

APLICAÇÃO DE MODELOS REGIONAIS E CONCEITUAIS  
EM ESTUDOS HIDROENERGÉTICOS. UMA ABORDAGEM  
BAYESIANA

JORGE MACHADO DAMAZIO

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 (M.Sc.).

Aprovado por:

  
\_\_\_\_\_  
JERSON KELMAN  
(Presidente)

  
\_\_\_\_\_  
PEDRO GUERRERO SALAZAR

  
\_\_\_\_\_  
BASÍLIO BRAGANÇA PEREIRA

  
\_\_\_\_\_  
JORGE XAVIER DA SILVA

RIO DE JANEIRO, RJ - BRASIL

AGOSTO DE 1980

**ii**

À Ruth

## AGRADECIMENTOS

Ao Prof. Jerson Kelman, seguro e eficiente orientador desta Tese.

À COPPE, na pessoa do Dr. Rui Carlos Vieira.

Ao CEPEL, na pessoa do Dr. João Barros, chefe do Departamento de Sistemas, pelo apoio e estímulo prestados.

Ao Projeto de Hidrologia e Climatologia da Amazônia, na pessoa do engenheiro Eduardo Basso, Diretor Internacional, pela gentil cessão dos arquivos em "malha quadrada".

Aos colegas do Departamento de Sistema do CEPEL, pelas diversas sugestões durante o desenvolvimento da Tese.

Às secretárias e desenhistas do CEPEL, pelos excelentes trabalhos de datilografia e execução gráfica respectivamente.

Finalmente, um agradecimento especial a meus pais, Paulo Rubens e Carmen Dora, pelo carinho com que sempre acompanharam meus estudos.

RESUMO

Esta tese tem como objetivo avaliar o efeito do uso de informações regionais e conceituais, bem como da inclusão de medidas da variação amostral, nos projetos de reservatórios. O assunto é desenvolvido através da apresentação de um estudo de caso, especificamente o aproveitamento de Carolina no rio Tocantins. A metodologia adotada é de caráter geral, podendo ser implementada em outros casos.

A teoria bayesiana é o instrumento básico usado. São apresentados, de forma sucinta, os principais resultados desta teoria para o caso em que as vazões médias anuais sejam independentes e normalmente distribuidas. Outros processos estocásticos geradores de vazões anuais não são considerados.

Os dados fisiográficos, pluviométricos e fluviométricos para a Bacia do Rio Tocantins são obtidos de arquivos elaborados pelo Projeto de Hidrologia e Climatologia da Amazônia (SUDAM) a partir do "método da malha quadrada". Estas informações alimentam um modelo de regressão multivariado com o objetivo de obter estimativas para a média e variância das vazões anuais no aproveitamento.

Um modelo conceitual do balanço hídrico é utilizado para completar o registro fluviométrico a partir de registros de precipitação. A série extendida é então empregada na obtenção de novas estimativas.

A capacidade de geração do aproveitamento hidroelétrico é definida pela curva que relaciona demanda energética com risco de não atendimento num ano qualquer. Adotando 5% como um risco aceitável o potencial energético do aproveitamento é avaliado para três casos:

- (i) Usando somente as informações amostrais;
- (ii) Informações amostrais e conceituais combinadas;
- (iii) Informações amostrais, conceituais e regionais combinadas.

O método de Monte-Carlo é aplicado para se obter intervalos de confiança para o risco.

ABSTRACT

This thesis aims to quantify the effect on reservoir design of regional and conceptual information as well as measures of sample variation. The subject is presented by means of the case study of Carolina hydro-electrical plant on the Tocantins River. The adopted method is general and may be developed for other cases.

Bayesian theory is the framework from which the main results are derived. Background material of this theory for independent and normally distributed mean annual flows, are presented in concise form. Other stochastic models are not considered.

The physiographic, precipitation and streamflow data for the Tocantins basin are obtained from files developed by the Projeto de Hidrologia e Climatologia da Amazônia (SUDAM) that used the "square grid method". A multivariate regression model has used these informations to estimate the mean and variance of the annual flows at Carolina.

A conceptual model of the water balance is utilized to extend the streamflow data through the use of precipitation data. Then the extended serie is used to obtain new estimates.

The generation of the hydro-electric plant is defined by the relation between annual power demand and the risk of falling below this target. Adopting the desired risk equal to 5%, the capacity of the plant is estimated for three cases:

- (i) Using only the sample information.
- (ii) Using the sample combined with the results yielding from the conceptual model.
- (iii) Using the sample combined with the results yielding from the conceptual and regional models.

The Monte-Carlo method is used to obtain confidence intervals for the risk.

# Í N D I C E

## CAPÍTULO I - INTRODUÇÃO

I.1	- Os Objetivos .....	1
I.2	- A Região Estudada .....	1
I.3	- O Arquivo da Malha Quadrada .....	3
I.4	- Comentários Sobre Análise de Reservatórios com Escassez de Dados .....	3

## CAPÍTULO II - INFERÊNCIA BAYESIANA

II.1	- Introdução .....	5
II.2	- Conceituação	
	II.2.1 - Teorema de Bayes .....	5
	II.2.2 - Estatísticas Suficientes .....	6
	II.2.3 - Distribuições Conjugadas Naturais ....	7
	II.2.4 - Estimação Bayesiana .....	8
II.3	- O Processo Normal Independente	
	II.3.1 - Definição .....	8
	II.3.2 - Estatísticas Suficientes .....	9
	II.3.3 - Distribuição Conjugada Natural .....	10
	II.3.4 - Distribuição Posterior dos Parâmetros ....	12
	II.3.5 - Distribuição Bayesiana de Vazões ....	13
II.4	- Sumário .....	14

**CAPÍTULO III - A DISTRIBUIÇÃO "A PRIORI" OBTIDA POR REGRESSÃO**  
**MULTIVARIADA LINEAR**

III.1 - Introdução .....	15
III.2 - Conceituação .....	15
III.3 - A Escolha do Modelo .....	18
III.4 - Aplicação	
III.4.1 - Precipitação .....	21
III.4.2 - A Escolha das Variáveis .....	25
III.4.3 - Hipótese A: Independência Espacial .....	25
III.4.4 - Hipótese B: Correlação Espacial .....	29
III.4.5 - Conclusão .....	33
III.5 - Sumário .....	34

**CAPÍTULO IV - O USO DE UM MODELO CONCEITUAL**

IV.1 - Introdução .....	35
IV.2 - O Modelo .....	36
IV.3 - Estimativa da Radiação Solar Absorvida .....	37
IV.4 - Aplicação .....	38
IV.5 - Sumário .....	46

**CAPÍTULO V - AVALIAÇÃO DA CAPACIDADE DE GERAÇÃO ENERGÉTICA**  
**DE UM APROVEITAMENTO HIDROELÉTRICO**

V.1 - Introdução .....	47
------------------------	----

V.2	- O Método .....	47
V.3	- Aplicação .....	50
	V.3.1 - Uso da Amostra de Vazões Medidas .....	51
	V.3.2 - Uso das Informações Fornecidas pelo Modelo Conceitual do Balanço Hídrico .....	53
	V.3.3 - Uso das Informações Fornecidas pelos Modelos de Regressão .....	55
	V.3.4 - Aplicação do Método de Monte Carlo .....	57
V.4	- Sumário .....	61
CAPÍTULO VI - <u>CONCLUSÕES</u> .....		62
ANEXO 1	- <u>DESCRICAÇÃO DOS ARQUIVOS DA MALHA QUADRADA DE-</u> <u>SENVOLVIDOS PELO PHCA NA BACIA DO RIO TOCANTINS</u> .....	64
ANEXO 2	- <u>APROXIMAÇÃO PARA A CORRELAÇÃO ENTRE ESTIMA-</u> <u>DORES DE MÉDIA E VARIÂNCIA</u> .....	70
ANEXO 3	- <u>NÚMERO DE ANOS EQUIVALENTES DE INFORMAÇÃO</u> <u>DO MODELO CONCEITUAL</u> .....	73
BIBLIOGRAFIA .....		76

## CAPÍTULO I

### INTRODUÇÃO

#### I.1      OS OBJETIVOS

A avaliação de recursos hídricos de áreas inexploradas, como a Amazônia, é frequentemente realizada com escassez de dados hidrológicos. Usualmente também não estão disponíveis outras informações potencialmente úteis tais como: dados topográficos, pedológicos, geológicos, etc... Felizmente estas informações dependem de medições que podem ser feitas num espaço de tempo relativamente curto (por exemplo, projeto RADAM), ao contrário do banco de dados hidrológicos, que só pode crescer lentamente.

Este estudo pretende atingir simultaneamente dois objetivos:

1- Avaliar o efeito da inclusão das incertezas envolvidas nas estimativas de parâmetros hidrológicos no que diz respeito a projetos de aproveitamentos hidroelétricos.

2- Quantificar a importância do uso de outras informações, além das hidrológicas, através do uso de um modelo conceitual e de técnicas de regressão multivariada.

#### I.2      A REGIÃO ESTUDADA

A região estudada, abrange uma área de 300.000 km<sup>2</sup> na região centro-oeste brasileira localizada entre as latitudes 5°S e 18°S, e entre as longitudes 53°W e 46°W, e engloba a bacia hidrográfica do alto e médio curso do rio Tocantins. O sistema de drenagem é formado pelos formadores do rio Tocantins, rio Almas e rio Maranhão, e pelos seguintes afluentes do rio Tocantins: Bagagem, Tocantinzinho, Paraná, Palma, Manoel Alves da Natividade, Balsas, Sono, Perdida, Manoel Alves Grande e Farinha, todos pela margem direita, e Santa Tereza e Crixás pela margem esquerda (Figura A.1, página 67).

A altitude na região varia de 100m a 1.500m, tendo um valor médio de 469m. Os solos em geral são espessos, podendo atingir 50m de profundidade. O tipo de vegetação dominante é o cerrado, ocorrendo formações florestais ao longo dos vales e em áreas de solos úmidos e ricos. Ao norte da região aparecem as matas de babaçu, indicando a transição para o ambiente amazônico. A temperatura média na área é de 24°C (Engevix<sup>1</sup>).

O clima, segundo a classificação de Köppen, é úmido tropical de verão úmido e inverno seco, correspondente às savanas tropicais. Este clima indica pequena variação sazonal da temperatura, com a temperatura do mês menos quente, maior que 18°C e que pelo menos um mês da estação seca tem altura de chuva inferior a 60mm (Engevix<sup>1</sup>).

O período chuvoso, de outubro a abril, coincide com o deslocamento da Zona Intertropical de Convergência (ITCZ) para o hemisfério sul. Neste período a massa de ar continental equatorial se expande e alcança a bacia do Tocantins. Estas massas de ar são bastante úmidas e instáveis, ocasionando diversos temporais. Durante os meses secos do inverno (junho, julho e agosto) a ITCZ está localizada no hemisfério Norte e a região é invadida por massas de ar continentais tropicais frias e secas. Neste período a nebulosidade é pequena e a temperatura sofre maiores amplitudes diárias. Chove em média na região 1.500m por ano, chovendo mais no extremo sul da bacia devido a ocorrência de chuvas orográficas (Engevix<sup>1</sup>).

Em 1975 a Engevix<sup>1</sup> terminou o inventário dos recursos hidro-energéticos do rio Tocantins e parte do rio Araguaia. Segundo o relatório, os trabalhos hidrológicos foram bastante dificultados pela falta de dados. Com efeito, para uma área inventariada de 545.000km<sup>2</sup>, haviam apenas 36 postos fluviométricos que juntos totalizam 248 anos. A divisão de quedas selecionada prevê 18 aproveitamentos localizados na área analisada por esta tese, que juntos representam 3.996MW firmes. Esta tese se concentrará no aproveitamento de Carolina Alto, no rio Tocantins, cujo reservatório foi projetado com  $36,88 \times 10^9 \text{ m}^3$  de volume útil. A área para o N.A. máximo normal é de 3.860km<sup>2</sup> e inundará 125km de rodovia, 5 cidades, 6 povoados e 65km<sup>2</sup> da reserva indígena Xerente.

