN, J., DAMÁZIO, J.M., "Synthetic Hydrology and Spillway Design". XX Congress International Association for Hydraulic Research, Moscou, 1983
- [34] KELMAN, J., DAMÁZIO, J.M., COSTA, J.P., "Geração de Séries Sintéticas de Jazões Diárias - Modelo DIANA", Revista Brasileira de Engenharia, Caderno de Recursos Hídricos, Vol, 1, N° 2, 1983.
- [35] GUMBEL, E.J., "Statistics of Extremes", Columbia University Press, 1958.