## I.3

O ARQUIVO DA MALHA QUADRADA

O processo de informações regionais pode ser facilitado pelo uso de um sistema racional de arquivos capaz de ser manuseado de forma simples por computadores. Isto pode ser alcançado retalhando a região numa malha de quadrados. O arquivo pode ser criado de forma que cada registro contenha todas as informações disponíveis para um certo quadrado, inclusive sua localização em relação ao referencial adotado. Este processo foi inicialmente aplicado no Canadá<sup>2</sup>. O governo brasileiro, através do Ministério do Interior e da Superintendência de Desenvolvimento da Amazônia (SUDAM), firmou convênio com o Programa das Nações Unidas para o Desenvolvimento (PNUD), para o desenvolvimento conjunto de estudos hidrometeorológicos na Amazônia através do Projeto de Hidrologia e Climatologia da Amazônia<sup>3</sup> (PHCA). Para se conhecer a viabilidade da aplicação, em projetos de redes de medição hidrológica na bacia Amazônica, de métodos baseados no arquivo da malha quadrada, foi desenvolvido pela PHCA<sup>4</sup> um estudo piloto na bacia do Tocantins a montante de Itaguatins. A região foi dividida em 2.485 quadrados com aproximadamente 11 km x 11 km. Foi criado um arquivo fisiográfico, contendo informações topográficas, de solo, de vegetação e de geologia superficial. Foi criado também um outro arquivo definindo as condições de drenagem na área. Os dois arquivos são descritos no anexo 1. A partir desses arquivos, operações tais como identificação de áreas de drenagem de quaisquer postos e cálculo de suas respectivas características médias são facilmente executadas por computadores. Grande parte dos resultados desta tese são baseados nas informações dos arquivos da PHCA.

## I.4

COMENTÁRIOS SOBRE ANÁLISE DE RESERVATÓRIO COM ESCASSEZ DE DADOS

Existe na literatura um grande número de contribuições usando métodos probabilísticos para se obter as relações entre o tamanho do reservatório, as propriedades estocásticas das afluências, a demanda e a garantia de atendimento. Boas revisões do assunto aparecem em Lloyd<sup>5</sup> e Gomide<sup>6</sup>. No entanto, existem regiões (Exemplo: norte do Canadá, Amazônia) onde os poucos dados

indicam incerteza quanto ao comportamento das afluências ao reservatório. Os resultados obtidos por Klemes<sup>7</sup> indicam que um histórico de 25 anos é incapaz de fornecer boas estimativas para o risco de não atendimento da demanda pelo reservatório. Esta é a motivação de numerosos trabalhos sobre a utilização de dados regionais e de modelos conceituais para tornar mais confiáveis as estimativas de parâmetros hidrológicos. Por relevância aos capítulos seguintes desta tese pode-se citar os trabalhos de Vicens<sup>8</sup>, sobre a metodologia bayesiana para o uso combinado de informações locais e regionais; Gomide<sup>9</sup> que discuti o uso de correlação entre estações e a aplicação de modelos lineares e Carey e Haan<sup>10</sup>, sobre o uso de modelos conceituais para a extensão de dados fluviométricos.

CAPÍTULO IIINFERÊNCIA BAYESIANAII.1 INTRODUÇÃO

A metodologia bayesiana, apresentada a seguir, permite usar todas as informações disponíveis (amostrais e não amostrais) tanto em problema de inferência quanto de estimação. A informação não amostral pode ser obtida de dados regionais, ou de considerações teóricas ou mesmo de julgamentos subjetivos do hidrólogo. Quando se deseja usar estimativas pontuais do parâmetros, o procedimento bayesiano leva em conta, formalmente, as consequências de estimativas erradas, usando funções-prejuízo. Uma boa referência sobre os métodos bayesianos pode ser encontrada em Raiffa e Schlaifer<sup>11</sup>.

II.2 CONCEITUAÇÃOII.2.1 Teorema de Bayes

A análise a seguir assume que se conhece o vetor  $z = (z_1, \dots, z_n)$  o qual contém  $n$  observações de vazão média anual em um certo posto. O modelo probabilístico é totalmente conhecido exceto nos valores de seus  $p$  parâmetros arrumados no vetor  $\theta = (\theta_1, \dots, \theta_p)$ . O valor da função densidade de  $z$  condicionada ao valor de  $\theta$  será chamado de verossimilhança e denotando  $\ell(z|\theta)$ . Existe ainda disponível uma função de densidade de probabilidade para  $\theta$ ,  $f'(\theta|Ir)$ , obtida de informações não amostrais  $Ir$ . Chamaremos  $f'(\theta|Ir)$  de distribuição "a priori" dos parâmetros.

O teorema de Bayes torna possível combinar as informações não amostrais e amostrais contidas em  $f'(\theta|Ir)$  e  $\ell(z|\theta)$  obtendo uma nova função de densidade de probabilidade,  $f''(\theta|Ir, z)$ , calculada por:

$$f''(\theta | Ir, z) = f'(\theta | Ir) \cdot \ell(z | \theta) \cdot N(z) \quad (2.1)$$

onde a constante  $N(z)$  é definida pela condição:

$$\int_{\theta} f''(\theta | Ir, z) d\theta = 1 \quad (2.2)$$

Chamaremos  $f''(\theta | Ir, z)$  de distribuição posterior dos parâmetros.

A função densidade de probabilidade bayesiana de uma observação futura de vazão média anual,  $q_f$ , é dada por:

$$\tilde{f}(q_f | Ir, z) = \int_{\theta} \ell(q_f | \theta) \cdot f''(\theta | Ir, z) d\theta \quad (2.3)$$

Esta distribuição inclui a aleatoriedade natural do processo e a incerteza restante nos valores dos parâmetros após o uso das informações em  $z$  e  $Ir$ .

### III.2.2 Estatísticas Suficientes

Chamaremos "kernels" as funções em  $\theta$ ,  $K'(\theta | Ir)$  e  $k(z | \theta)$  se elas são proporcionais respectivamente à distribuição "a priori" dos parâmetros e à função verossimilhança.

$$f'(\theta | Ir) = K'(\theta | Ir) \cdot N(Ir) \quad (2.4)$$

$$\ell(z | \theta) = k(z | \theta) \cdot N(z) \quad (2.5)$$

O teorema de Bayes pode ser escrito por:

$$f''(\theta | Ir, z) \propto K'(\theta | Ir) \cdot k(z | \theta) \quad (2.6)$$

onde a constante de proporcionalidade pode ser sempre calculada pela condição (2.2).

Conforme a equação (2.6) nada se lucra ao se explicitar vetores amostrais com mesma kernel  $k(z | \theta)$ . Neste caso pode

existir um outro vetor  $y = \tilde{y}(z)$ , possivelmente mais simples que  $z$ , que ainda contenha todas as informações relevantes para o problema, ou seja, se a função verossimilhança pode ser escrita como:

$$\ell(z|\theta) = k(\tilde{y}(z)|\theta) \cdot p(z) \quad (2.7)$$

então  $y = \tilde{y}(z)$  é um vetor suficiente, e os seus componentes estatísticas suficientes.

Seja um vetor suficiente  $y = \tilde{y}(z)$  e seja  $y_1 = \tilde{y}(z_1, \dots, z_i)$  e  $y_2 = \tilde{y}(z_{i+1}, \dots, z_n)$ . Então é possível encontrar a operação:

$$y = y_1 * y_2 \quad (2.8)$$

tal que:

$$k(y|\theta) \propto k(y_1|\theta) \cdot k(y_2|\theta) \quad (2.9)$$

### II.2.3 Distribuições Conjugadas Naturais

Quando, as informações não amostrais não especificam uma função densidade de probabilidade para  $\theta$ , informando apenas conceitos tais como "melhor estimativa" de  $\theta$  e "erro médio" desta estimativa, justifica-se o uso conveniente de uma família paramétrica  $\{h(\theta)\}$  de funções densidade de probabilidade. A escolha da distribuição "a priori" é feita ajustando-se os parâmetros de forma a corresponder às informações não amostrais. Famílias convenientes são aquelas que simplificam os cálculos de obtenção da distribuição posterior dos parâmetros e são capazes de expressar qualquer tipo de informação não amostral. Uma das famílias mais usadas são as chamadas distribuições conjugadas naturais. Cada membro desta família é indexada por um elemento do conjunto  $Y$  dos vetores suficientes (chamado vetor dos parâmetros priorísticos) e possui uma kernel similar a da verossimilhança do processo. Se a distribuição escolhida for indexada

da por  $y'$  e o vetor amostral nos fornecer as estatísticas suficientes  $y$ , a distribuição posterior dos parâmetros pertencerá a  $\{h(\theta)\}$  e será indexada por outro elemento de  $Y$ ,  $y''$  calculado por (2.8), chamado vetor dos parâmetros posteriores.

#### II.2.4 Estimação Bayesiana

A metodologia bayesiana encara a questão da estimação de  $\theta$  como um problema de otimização. O objetivo é obter uma estimativa ótima de  $\theta$ ,  $\hat{\theta}$ , que minimiza o valor esperado de uma função prejuízo  $C(\theta, \hat{\theta})$ . Seja o caso simples de se obter a estimativa ótima de um parâmetro  $\theta_1$ , que minimiza o valor esperado da função prejuízo:

$$C(\theta_1, \hat{\theta}_1) = (\theta_1 - \hat{\theta}_1)^2 \quad (2.10)$$

Neste caso a solução é:

$$\theta_1^* = E(\theta_1) = \int f''(\theta_1 | I_r, z) \cdot \theta_1 \delta\theta_1 \quad (2.11)$$

#### II.3 O PROCESSO NORMAL INDEPENDENTE

Nesta seção se apresentará resultados necessários para a análise bayesiana de processos normais independentes.

##### II.3.1 Definição

Define-se um processo normal independente como um gerador de valores quaisquer (exemplo: vazão anual),  $z_1, z_2, \dots$ , aleatórios, independentes e igualmente distribuídos, com função densidade de probabilidade, de parâmetros  $\theta = (\mu, \sigma)$  dada por:

$$f_N(z_i | \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \cdot \left( \frac{z_i - \mu}{\sigma} \right)^2 \right] \quad (2.12)$$

A função verossimilhança de um vetor  $z = (z_1, \dots, z_n)$  é dada por:

$$\ell(z | \theta) = \left( \frac{1}{\sigma \sqrt{2\pi}} \right)^n \exp \left[ -\frac{1}{2} \frac{\sum (z_i - \mu)^2}{\sigma^2} \right] \quad (2.13)$$

Defina-se agora as estatísticas

$$\bar{z} = \frac{1}{n} \sum z_i \quad (2.14)$$

$$s^2 = \frac{1}{n-1} \sum (z_i - \bar{z})^2 \quad n > 1$$

$$s^2 = 0 \quad n = 1 \quad (2.15)$$

$$v = n - 1 \quad (2.16)$$

### II.3.2 Estatísticas Suficientes

A função em (2.13) pode ser escrita por:

$$\ell(z | \theta) = (2\pi)^{-\frac{n}{2}} \cdot \frac{1}{\sigma^{\delta(n)}} \cdot \frac{1}{\sigma^{(n-1)}} \cdot \exp \left[ -\frac{1}{2} \frac{s^2 v}{\sigma^2} \right].$$

$$\cdot \exp \left[ -\frac{1}{2} n \frac{(\bar{z} - \mu)^2}{\sigma^2} \right] \quad (2.17)$$

onde:

$$\delta(n) = \begin{cases} 0 & \text{se } n = 0 \\ 1 & \text{se } n > 0 \end{cases} \quad (2.18)$$

Uma kernel de  $\ell(z|\theta)$  é dada por:

$$K(z|\theta) = \frac{1}{\sigma^{\delta(n)}} \cdot \frac{1}{\sigma^{(n-1)}} \cdot \exp\left[-\frac{1}{2} \frac{s_v^2}{\sigma^2}\right] \cdot \exp\left[-\frac{1}{2} n \frac{(\bar{z}-\mu)^2}{\sigma^2}\right] \quad (2.19)$$

por (2.7) está claro que,

$$\tilde{y}(z) = (\bar{z}, s^2, n, v) \quad (2.20)$$

forma um vetor suficiente.

Observa-se que a estatística  $v$  é uma estatística redundante. Sua inclusão no vetor suficiente tem como objetivo uniformizar o tratamento entre informações (amostrais ou não amostrais) que só se referem a um dos dois parâmetros. Ou seja, quando  $\bar{z}$  é desconhecido, ou irrelevante,  $n = 0$  e  $v \neq 0$ . Quando  $s^2$  é desconhecido, ou irrelevante,  $v = 0$  e  $n \neq 0$ .

### II.3.3 Distribuição\_Conjugada\_Natural

A distribuição conjugada natural do processo independente normal de parâmetros  $(\mu, \sigma)$  é uma distribuição bivariada, produto de uma normal com uma gama-2 invertida (IG2):

$$f'(\mu, \sigma | \text{Ir}) = f'_N(\mu | \bar{z}', \sigma / \sqrt{n'}) \cdot f'_{\text{IG2}}(\sigma | s', v') \quad (2.21)$$

onde  $\bar{z}'$ ,  $n'$ ,  $v'$  e  $s'^2$  são os parâmetros priorísticos.

A distribuição "a priori" marginal para  $\mu$ , obtida integrando  $\sigma$  em (2.21) é uma distribuição student generalizada com momentos:

$$E(\mu | Ir) = \bar{z}' \quad v' > 1 \quad (2.22)$$

$$V(\mu | Ir) = \frac{v' s'^2}{n' (v' - 2)} \quad v' > 2 \quad (2.23)$$

A distribuição "a priori" marginal para  $\sigma$ , obtida integrando  $\mu$  em (2.21) é uma gama-2 invertida. Usando a teoria de distribuições de funções, a distribuição "a priori" marginal para  $\sigma^2$  é uma gama invertida. Os momentos de  $\sigma$  e  $\sigma^2$  são dados por:

$$E(\sigma | Ir) = \left(\frac{v'}{2}\right)^{\frac{1}{2}} s' \cdot \frac{\Gamma[\frac{v'-1}{2}]}{\Gamma[v'/2]} \quad v' > 1 \quad (2.24)$$

$$V(\sigma | Ir) = \frac{v' s'^2}{v'-2} - E^2[\sigma | Ir] \quad v' > 2 \quad (2.25)$$

$$E(\sigma^2 | Ir) = \frac{v' s'^2}{v'-2} \quad v' > 2 \quad (2.26)$$

$$V(\sigma^2 | Ir) = \frac{2 \cdot v'^2 s'^4}{(v'-2)^2 (v'-4)} \quad v' > 4 \quad (2.27)$$

Quando são conhecidos os valores de  $E(\mu | Ir)$ ,  $V(\mu | Ir)$ ,  $E(\sigma^2 | Ir)$  e  $V(\sigma^2 | Ir)$ , as equações (2.22), (2.23) e (2.26), (2.27) fornecem:

$$\bar{z}' = E[\mu | Ir] \quad (2.28)$$

$$n' = \frac{E(\sigma^2 | Ir)}{V(\mu | Ir)} \quad (2.29)$$

$$v' = 2 \cdot \frac{E^2[\sigma^2 | Ir]}{V[\sigma^2 | Ir]} \quad (2.30)$$

$$s'^2 = \frac{v'-2}{v'} \cdot E[\sigma^2 | Ir] \quad (2.31)$$

### II.3.4 Distribuição Posterior dos Parâmetros

Conhecidos um vetor com as estatísticas suficientes,  $y = (\bar{z}, s^2, n, v)$  e os parâmetros de uma distribuição "a priori" de  $\theta$ ,  $y' = (\bar{z}', s'^2, n', v')$  a operação (2.8) que define os parâmetros da distribuição posterior de  $\theta$ ,  $y'' = (\bar{z}'', s''^2, n'', v'')$  é dada por:

$$\bar{z}'' = \frac{n' \bar{z}' + n \bar{z}}{(n'+n)} \quad (2.32)$$

$$n'' = n' + n \quad (2.33)$$

$$v'' = [v' + \delta(n')] + [v + \delta(n)] - \delta(n'') \quad (2.34)$$

$$s''^2 = \frac{v' s'^2 + n' \bar{z}'^2 + v s^2 + n \bar{z}^2 - n'' \bar{z}''^2}{v''} \quad (2.35)$$

A distribuição posterior dos parâmetros, da mesma forma que a "a priori", é um produto de uma normal com um gama-2 invertida.

$$f''(\mu, \sigma | y, Ir) = f''_N(\mu | \bar{z}^n, \frac{\sigma}{\sqrt{n^n}}) \cdot f''_{IG2}(\sigma | s^n, v^n) \quad (2.36)$$

As distribuições posteriores marginais de  $\mu$ ,  $\sigma$  e  $\sigma^2$  são da mesma forma que as correspondentes distribuições "a priori" e com momentos dados pelas fórmulas (2.22) a (2.27), apenas que são usadas os parâmetros posteriores  $\bar{z}^n$ ,  $n^n$ ,  $v^n$  e  $s^n$ .

Os parâmetros posteriores  $n^n$  e  $v^n$  podem ser considerados como o número de observações equivalentes total disponível, referentes respectivamente a  $\mu$  e  $\sigma$ . Quando não existem informações amostrais, ( $n = 0$ ,  $v = 0$ ) os parâmetros posteriores são iguais aos priorísticos. Quando não existem informações amostrais ( $n' = 0$ ,  $v' = 0$ ) os parâmetros posteriores são iguais às estatísticas suficientes.

### II.3.5 Distribuição Bayesiana de Vazões

A função densidade de probabilidade bayesiana de uma vazão futura dada por:

$$\tilde{f}_{q_f}(q_f | \bar{z}^n, s^n, v^n, n^n) = \int_0^{+\infty} \int_{-\infty}^{+\infty} f_N(q_f | \sigma; \mu) \cdot f''_N(\mu | \bar{z}^n, \frac{\sigma}{\sqrt{n^n}}) \cdot f''_{IG2}(\sigma | s^n, v^n) \delta\mu \delta\sigma \quad (2.37)$$

é uma student generalizada de momentos:

$$E(q_f | Ir, z) = \bar{z}^n \quad (2.38)$$

$$V(q_f | Ir, z) = s''^2 \frac{v''}{(v''-2)} \cdot \frac{n''+1}{n''} \quad (2.39)$$

Como esta distribuição de probabilidade leva em conta a incerteza restante nos valores de  $\mu$  e  $\sigma$ , a incerteza numa estimativa da próxima vazão, dada pela variância em (2.39), é maior que a obtida de uma análise que usasse estimativas. Seguindo a abordagem bayesiana apresentada no item II.2.4, a estimativa ótima de  $\sigma^2$  seria,  $s''^2 \frac{v''}{(v''-2)}$ , portanto menor que a variância dada por (2.39).

#### II.4 SUMÁRIO

Este capítulo indica o método bayesiano como uma ferramenta útil em problemas de inferência e estimação de parâmetros a partir de informações amostrais e não amostrais. A estrutura geral do método é apresentada definindo os conceitos de verossimilhança, distribuição "a priori" e posterior, distribuição bayesiana e função prejuízo. É mostrado como os conceitos de kernels e estatísticas suficientes levam a uma classe especial de distribuições "a priori" que simplificam os cálculos, chamadas distribuições conjugada naturais. Finalmente, são apresentados resultados úteis na análise bayesiana de processos geradores de variáveis aleatórias, independentes, de distribuição normal.

CAPÍTULO IIIA DISTRIBUIÇÃO "A PRIORI" OBTIDA  
POR REGRESSÃO MULTIVARIADA LINEARIII.1 INTRODUÇÃO

Frequentemente hidrólogos procuram relacionar funcionalmente parâmetros hidrológicos com características físico-meteorológicas através de equações simples, válidas apenas para regiões definidas.

Neste capítulo, o uso de modelos de predição, obtidos por regressão multivariada linear, são apresentados como uma metodologia para a construção de distribuições "a priori" dos parâmetros de séries temporais de vazões anuais.

III.2 CONCEITUAÇÃO

As variáveis que fazem parte de modelos de regressão multivariada linear se classificam em variáveis independentes e variáveis dependentes ou variáveis-resposta. O objetivo da análise é conhecer o efeito nas variáveis-resposta, causado por alterações nas variáveis independentes (ou em subconjuntos delas) de modo que se possa fazer previsões com nível de incerteza conhecido.

O modelo proposto é:

$$Y = X\beta + \epsilon \quad (3.1)$$

onde,

$Y$  = vetor ( $n \times 1$ ) contendo  $n$  observações da variável-resposta.

$\beta$  = vetor ( $p \times 1$ ) de coeficientes, desconhecido a priori e cuja estimativa denotaremos  $b$ .

$X$  = matriz ( $n \times p$ ) contendo  $n$  observações das  $p$  variáveis independentes consideradas.

$\varepsilon$  = vetor ( $n \times 1$ ) contendo as  $n$  perturbações, desconhecidas a priori cuja estimativa denotaremos  $\underline{e}$ .

A inclusão do vetor de perturbações na equação (3.1) é devida a erros de medição, à inadequacidade do modelo e ou, à aleatoriedade intrínseca da natureza.

Propõe-se então:

$$E(\varepsilon) = 0 \quad (3.2)$$

$$V(\varepsilon) = I\sigma^2 \quad (3.3)$$

Sendo  $I$  a matriz identidade ( $n \times n$ ), e  $\sigma^2$  a variância de cada componente de  $\varepsilon$ , desconhecida a priori, cuja estimativa denotaremos  $s_r^2$ .

O estimador  $b$ , do vetor  $\beta$  que minimiza  $\sum(e_i^2)$ , é chamada de estimador de mínimos quadrados, sendo dado por:

$$b = (X^T X)^{-1} X^T Y \quad (3.4)$$

Se o modelo é correto, tem-se:

$$E(b) = \beta \quad (3.5)$$

$$V(b) = (X^T X)^{-1} \cdot \sigma^2 \quad (3.6)$$

Neste caso,  $b$  é o estimador linear, não tendencioso e de menor variância de  $\beta$ . O estimador não-tendenciosos de  $\sigma^2$  é dado por:

$$s_r^2 = \frac{e^T e}{n-p} \quad (3.7)$$

Uma estimativa para a variável dependente associada aos valores das variáveis independentes de um vetor ( $l \times p$ ),  $X_o$  é dada por:

$$\hat{Y}_o = X_o b \quad (3.8)$$

De momentos, dadas pelas fórmulas:

$$E(\hat{Y}_O) = X_O \beta \quad (3.9)$$

$$V(\hat{Y}_O) = X_O (X^T X)^{-1} X_O^T \sigma^2 \quad (3.10)$$

que podem ser usadas substituindo-se  $\beta$  e  $\sigma^2$  por  $b$  e  $s_r^2$ , dadas por (3.4) e (3.7). Quando se deseja levar em conta a incerteza em  $\sigma^2$ , deve-se ainda multiplicar  $s_r^2$  pelo fator  $\frac{n-p}{n-p-2}$ .

Com a hipótese de normalidade das perturbações, é possível construir intervalos de confiança para os coeficientes e testar hipótese de nulidade. Caso se tenha conhecimento do valor de  $\sigma^2$ , pode se testar o ajuste dos dados ao modelo assumido, comparando  $\sigma^2$  com  $s_r^2$ .

Pode acontecer que o sistema de equações (3.4) se forme com equações dependentes entre si fazendo com que  $X^T X$  seja singular. Neste caso deve-se usar um número menor de variáveis independentes no modelo.

O modelo proposto pode ser modificado de forma a incluir diferentes hipóteses sobre a matriz de covariância das perturbações. A equação (3.3) generalizada é escrita:

$$V(\varepsilon) = V \sigma^2 \quad (3.11)$$

onde  $\sigma^2$  é um escalar desconhecido e  $V$  uma matriz ( $n \times n$ ) conhecida. A equação (3.11) permite modelar correlações e heteroscedâcia nas observações.

O estimador linear, não-tendencioso, de menor variância de  $\beta$ , passa a ser dado por:

$$b = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \quad (3.12)$$

cujas matrizes de média e covariância são:

$$E(b) = \beta \quad (3.13)$$

$$V(b) = (X^T V^{-1} X)^{-1} \sigma^2 \quad (3.14)$$

O estimador não-tendencioso de  $\sigma^2$  é então:

$$s_r^2 = \frac{\mathbf{e}^T \mathbf{V}^{-1} \mathbf{e}}{n-p} \quad (3.15)$$

As equações (3.8) e (3.9) continuam válidas e a equação (3.10) passa a ser escrita :

$$\mathbf{V}(\hat{\mathbf{Y}}_o) = \mathbf{x}_o (\mathbf{x}^T \mathbf{V}^{-1} \mathbf{x})^{-1} \mathbf{x}_o^T \sigma^2 \quad (3.16)$$

### III.3 A ESCOLHA DO MODELO

A fase menos objetiva, por isso mais sujeita a controversias, da aplicação de modelos multivariados lineares de regressão é a escolha do modelo.

Escolhido o parâmetro hidrológico que atuará como variável-resposta (exemplo: variância do escoamento médio anual) é necessário que se faça uma prévia avaliação das variáveis fí-sico-meteorológicas capazes de influirem no comportamento hidrológico. Mesmo quando o número de variáveis independentes pré-selecionadas for pequeno, a quantidade de modelos aplicáveis é bem grande se considerarmos transformações que podem ser aplicada nos dados.

Um dos procedimentos estatísticos para a definição das variáveis independentes que devem ser incluídas numa equação de regressão é o "Stepwise Regression Method", descrito adiante.

Seja  $Y$  a variável-resposta e  $X_1, X_2, \dots, X_m$  as variáveis candidatas a participar da equação de predição para  $Y$ . Estão incluídas todas as transformações que, se espera, sejam importantes. O método começa ajustando uma equação do tipo  $\hat{Y} = f(X_j)$ , onde  $X_j$  é a variável que tem a maior correlação com  $Y$ . Em seguida são ajustadas, sequencialmente, equações da forma  $\hat{Y} = f(X_j, X_k)$ ,  $\hat{Y} = f(X_j, X_k, X_e)$ , sendo que a variável escolhida para entrar na equação num certo passo é aquela de maior correlação parcial com  $Y$ , dado as variáveis independentes que já estão na equação.

Assim  $X_k$  é a variável de maior correlação parcial com  $Y$ , dado  $X_j$ . A cada passo testa-se a significância de todos os coeficientes pelo teste F geral e a significância de cada coeficiente pelo seu correspondente teste F parcial. No teste F geral a hipótese nula é:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0.0 \quad (3.17)$$

sendo  $p$  o número de variáveis independentes na equação. A estatística do teste é calculada pelo quociente, entre a variância explicada pelas  $p$  variáveis na equação,  $MSR$ , e a variância residual,  $s_r^2$ .

$$F_g^* = \frac{MSR}{s_r^2} \quad (3.18)$$

$MSR$  é dado por:

$$MSR = (b^T X^T Y) / p \quad (3.19)$$

Se  $F_g^*$  é maior que  $F(p, n-p, 1-\alpha)$  obtido de tabelas da distribuição F, então a hipótese nula (3.17) pode ser rejeitada com nível de confiança de  $(1-\alpha)$ .

Nos testes F parciais a hipótese nula é:

$$H_0 : \beta_q = 0 \quad (3.20)$$

e a estatística do teste é dado por:

$$F_q^* = \frac{SS(b_q | b_j, b_k, \dots)}{s_r^2} \quad (3.21)$$

onde,  $SS(b_q | b_j, b_k, \dots)$  é a variância explicada pela variável  $q$ , calculada pela diferença entre os numeradores da equação (3.19) para a regressão que inclua a variável  $q$  e para aquela que não

inclua a variável  $q$ . Se  $F_q^*$  é maior que  $F(1, n-p, 1-\alpha)$ , obtido de tabelas da distribuição F, então a hipótese nula (3.20) pode ser rejeitada com nível de confiança de  $(1-\alpha)$ .

Caso algum coeficiente torne-se não significativo a variável correspondente é retirada da regressão. O processo termina quando nenhuma variável pode ser adicionada ou removida da regressão.

Em hidrologia a aplicação de transformações logarítmicas nos dados é feita com o objetivo de se obter um certo tipo de comportamento dos resíduos, por exemplo homoscedáscia.

Estes modelos são escritos da mesma forma que (3.1) definindo as variáveis por:

$Y$  = vetor ( $n \times 1$ ) contendo  $n$  observações do logaritmo da variável-resposta  $Z$ .

$X$  = matriz ( $n \times p$ ) contendo  $n$  observações do logaritmo das  $p$  variáveis independentes  $W$ .

Em unidades originais o modelo é definido por:

$$Z = W_1^{\beta_1} W_2^{\beta_2} W_3^{\beta_3} \dots W_p^{\beta_p} \dots \gamma \quad (3.22)$$

sendo  $\gamma$  uma perturbação de distribuição log-normal com média  $e^{\sigma^2/2}$  e variância  $e^{\sigma^2}(e^{\sigma^2} - 1)$ . Os resultados obtidos da análise do modelo linear podem ser transferidos para as unidades originais através das fórmulas:

$$E(\hat{Z}_O) = \exp(\hat{Y}_O + 0,5 \cdot \sigma^2) \quad (3.23)$$

$$V(\hat{Z}_O) = \exp(\sigma^2 + 2\hat{Y}_O + V[\hat{Y}_O]) \cdot \exp(V[\hat{Y}_O]-1) \quad (3.24)$$

Neste trabalho o processo adotado para a escolha de um modelo foi feito por etapas:

- (i) Escolha das variáveis independentes através do "stepwise regression method" admitindo comportamento

simplificado das perturbações.

- (ii) Avançar hipótese mais realísticas na matriz de covariância das perturbações.

O nível de significância adotado para os testes F foi de 95%.

Frisa-se que os estimadores dados por (3.9) são tendenciosos quando o modelo proposto não é o verdadeiro. Quando, apesar do modelo ser o correto, a hipótese sobre a matriz de covariância das perturbações não é real, as estimativas são não-tendenciosas, porém não são mais de menor variância.

### III.4. APLICAÇÃO

O objetivo da aplicação é obter a partir do arquivo da malha quadrado do PHCA e dos dados hidro-meteorológicos registrados até 1974, equações que relacionam características fisiográficas e meteorológicas com a média e a variância anual da vazão líquida na bacia do Tocantins. Não fazem parte desta aplicação os dados sobre tipo de solo, vegetação e geologia superficial por que segundo o PHCA<sup>12</sup> estas informações não são seguras e produzem resultados pouco confiáveis.

Preliminarmente é ajustado um modelo de predição para a média da precipitação anual pontual na área do estudo.

#### III.4.1 Precipitação

Os postos pluviométricos usados (Figura A.3, pg.69) estão apresentados na tabela (3.1) com o código da estação, localização em relação ao referencial da malha quadrada, e período coletado. Cada estação foi associada às características físicas do quadrado correspondente à sua localização. A escolha das variáveis foi feita pelo "stepwise regression method" admitindo independência e variância inversamente proporcional ao número de anos de precipitação medidos. A matriz de variância adotada para  $\epsilon$  foi:

$$V(\varepsilon) = \begin{bmatrix} \frac{1}{n_1} & 0 & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \dots & \frac{1}{n_{50}} \end{bmatrix} \cdot \sigma^2 \quad (3.35)$$

sendo  $n_i$  o número de anos medidos na estação  $i$ , e  $\sigma^2$  a variância da precipitação anual assumida igual para todas as estações. A média ponderada pelo comprimento do registro das variâncias amostrais de cada estação foi comparada com o valor de  $s_r^2$  dos modelos ajustados. As equações de predições resultantes foram:

$$\text{MODELO 1} \quad (s_r^2/\hat{\sigma}^2 = 1.41)$$

$$P = 1655.25 - 0.757 B_n - 0.223 Be + 0.999 S \quad (3.26)$$

$$\text{MODELO 2} \quad (s_r^2/\hat{\sigma}^2 = 1.34)$$

$$P = 2857.46 \text{ ALT}^{0.381} \text{ En}^{-0.378} \text{ Be}^{-0.021} \quad (3.27)$$

sendo  $P$ , média da precipitação anual em mm/ano e  $\hat{\sigma}^2$ , média ponderada das variâncias amostrais (para o modelo 2,  $\hat{\sigma}^2$  é a média ponderada das variâncias dos logarítmicos da chuva anual). Os outros símbolos são definidos no anexo 1.

O modelo 2 foi escolhido por ter um valor de  $s_r^2/\hat{\sigma}^2$  mais próximo da unidade.

A equação (3.27) foi usada para se obter estimativa da média da precipitação anual em todos os quadrados do arquivo da malha quadrada. A figura (3.1) apresenta o mapa de precipitação resultante. A precipitação anual em cada bacia na região foi calculada com a média dos valores obtidos por (3.27) em todos os quadrados da bacia correspondente.

TABELA 3.1: - ESTAÇÕES PLUVIOMÉTRICAS

ESTAÇÃO	CÓDIGO	COLUNA	LINHA	NÚMERO DE ANOS	PERÍODO
Alto Paraíso	1000	46	30	4	70-73
Alvorada	1005	30	46	2	72-73
Alv. do Norte	1010	55	27	3	69, 72-73
Campinacu	1015	36	34	4	68-69, 72-73
Carolina	1020	45	97	25	49-73
Cavalcante	1025	46	33	5	69-73
Ceres	1030	25	18	3	64, 72-73
Colinas	1035	40	31	5	69-73
Colinas de Goiás	1040	36	91	2	72-73
Damianópolis	1045	59	26	5	69-73
Dianópolis	1050	53	54	2	72-73
Estrela do Norte	1055	31	33	2	72-73
Fátima	1060	31	63	1	72
Faz. Ingazeiro	1065	56	32	2	72-73
Faz. Lobeira	1070	37	55	2	72-73
Flores de Goiás	1075	46	25	3	69, 72-73
Formosa	1080	47	16	25	49-73
Goiânia	1085	30	18	3	71-73
Goiatins	1090	48	93	2	72-73
Gurupi	1095	31	53	2	72-73
Itaguatins	1100	46	113	3	69, 72-73
Itapuranga	1105	21	15	5	69-73
Mir. do Norte	1110	37	75	3	69, 72-73
M. Alegre	1115	52	40	3	69, 72-73
Niquelândia	1120	36	26	5	69-73
Nova-Roma	1125	52	32	2	72-73
Novo Acordo	1130	43	70	2	72-73
Paraíso do Norte	1135	32	69	2	72-73
Paraná	1140	44	42	25	49-73
Pedro Afonso	1145	39	81	9	49-57
Peixe	1150	35	50	4	70-73
Petrolina	1155	28	10	1	64
Pirenópolis	1160	31	13	10	49-52, 55, 69-73
Ponte Paraná	1165	48	36	3	69, 72-73
Ponte Quebralinha	1170	32	21	3	69, 72-73
Porto Gilândia	1175	43	63	3	69, 72-73
Porto Nacional	1180	36	63	25	49-73
Porto Real	1185	41	79	5	69-73
Porto Rio Bagagem	1190	38	27	2	72-73
Porto Uruaçu	1195	29	25	6	64, 69-73
Rubintaba	1200	23	19	1	64
Sama	1205	38	36	3	69, 72-73
Santa Sé	1210	46	17	3	69, 72-73
São Domingos	1215	58	37	3	69, 72-73
São Félix	1220	40	34	2	70-71
S. João da Aliança	1225	46	24	3	71-73
Taguatinga	1230	57	48	22	49-59, 61-70, 73
Tocantinópolis	1235	46	107	5	69-73
Tupirantins	1240	39	87	3	69, 72-73
Uruanã	1245	23	15	3	71-73

Obs.: Os totais anuais foram calculados para o ano hidrológico que inicia em Outubro.

FIGURA 3.1 - PRECIPITAÇÃO ANUAL (mm) NA BACIA DE ITAGUATINS

P = 2857,46 ALT<sup>0,381</sup> En<sup>-0,378</sup> Be<sup>-0,021</sup>

### III.4.2 A Escolha das Variáveis

Foram usados dados de 24 estações fluviométricas (Figura A.2, pag. 68) listadas na tabela (3.2). As possíveis variáveis independentes consideradas foram: a área de drenagem em  $\text{km}^2$  (Ad), a precipitação média anual na bacia (PA), e as características físicas do arquivo do PHCA, calculadas para cada bacia pela média da característica nos quadrados formadores de sua área de drenagem. O "stepwise regression method", aplicado sob a hipótese de independência e homoscedácia, escolheu para prever a variância, as variáveis S e Ad, e para prever a média as variáveis Be, Ds, En, AZ. Note-se que a precipitação, obtida no item anterior, foi rejeitada nos dois casos.

### III.4.3 Hipótese A: Independência Espacial

Mantida a hipótese de independência espacial, a matriz de covariância dos resíduos continua como uma matriz onde os elementos fora da diagonal principal são nulos. Cada elemento  $i$  da diagonal principal deve ser proporcional a variância da  $i$ -ésima observação.

Para a variância, a matriz V da equação (3.11) se escreve como:

$$V_{\sigma^2} = \begin{bmatrix} \frac{2\sigma_1^4}{n_1} & 0 & \dots & 0 \\ 0 & \frac{2\sigma_2^4}{n_2} & \dots & 0 \\ 0 & \dots & \frac{2\sigma_i^4}{n_i} & \dots & 0 \\ 0 & \dots & \dots & \frac{2\sigma_N^4}{n_N} & \end{bmatrix} \quad (3.28)$$

onde  $n_i$  é o tamanho do registro no posto  $i$ .

TABELA 3.2: - ESTAÇÕES FLUVIOMÉTRICAS

ESTAÇÃO	CÓDIGO	COLUNA	LINHA	NÚMERO DE ANOS	PERÍODO
P.Quebra Linha	2005	33	20	3	67-68, 71
Jaraguá	2010	27	13	2	71, 72
Uruanã	2020	23	15	4	67-68, 71-72
Ceres	2025	25	18	2	71, 72
Porto Uruaçu	2050	30	25	4	67-68-71-72
P.Rio Bagagem	2070	38	27	2	71-72
Tocantinzinho	2090	39	28	1	72
São Félix	2105	39	36	7	61-66, 71
Nova Roma	2150	52	32	1	72
Ponte Paraná	2160	48	36	2	71-72
Barra Palma	2189	42	44	1	72
Paranã	2190	42	44	1	72
Peixe	2205	35	50	2	71-72
Jacinto	2215	34	51	1	72
Faz.Lobeira	2225	37	55	2	71-72
Porto Nacional	2235	36	63	24	49-72
Mir. do Norte	2250	36	75	4	69-72
Porto Gilândia	2260	43	63	4	69-72
Novo Acordo	2270	43	70	1	72
Porto Real	2290	40	79	4	69-72
Tupirantins	2310	39	87	4	69-72
Goiatins	2325	48	93	1	72
Tocantinopolos	2360	46	107	18	55-72
Itaguatins	2370	46	113	4	69-72

Obs.: As estatísticas anuais foram calculadas para o ano hidrológico, considerado começando em Outubro. As estações com 1 ano hidrológico não foram usadas no modelo para a variância anual.

Como os valores na diagonal da matriz em (3.28) são desconhecidos deve-se aplicar a equação (3.12) de forma iterativa até que os valores obtidos em (3.8) para  $\hat{\sigma}_i^2$   $i = 1, \dots, N$ , converjam para os valores usados em (3.28).

Para a média a matriz  $V$  na equação (3.11) se escreve como:

$$V_\mu = \begin{bmatrix} \frac{\sigma_1^2}{n_1} & \dots & 0 & \dots & \dots & 0 \\ 0 & \frac{\sigma_2^2}{n_2} & 0 & \dots & \dots & 0 \\ 0 & \dots & \frac{\sigma_i^2}{n_i} & \dots & \dots & 0 \\ 0 & \dots & \dots & \frac{\sigma_N^2}{n_N} & \dots & \dots \end{bmatrix} \quad (3.29)$$

Os valores de  $\sigma_i^2$  na diagonal de (3.29) devem ser aqueles obtidos pela equação resultante no processo iterativo do modelo para a variância.

Pode-se notar que as matrizes  $V_\sigma^2$  e  $V_\mu$  definidas acima levam em conta apenas os erros de origem amostral. Neste caso, um valor obtido por (3.15), para  $s_r^2$  muito maior que a unidade indica uma fonte adicional para a aleatoriedade dos resíduos não explicada pelo modelo. Um valor de  $s_r^2$  menor que a unidade deve ser encarado como variação amostral.

Na aplicação do processo iterativo do modelo da variância, com as variáveis independentes  $S$  e  $Ad$ , não foi alcançada a convergência em 500 iterações. Abandonou-se então a variável  $S$  (de menor correlação com a variância) e obteve-se a equação de predição.

$$\sigma_q^2 = 5956.24 + 0.0501 Ad \quad (3.30)$$

onde  $\sigma_q^2$  é a variância do escoamento anual em mm. O valor obtido para  $s_r^2$  foi 1.48. Levando-se em conta a incerteza no parâme-

tro  $\sigma^2$ , (3.9) e (3.16) fornecem para Carolina:

$$E(\sigma_q^2) = 19\ 822 \quad (3.31)$$

$$V(\sigma_q^2) = 28\ 286\ 908 \quad (3.32)$$

Para a média, obteve-se a equação de predição:

$$\mu_q = -1043.58 - 0.42Be + 0.73Ds + 0.1En + 2.4Az \quad (3.33)$$

onde  $\mu_q$  é a média do escoamento anual em mm.

O valor obtido para  $s_r^2$  foi 0.826. Neste caso usaremos  $s_r^2 = 1$  e a correção da incerteza de  $\sigma^2$  em (3.16). Para Carolina, o modelo fornece:

$$E(\mu_q) = 411 \quad (3.34)$$

$$V(\mu_q) = 211 \quad (3.35)$$

Para se obter os momentos de  $\sigma_x^2$  e  $\mu_x$ , respectivamente variância e média da vazão anual em  $m^3/s$ , as fórmulas são:

$$E(\sigma_x^2) = A_d^2 \cdot K^2 \cdot E(\sigma_q^2) \quad (3.36)$$

$$V(\sigma_x^2) = A_d^4 \cdot K^4 \cdot V(\sigma_q^2) \quad (3.37)$$

$$E(\mu_x) = A_d \cdot K \cdot E(\mu_q) \quad (3.38)$$

$$V(\mu_x) = A_d^2 \cdot K^2 \cdot V(\mu_q) \quad (3.39)$$

onde,  $K = 3.17097 \times 10^{-5}$  é a constante de conversão de unidades. Logo para Carolina:

$$E(\sigma_x^2) = 1528697 \quad (3.40)$$

$$V(\sigma_x^2) = 1.68122 \times 10^{11} \quad (3.41)$$

$$E(\mu_x) = 3610 \quad (3.42)$$

$$V(\mu_x) = 16246 \quad (3.43)$$

Os parâmetros da distribuição "a priori" da média e variância da vazão anual, obtida substituindo os resultados acima em (2.28) a (2.31), são:

$$\bar{z}'_A = 3610 \quad (3.44)$$

$$n'_A = 94 \quad (3.45)$$

$$v'_A = 27,8 \quad (3.46)$$

$$s'_A = 1418731 \quad (3.47)$$

### III.4.4 Hipótese B: Correlação Espacial

A hipótese da existência de correlação entre os resíduos é incluída no modelo considerando as matrizes  $V_{\sigma^2}$  e  $V_{\mu}$  formada pelos elementos  $v_{\sigma^2}(i,j)$  e  $v_{\mu}(i,j)$ :

$$v_{\sigma^2}(i,j) = r_{\sigma^2}(i,j) \cdot 2 \frac{\sigma_i^2 \sigma_j^2}{(n_i \cdot n_j)^{0,5}} \quad (3.48)$$

$$v_{\mu}(i,j) = r_{\mu}(i,j) \cdot \frac{\sigma_i \cdot \sigma_j}{(n_i \cdot n_j)^{0,5}} \quad (3.49)$$

sendo  $r_{\sigma^2}(i,j)$  e  $r_{\mu}(i,j)$ , respectivamente, a correlação entre as variâncias amostrais do posto i e do posto j e a correlação entre as médias amostrais do posto i e do posto j calculadas por:

$$r_{\sigma^2}(i,j) = \frac{n}{(n_i \cdot n_j)^{0,5}} \cdot r^2(i,j) \quad (3.50)$$

$$r_{\mu}(i,j) = \frac{n}{(n_i \cdot n_j)^{0,5}} \cdot r(i,j) \quad (3.51)$$

onde n é o número de anos em comum do registro do posto i e do posto j, e  $r(i,j)$  é a correlação entre as vazões anuais do posto i e j.

A equação (3.50) e (3.51) estão justificadas no anexo 2.

Os coeficientes de correlação entre vazões anuais de dois postos foram calculados pela fórmula:

$$r(d) = 0.9681 e^{-0.0014d} + 0.0319 \quad (3.52)$$

onde  $d$  é a distância entre os postos em km. Esta equação foi obtida de um estudo, feito no Departamento de Sistemas do CEPEL, de postos fluviométricos das regiões Sul e Sudeste do Brasil. A figura (3.2) ilustra o ajuste obtido pela equação (3.52).

Novamente, o ajuste do modelo para a variância deve ser feito de forma iterativa e o seu resultado usado na formulação do modelo para a média.

Como o processo iterativo da variância não convergiu em 500 iterações quando aplicado com as variáveis independentes  $S$  e  $Ad$ , abandonou-se a variável  $S$  e obteve-se a equação de predição:

$$\sigma_q^2 = 5955.2 + 0.0501 Ad, \quad (3.53)$$

praticamente idêntica a (3.30).

O valor obtido para  $s_r^2$  foi 1.48. Levando-se em conta a incerteza no parâmetro  $\sigma^2$ , (3.9) e (3.16) fornecem para Carolina:

$$E(\sigma_q^2) = 19829 \quad (3.54)$$

$$V(\sigma_q^2) = 28325238 \quad (3.55)$$

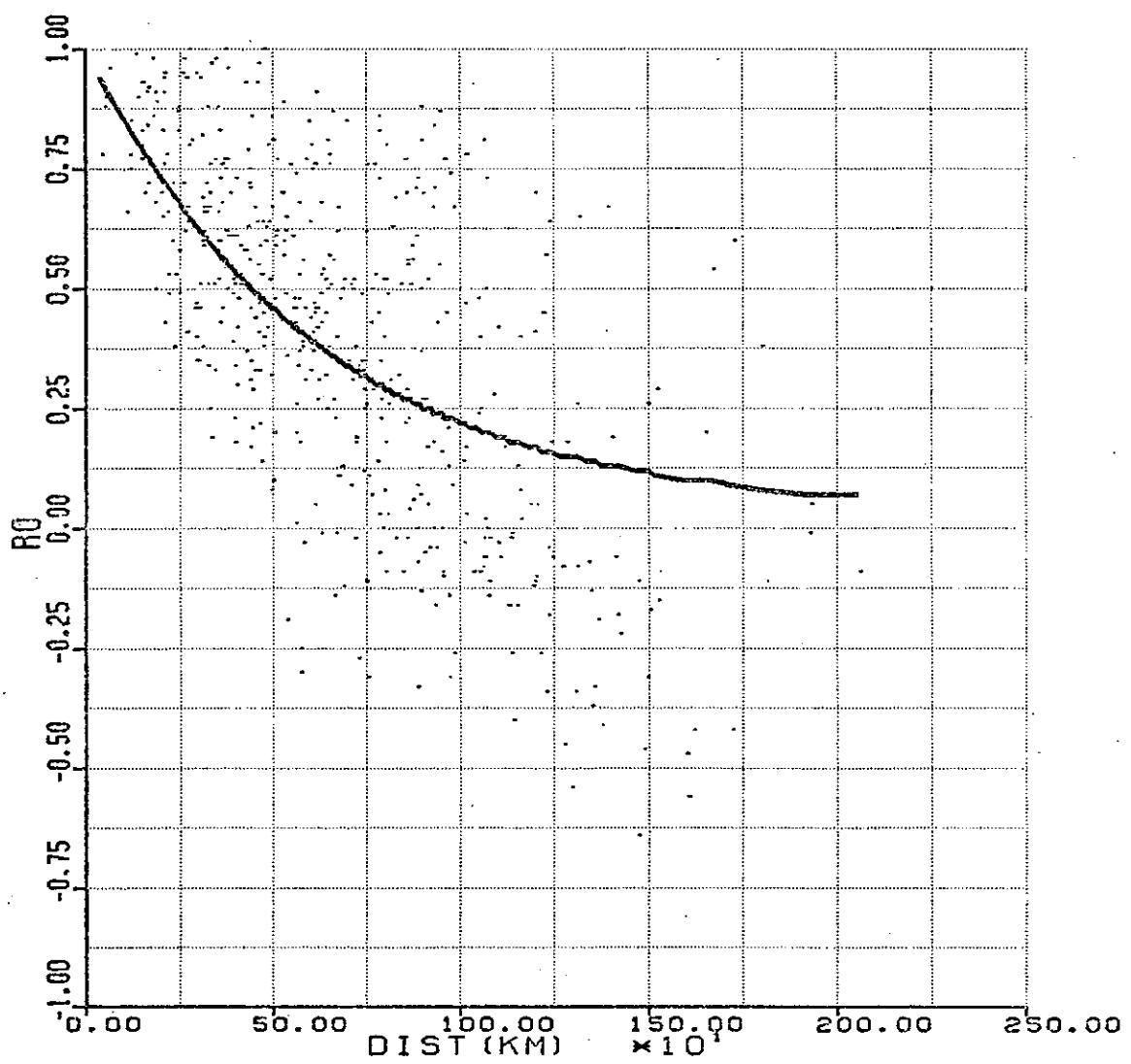


FIGURA 3.2 - CORRELAÇÃO ESPACIAL DE VAZÃO ANUAL  
NAS REGIÕES SUL E SUDESTE DO BRASIL

$$r(d) = 0,9681 e^{-0,0014d} + 0,0319$$

Para a média, obteve-se a equação de predição:

$$\mu_q = 790.7 - 0.54Be + 0.64Ds + 0.12En + 1.64AZ \quad (3.56)$$

O valor obtido para  $s_r^2$  foi 3.4, estatisticamente diferente de 1.00, indicando outra fonte de aleatoriedade além da amostral. Levando-se em conta a incerteza no parâmetro  $\sigma^2$ , (3.9) e (3.16) fornecem para Carolina:

$$E(\mu_q) = 375 \quad (3.57)$$

$$V(\mu_q) = 1358 \quad (3.58)$$

Os momentos da variância e média anual em  $m^3/s$  são dados por:

$$E(\sigma_x^2) = 1528697 \quad (3.59)$$

$$V(\sigma_x^2) = 1.6835 \times 10^{11} \quad (3.60)$$

$$E(\mu_x) = 3291 \quad (3.61)$$

$$V(\mu_x) = 104703 \quad (3.62)$$

Os parâmetros da distribuição "a priori" da média e variância da vazão anual, obtidos substituindo os resultados acima em (2.28) a (2.31) são:

$$\bar{z}_B' = 3291 \quad (3.63)$$

$$n_B' = 14.6 \quad (3.64)$$

$$v_B' = 27.8 \quad (3.65)$$

$$s_B'^2 = 1418719 \quad (3.66)$$

### III.4.5 Conclusão

Da tabela 3.3, que apresenta os parâmetros priorísticos obtidos pelos modelos A e B, pode-se concluir que o modelo A é mais otimista quanto ao conteúdo da informação para a média.

PARÂMETRO PRIORÍSTICO	MODELO A (INDEPENDÊNCIA ESPACIAL)	MODELO B (CORRELAÇÃO ESPACIAL)
$z'$	3610	3291
$n'$	94	14,6
$v'$	27,8	27,8
$s'^2$	1418731	1418719

TABELA 3.3 - COMPARAÇÃO DOS MODELOS

Como a existência de correlação espacial é mais razoável, recomenda-se o modelo B. Note-se que o valor de  $s_r^2 = 3.4$  para a equação da média, fornecido pelo modelo B, indica que a incerteza da média pode ser diminuída transferindo maior quantidade de informação. Isto é conseguido pelo modelo A às custas de uma hipótese pouco realista (independência espacial).

III .5 SUMÁRIO

Este capítulo apresenta a teoria necessária para a aplicação de modelos de regressão lineares multivariados na construção de distribuições "a priori". A metodologia foi aplicada na bacia de Itaguatins com o uso do arquivo fisiográfico desenvolvido pelo PHCA.

O estudo considerou duas hipóteses:

- a) As vazões anuais na região tem distribuição normal e são independentes serialmente e espacialmente.
- b) As vazões anuais na região tem distribuição normal, são independentes serialmente e seguem uma estrutura de correlação espacial função da distância.

Ao ajustar-se os modelos para a variância da vazão média anual verificou-se que, neste caso, os resultados não são sensíveis à estrutura de correlação espacial. Por outro lado, os resultados para a média mostraram uma diferença entre os valores preditos pelos dois modelos de 10% e o conteúdo de informação baixou de 94 anos (modelo A) para 14,6 anos (modelo B). O modelo B é considerado mais verossímil e portanto é recomendado.

Para que a informação regional possa ser usada para reduzir o nível de incerteza hidrológico em Carolina, foram calculados os parâmetros da distribuição "a priori" da média e da variância das vazões anuais.

Como subproduto, foi ajustado um modelo de previsão para a média da precipitação anual, cuja aplicação resultou no mapa apresentado na figura 3.1.

CAPÍTULO IVO USO DE UM MODELO CONCEITUALIV.1 INTRODUÇÃO

A equação da conservação de massa, aplicada a uma bacia hidrográfica, por unidade de área, é dada por:

$$\begin{array}{cccccc} \Delta t & & \Delta t & & \Delta t & \\ \int p(t) dt - \int e(t) dt = & & \int r(t) dt + \int \frac{\delta V(t)}{\delta t} dt & & & (4.1) \\ o & & o & & o & \end{array}$$

onde  $p(t)$  é a intensidade de precipitação,  $e(t)$  a taxa de evaporação,  $r(t)$  a vazão e  $V(t)$  o armazenamento. A simulação da equação (4.1) para intervalos curtos exige um grande esforço computacional devido a interdependência entre o armazenamento, a vazão básica e a evaporação. Para intervalos de tempo suficientemente longos a variação do volume armazenado pode ser desprezado.

O balanço hídrico de uma bacia é feita através das relações entre os valores médios de longo termo dos processos de evaporação, precipitação, infiltração, etc... Como em regiões com escassez de dados hidrológicos o elemento mais bem conhecido normalmente é a precipitação, o conhecimento das relações entre os diversos componentes do balanço hídrico e a precipitação tem uma maior importância prática. Em geral, estas relações são válidas para intervalos de tempo suficientemente longos de forma que as variações das condições da bacia, e eventualmente de sua forma de atuação, se anulem. Por exemplo, se considerarmos estas relações válidas anualmente, um registro de precipitações anuais podem ser transformado em um registro de vazões anuais e a distribuição de probabilidades da vazão anual pode ser obtida de uma distribuição de probabilidade de precipitação anual.

Como os sistemas naturais são por demais complexos, com características que variam no tempo e no espaço, a sua representação tem que ser simplificada. Em troca, quanto maior a simplificação do modelo conceitual adotado, maior a distância com a realidade impossibilitando medidas diretas dos parâmetros.

Em seguida é apresentada o modelo conceitual do balanço hídrico desenvolvido por S. Solomon<sup>13</sup> para regiões tropicais e equatoriais. O modelo é bastante simples e tem apenas um parâmetro: radiação solar absorvida. O sucesso deste modelo em regiões tropicais e equatoriais deve-se ao fato que a radiação solar absorvida praticamente não varia de ano para ano nestas regiões. A sua aplicação na bacia de Carolina servirá para se construir uma série sintética de vazões anuais a partir de dados pluviométricos na região.

#### IV.2 O MODELO

Nos processos dinâmicos que ocorrem dentro das bacias hidrográficas tem grande importância a relação entre a capacidade que o sistema tem de transferir a água do solo para a atmosfera, e a umidade disponível para evaporação. Grande parte dos modelos conceituais hidrológicos definem "evaporação potencial" como a evaporação real que aconteceria caso houvesse um constante fornecimento de umidade para o processo. Em situações nas quais a bacia esteja totalmente úmida esta seria a taxa de evaporação. Porém, quando restringe-se a evaporação, por exemplo por uma deficiência de umidade do solo, provoca-se um aumento da temperatura, maiores velocidades de vento e umidades de ar decrescentes com o tempo. Obviamente aumenta-se o valor da "evaporação potencial".

Por outro lado, alimentando-se continuamente a evaporação, aumenta a umidade do ar, diminui a velocidade dos ventos, reduzindo-se a "evaporação potencial". O princípio de Bouchet<sup>14</sup> diz que a "evaporação potencial" é igual a parcela que resta da radiação solar absorvida depois da ocorrência da evaporação sendo o fator de controle a umidade disponível. Este princípio po-

de ser escrito por:

$$W = (1-a) R_i \quad (4.2)$$

$$E_p = W - E \quad (4.3)$$

onde  $E_p$  é a evaporação potencial,  $a$  o albedo da superfície,  $R_i$  a incidência de energia solar,  $E$  a evaporação e  $W$  a radiação solar absorvida. O controle exercido pela umidade disponível na divisão da energia em evaporação e evaporação potencial é influenciado pelas características da bacia, tais como: declividade, tipo de vegetação, distribuição espaço-temporal da umidade disponível. Solomon propôs o uso da equação empírica semelhante a de Turc que se escreve.

$$E = \frac{P \cdot E_p}{(E_p^2 + P^2)^{0.5}} \quad (4.4)$$

onde  $P$  é a umidade disponível, ou seja, a precipitação quando se considera intervalos de tempos longos.

Combinando (4.3) e (4.4) obtém-se uma equação relacionando evaporação e precipitação.

$$P = \frac{E \cdot (W - E)}{(W(W - 2E))^{1/2}} \quad (4.5)$$

Conhecida a precipitação e evaporação facilmente calcula-se a vazão, como:

$$R = P - E \quad (4.6)$$

#### IV.3 ESTIMATIVA DA RADIAÇÃO SOLAR ABSORVIDA

Normalmente o número de estações climatológicas que medem insolação incidente e albedo é muito pequeno ou nulo. Por isso, tem-se desenvolvido relações entre as condições atmosféricas (exemplo: horas de sol) e a percentagem da insolação no topo da atmosfera que chega à superfície do solo. Uma das mais conhecidas é a equação de Penman:

das é a relação de Penman:

$$\frac{R_i}{R_a} = 0,18 + 0,55 \frac{S}{S_A} \quad (4.7)$$

onde  $R_a$  é a insolação no topo da atmosfera,  $S_A$  o número máximo de horas de sol e  $S$  o número real de horas de sol. Tanto  $R_a$  e  $S_a$  estão tabeladas em função da latitude (tabela(4.1) e (4.2)). O valor médio de  $S$  pode ser estimado de dados de estações climatológicas convencionais. Para a estimativa do albedo existem tabelas reportando medições para diversos tipos de superfície. Um exemplo é a tabela (4.3). Conforme pode-se notar existe grande variação nos valores medidos, o que dificulta as estimativas de albedo de bacias naturais.

Uma outra abordagem seria obter o valor da radiação solar absorvida a partir do registro de vazão e precipitação. Conforme Solomon<sup>13</sup>, 2 a 5 anos são suficientes para uma estimativa bem segura da radiação solar absorvida em zonas tropicais equatoriais.

A solução de (4.5) para  $W$  é:

$$W = E \left( 1 + \frac{P}{(P^2 - E^2)^{0,5}} \right) \quad (4.8)$$

#### IV.4 APLICAÇÃO

O arquivo de dados de precipitação usado (tabela (3.1)) abrange o período de 1949 a 1974. As medições de vazão em Carolina (fig.A.1,pg.67) cobrem os anos hidrológicos de 1962 a 1972. Os valores de precipitação na bacia de Carolina foram calculados através do arquivo da malha quadrada. Em cada ano, a precipitação de cada quadricula da bacia de Carolina foi estimada como a média das precipitações medidas, ponderadas pelo inverso da distância entre o quadrado e o local da medição. Usando os valores médios do período 62-72 de precipitação e evaporação (tabela (4.4)) em (4.8) obtém-se  $W = 2571 \text{ mm/ano}$ . Este valor corresponderia a

Lat	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
50°	17.5	14.7	10.9	7.0	4.2	3.1	3.5	5.5	8.9	12.9	16.5	18.2
48	17.6	14.9	11.2	7.5	4.7	3.5	4.0	6.0	9.3	13.2	16.6	18.2
46	17.7	15.1	11.5	7.9	5.2	4.0	4.4	6.5	9.7	13.4	16.7	18.3
44	17.8	15.3	11.9	8.4	5.7	4.4	4.9	6.9	10.2	13.7	16.7	18.3
42	17.8	15.5	12.2	8.8	6.1	4.9	5.4	7.4	10.6	14.0	16.8	18.3
40	17.9	15.7	12.5	9.2	6.6	5.3	5.9	7.9	11.0	14.2	16.9	18.3
38	17.9	15.8	12.8	9.6	7.1	5.8	6.3	8.3	11.4	14.4	17.0	18.3
36	17.9	16.0	13.2	10.1	7.5	6.3	6.8	8.8	11.7	14.6	17.0	18.2
34	17.8	16.1	13.5	10.5	8.0	6.8	7.2	9.2	12.0	14.9	17.1	18.2
32	17.8	16.2	13.8	10.9	8.5	7.3	7.7	9.6	12.4	15.1	17.2	18.1
30	17.8	16.4	14.0	11.3	8.9	7.8	8.1	10.1	12.7	15.3	17.3	18.1
28	17.7	16.4	14.3	11.6	9.3	8.2	8.6	10.4	13.0	15.4	17.2	17.9
26	17.6	16.4	14.4	12.0	9.7	8.7	9.1	10.9	13.2	15.5	17.2	17.8
24	17.5	16.5	14.6	12.3	10.2	9.1	9.5	11.2	13.4	15.6	17.1	17.7
22	17.4	16.5	14.8	12.6	10.6	9.6	10.0	11.6	13.7	15.7	17.0	17.5
20	17.3	16.5	15.0	13.0	11.0	10.0	10.4	12.0	13.9	15.8	17.0	17.4
18	17.1	16.5	15.1	13.2	11.4	10.4	10.8	12.3	14.1	15.8	16.8	17.1
16	16.9	16.4	15.2	13.5	11.7	10.8	11.2	12.6	14.3	15.8	16.7	16.8
14	16.7	16.4	15.3	13.7	12.1	11.2	11.6	12.9	14.5	15.8	16.5	16.6
12	16.6	16.3	15.4	14.0	12.5	11.6	12.0	13.2	14.7	15.8	16.4	16.5
10	16.4	16.3	15.5	14.2	12.8	12.0	12.4	13.5	14.8	15.9	16.2	16.2
8	16.1	16.1	15.5	14.4	13.1	12.4	12.7	13.7	14.9	15.8	16.0	16.0
6	15.8	16.0	15.6	14.7	13.4	12.8	13.1	14.0	15.0	15.7	15.8	15.7
4	15.5	15.8	15.6	14.9	13.8	13.2	13.4	14.3	15.1	15.6	15.5	15.4
2	15.3	15.7	15.7	15.1	14.1	13.5	13.7	14.5	15.2	15.5	15.3	15.1
0	15.0	15.5	15.7	15.3	14.4	13.9	14.1	14.8	15.3	15.4	15.1	14.8

Tabela 4.1 - INSOLAÇÃO NO TOPO DA ATMOSFERA EM mm/DIA

PARA O HEMISFÉRIO SUL.

(FAO, "Crop Water Requirements")

Lats	Jul	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun
50°	8.5	10.1	11.8	13.8	15.4	16.3	15.9	14.5	12.7	10.8	9.1	8.1
48	8.8	10.2	11.8	13.6	15.2	16.0	15.6	14.3	12.6	10.9	9.3	8.3
46	9.1	10.4	11.9	13.5	14.9	15.7	15.4	14.2	12.6	10.9	9.5	8.7
44	9.3	10.5	11.9	13.4	14.7	15.4	15.2	14.0	12.6	11.0	9.7	8.9
42	9.4	10.6	11.9	13.4	14.6	15.2	14.9	13.9	12.6	11.1	9.8	9.1
40	9.6	10.7	11.9	13.3	14.4	15.0	14.7	13.7	12.5	11.2	10.0	9.3
35	10.1	11.0	11.9	13.1	14.0	14.5	14.3	13.5	12.4	11.3	10.3	9.8
30	10.4	11.1	12.0	12.9	13.6	14.0	13.9	13.2	12.4	11.5	10.6	10.2
25	10.7	11.3	12.0	12.7	13.3	13.7	13.5	13.0	12.3	11.6	10.9	10.6
20	11.0	11.5	12.0	12.6	13.1	13.3	13.2	12.8	12.3	11.7	11.2	10.9
15	11.3	11.6	12.0	12.5	12.8	13.0	12.9	12.6	12.2	11.8	11.4	11.2
10	11.6	11.8	12.0	12.3	12.6	12.7	12.6	12.4	12.1	11.8	11.6	11.5
5	11.8	11.9	12.0	12.2	12.3	12.4	12.3	12.3	12.1	12.0	11.9	11.8
0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0	12.0

Tabela 4.2 - NÚMERO MÁXIMO POSSÍVEL DE HORAS DE SOL, PARA O HEMISFÉRIO SUL  
 (FAO, "Crop Water Requirements")

SUPERFÍCIE	ALBEDO	SUPERFÍCIE	ALBEDO
solos seco preto	0.14	areia fina brilhante	0.37
solos úmido preto	0.08	rocha	0.12 - 0.15
solos seco cinza	0.25 - 0.30	áreas urbanizadas	0.15 - 0.25
solos úmido cinza	0.10 - 0.12	neve	0.40 - 0.85
barro seco azul	0.23	água	0.03 - 0.40
barro úmido azul	0.16	grama alta e densa	0.18 - 0.20
barro de deserto	0.29 - 0.31	grama seca	0.19
areia amarela	0.35	grama verde	0.26
areia branca	0.34 - 0.40	campos de arroz	0.12
areia de rio	0.43	campos de batatas	0.19

Tabela 4.3 - ALBEDO DE SUPERFÍCIES NATURAIS  
 (Eagleson, "Dynamic Hydrology")

ANO	PRECIPITAÇÃO (mm)	VAZÃO (mm)	EVAPORAÇÃO (mm)
62-63	1404	341	1063
63-64	1502	472	1030
64-65	1401	440	961
65-66	1558	430	1128
66-67	1312	383	929
67-68	1606	436	1170
68-69	1393	421	972
69-70	1523	486	1037
70-71	1323	280	1043
71-72	1434	349	1085
72-73	1532	393	1139
MÉDIA	1453	403	1050

Tabela 4.4 - BALANÇO HÍDRICO ANUAL EM CAROLINA

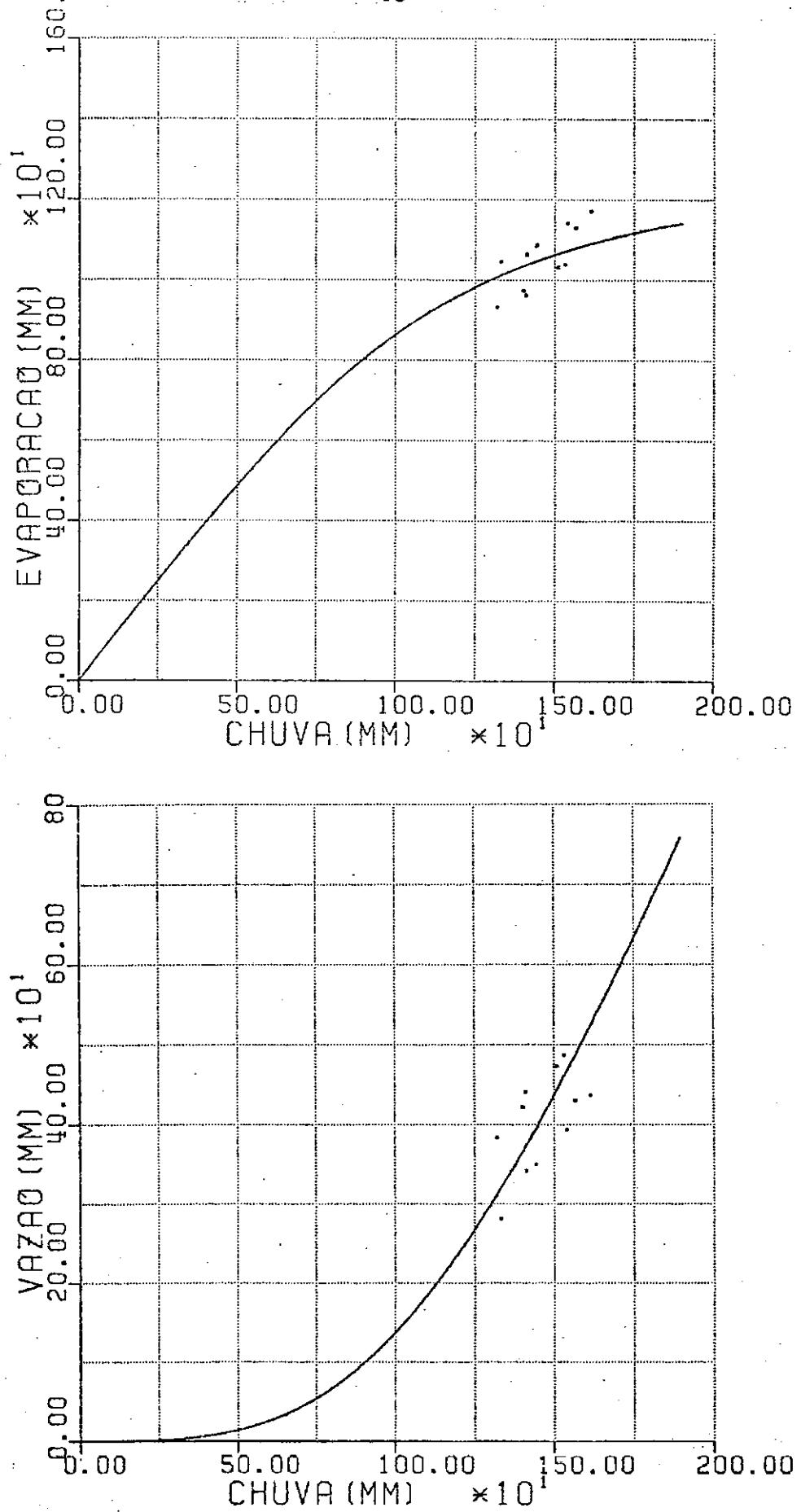


FIGURA 4.1 - BALANÇO HÍDRICO DA BACIA DE CAROLINA

ANO	PRECIPITAÇÃO	VAZÃO (mm)	VAZÃO M <sup>3</sup> /S
49-50	1419	378	3322
50-51	1027	242	2125
51-52	1331	319	2799
52-53	1299	298	2619
53-54	1233	257	2259
54-55	1143	206	1810
55-56	1465	411	3608
56-57	1783	659	5788
57-58	1411	373	3276
58-59	1402	367	3219
59-60	1463	410	3596
60-61	1244	264	2321
61-62	1285	289	2540
73-74	1708	598	5252

Tabela 4.5 - SIMULAÇÃO ANUAL EM CAROLINA  
 $W = 2571 \text{ mm/ano}$

um albedo de 0.12 para a bacia se aceitarmos  $R_i = 2918 \text{ mm/ano}$  obtido dos mapas de radiação solar incidente no Brasil publicado pelo Instituto de Pesquisas Espaciais<sup>15</sup>. Solomon<sup>13</sup> analisou dados equatoriais e tropicais de todo o mundo encontrando albedos entre 0.10 e 0.15.

A figura (4.1) apresenta as relações entre precipitação vazão e evaporação fornecidas pelo modelo para  $W=2571\text{mm/ano}$ . Este modelo permitiu transformar os dados de precipitação na bacia de Carolina na série de vazões anuais da tabela (4.5) que pode ser encarada como uma informação amostral. As estatísticas suficientes (ver capítulo 2), correspondente a série da tabela (4.5) estão apresentados na tabela (4.6).

As estatísticas  $n$  e  $v$ , que indicam o número de anos equivalentes de informação para  $\mu$  e  $\sigma$  respectivamente (ver capítulo 2), foram calculados por:

$$n = 0.34 N \quad (4.9)$$

$$v = 0.72 (N-1) \quad (4.10)$$

onde  $N$  é o comprimento da série e os fatores de redução 0.890 e 0.671 foram obtidos empiricamente comparando médias e variâncias amostrais de séries observadas e calculadas pelo modelo, para o posto de Porto Nacional localizado a montante de Carolina (ver anexo 3).

ESTATÍSTICAS	SÉRIE SIMULADA
$\bar{z}$	3181
$n$	4.76
$v$	9.36
$s^2$	1302 065

TABELA 4.6 - ESTATÍSTICAS SUFICIENTES DA SÉRIE SIMULADA

## IV.5

SUMÁRIO

Este capítulo define como balanço hídrico de uma região as relações entre os valores médios de longo termo dos processos continentais do ciclo hidrológico. É frisada a importância prática do conhecimento das relações entre a precipitação e os outros componentes do balanço hídrico em situações de escassez de dados hidrológicos, já que, na maioria das vezes, a precipitação é o elemento mais bem conhecido. Em seguida é mostrado o desenvolvimento de um modelo conceitual do balanço hídrico, devido a S.Solomón, que basicamente utiliza o princípio de Bouchet e a equação empírica de Turc. O princípio de Bouchet diz que a energia líquida disponível na bacia num certo período se repartirá em duas parcelas: evaporação real e evaporação potencial. A equação de Turc condiciona esta divisão ao valor da umidade disponível. A aplicação deste modelo na bacia de Carolina permitiu transformar um registro de precipitação numa série de variações anuais de 14 anos.

CAPÍTULO V

AVALIAÇÃO DA CAPACIDADE DE GERAÇÃO  
ENERGÉTICA DE UM APROVEITAMENTO HIDROELÉTRICO

v.1 INTRODUÇÃO

Este capítulo analisa o efeito do uso de informações regionais e meteorológicas, bem como da imprecisão existente nas estimativas de propriedades probabilísticas, na avaliação do mercado de energia capaz de ser suprido por um determinado aproveitamento hidroelétrico.

v.2 O MÉTODO

Seja um reservatório de volume útil igual a  $v_u$ . O estado do reservatório, a cada instante  $t$ , é caracterizado pelo nível de armazenamento  $v(t)$ . Considera-se  $K$  possíveis estados, definidos por:

ESTADO	VOLUME ARMAZENADO
1	$v(t) = 0$
:	:
i	$(i-2) \Delta v < v(t) < (i-1) \Delta v$
:	:
K	$v(t) = v_u$

sendo  $\Delta v = v_u / (K-2)$ . A regra de operação do reservatório é tal que a vazão turbinada deverá atender, sempre que possível, as necessidades energéticas definidas pelo mercado  $M$ . Em anos nos quais as disponibilidades de água armazenada mais afluência não são suficientes para satisfazer o mercado, o reservatório é esvaziado de forma a gerar o máximo de energia possível. Define-se como risco associado ao mercado  $M$ , a probabilidade do reser-

vatório ocupar o estado vazio, ou seja, de que a demanda energética não seja atendida.

A energia gerada pela usina durante um ano no qual o nível de armazenamento evolui do estado  $i$  para o estado  $j$  é calculada considerando-se uma altura de queda equivalente dada pela média das alturas dos estados  $i$  e  $j$ . As alturas de queda de cada estado são obtidas diretamente da curva cota  $x$  - volume do reservatório considerando-se o nível do canal de fuga e o rendimento constantes. A evaporação é levada em conta através de uma taxa anual fixa dada em mm/ano. Finalmente as vazões anuais afluentes ao reservatório se comportam como um processo estocástico independente e estacionário, com distribuição de probabilidades marginal normal  $f_x(x|\theta = (\mu_x, \sigma_x^2))$ . São disponíveis informações sobre os parâmetros em  $\theta$ .

Quando  $\theta$  é perfeitamente conhecido, é possível montar, para cada mercado  $M$  a matriz  $A(M)$  de dimensão  $K \times K$ , cujos elementos  $a_M(j,i)$  indicam a probabilidade de que se o estado do reservatório ao início de qualquer ano é  $i$ , o estado ao final do ano seja  $j$ . Seja o vetor  $\pi^T(M) = (\pi_1(M), \pi_2(M), \dots, \pi_K(M))$ , onde  $\pi_i(M)$  é a probabilidade de que em algum ano o reservatório ocupe o estado  $i$ , quando o mercado é  $M$ , e portanto  $\pi_1(M)$  indica o risco associado ao mercado  $M$ . Este vetor é a solução do sistema de equações:

$$\pi(M) = A(M) \pi(M) \quad (5.1)$$

$$\sum_{i=1}^K \pi_i(M) = 1 \quad (5.2)$$

O sistema (5.1) e (5.2) possue  $K + 1$  equações para  $K$  incógnitas. Para resolvê-lo substitui-se (5.1) por:

$$(A(M) - I) \pi(M) = 0 \quad (5.3)$$

onde,  $I$  é a matriz de identidade. Substituindo-se a primeira equação do sistema (5.3) por (5.2) obtém-se:

$$\pi(M) = (B(M))^{-1} C^T \quad (5.4)$$

onde,  $B(M)$  é a matriz  $(A(M) - I)$  com a primeira linha modificada para  $(1, 1, \dots)$  e  $C$  o vetor  $(1, 0, 0, \dots)$ .

A distribuição de probabilidades do estado do reservatório dada por (5.4) é uma aproximação discreta da distribuição contínua. A precisão é uma função do número de estados discretos. Doran<sup>16</sup> recomenda 5 a 10 estados como valores suficientes para se obter o risco de deficit para os casos usuais da prática. Neste trabalho foi adotada o procedimento de tornar a malha de discretização cada vez mais fina até que o resultado não fosse mais sensível a um maior detalhamento. Em alguns casos foi necessário usar 18 estados para que a variação de risco de deficit fosse menor que 0,001.

Quando  $\theta$  é desconhecido, estando disponível a distribuição posterior dos parâmetros  $f''(\theta)$ , o vetor  $\pi(M)$  é estimado tirando valor esperado em (5.4):

$$\hat{\pi}(M) = E[\pi(M)] = \int_{\theta} \pi(M|\theta) f''(\theta) d\theta \quad (5.5)$$

A resolução da integral em (5.5) pode ser feita pelo método de Monte-Carlo resolvendo-se (5.4) para um número grande de vetores de parâmetros  $\theta_i$ ,  $i = 1, \dots$ , gerados conforme  $f''(\theta)$  e computando-se a média dos valores de  $\pi(M)$  assim obtidos. A seguir, descreve-se como obter uma aproximação para (5.5), resolvendo-se apenas uma vez a equação (5.4).

O valor esperado em (5.4) pode ser escrito por:

$$\hat{\pi}(M) = E[(B(M))^{-1} C^T] = E[(B(M))^{-1}]C^T = DC^T$$

sendo D o valor esperado da inversa de  $B(M)$ . Se aceitarmos a aproximação de primeira ordem de D dada por:

$$D \approx (E[B(M)])^{-1} = (B'(M))^{-1} \quad (5.6)$$

sendo  $B'(M)$  obtida usando a distribuição bayesiana de uma va-  
zão futura (ver capítulo 2) para a construção de  $(A(M)-I)$  e  
substituindo a primeira linha por  $(1,1,1,\dots)$ , a equação (5.5) po-  
de ser escrita por:

$$\hat{\pi}_{(M)} = (B'(M))^{-1} C^T \quad (5.7)$$

Para este trabalho interessa saber se a aproximação de  
primeira ordem em (5.6) é suficiente para fornecer boas estima-  
tivas para o risco. Por enquanto, admite-se que a aproximação  
é boa, adiando-se para item (V.3.4) a constatação, deste fato.

### V.3 APLICAÇÃO

O aproveitamento hidroelétrico de Carolina está descri-  
to em Engevix<sup>1</sup>. As características que interessam a esta apli-  
cação são:

N.A	máximo do reservatório	-	197,5 m
N.A	normal a jusante	-	144,6 m
Volume utilizado no local		-	$36,88 \times 10^9 \text{ m}^3$
Taxa de Evaporação		-	1009 mm/ano
Rendimento		-	0,83
Área de Drenagem		-	276897 km <sup>2</sup>

além das curvas cota x volume x área. No relatório acima a e-  
nergia firme do aproveitamento de Carolina foi avaliada em  
952 Mw. Nesta aplicação, a capacidade de geração de Carolina  
será fixada pelo valor do mercado associado ao risco de 5%. Pa-  
ra alcançar os objetivos desta tese, a avaliação será feita con-  
siderando-se:

- (i) Somente as informações amostrais locais.

- (ii) Uso das informações fornecidas pelo modelo conceitual do balanço hídrico (capítulo 4).
- (iii) Uso das informações fornecidas pelos modelos de regressão (capítulo 3).

O fluxograma na figura (5.1) esclarece a manipulação das informações existentes.

#### V.3.1 Uso da Amostra de Vazões Medidas

Considerou-se o histórico de vazões ao posto de Carolina nos períodos de 1962 a 1972 apresentados em Engevix<sup>1</sup>. As estatísticas suficientes da série anual são dadas por:

$$\bar{z} = 3537 \quad (5.8)$$

$$n = 11 \quad (5.9)$$

$$s^2 = 290727 \quad (5.10)$$

$$v = 10 \quad (5.11)$$

Os momentos de  $\mu_x$  e  $\sigma_x^2$ , fornecidos pela amostra são dados usando-se as estatísticas acima como parâmetro de uma distribuição normal - gama invertida cujos momentos são fornecidos por (2.22) a (2.27). Logo:

$$E(\mu_x) = \bar{z} = 3537 \quad (5.12)$$

$$E(\sigma_x^2) = \frac{vs^2}{v-2} = 363409 \quad (5.13)$$

$$V(\mu_x) = \frac{vs^2}{n(v-2)} = 33037 \quad (5.14)$$

$$V(\sigma_x^2) = \frac{2v^2 s^4}{(v-2)^2 (v-4)} = 4,402 \times 10^{10} \quad (5.15)$$

A curva 1 da figura (5.2) mostra a relação entre o mercado e o risco fornecida pela equação (5.4) usando-se os valores em (5.12) e (5.13) como se fossem os verdadeiros. O mercado correspondente ao risco de 5% na curva 1 é 1175 Mw, ou seja, de acordo com as simplificações utilizadas até aqui e baseando-se nas estimativas de  $\mu_x$  e  $\sigma_x^2$  fornecidas pelo histórico de 11 anos de vazões afluentes ao local do aproveitamento, a capacidade de geração em Carolina é estimada em 1175 Mw.

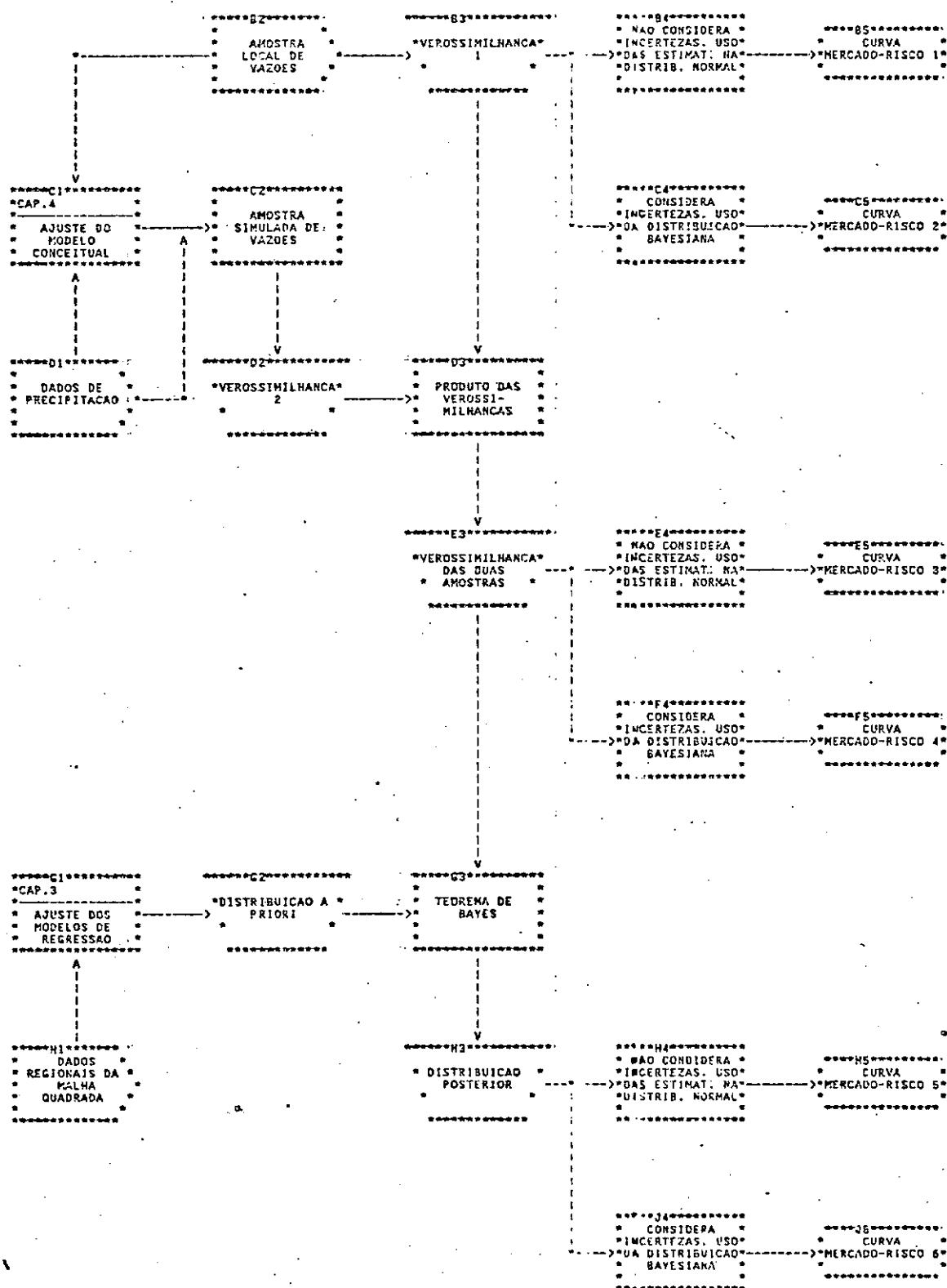


FIGURA 5.1- MANIPULAÇÃO DOS DADOS

A inclusão das incertezas nos parâmetros pode ser feita resolvendo a equação (5.7) através do uso da distribuição bayesiana de uma vazão futura dada por (2.37). Os momentos da distribuição bayesiana fornecida pela amostra são dados usando-se as estatísticas suficientes como parâmetros. Logo:

$$E [qf|z] = \bar{z} = 3537 \quad (5.16)$$

$$V [qf|z] = s^2 \frac{v}{(v-2)} \cdot \frac{n+1}{n} = 396446 \quad (5.17)$$

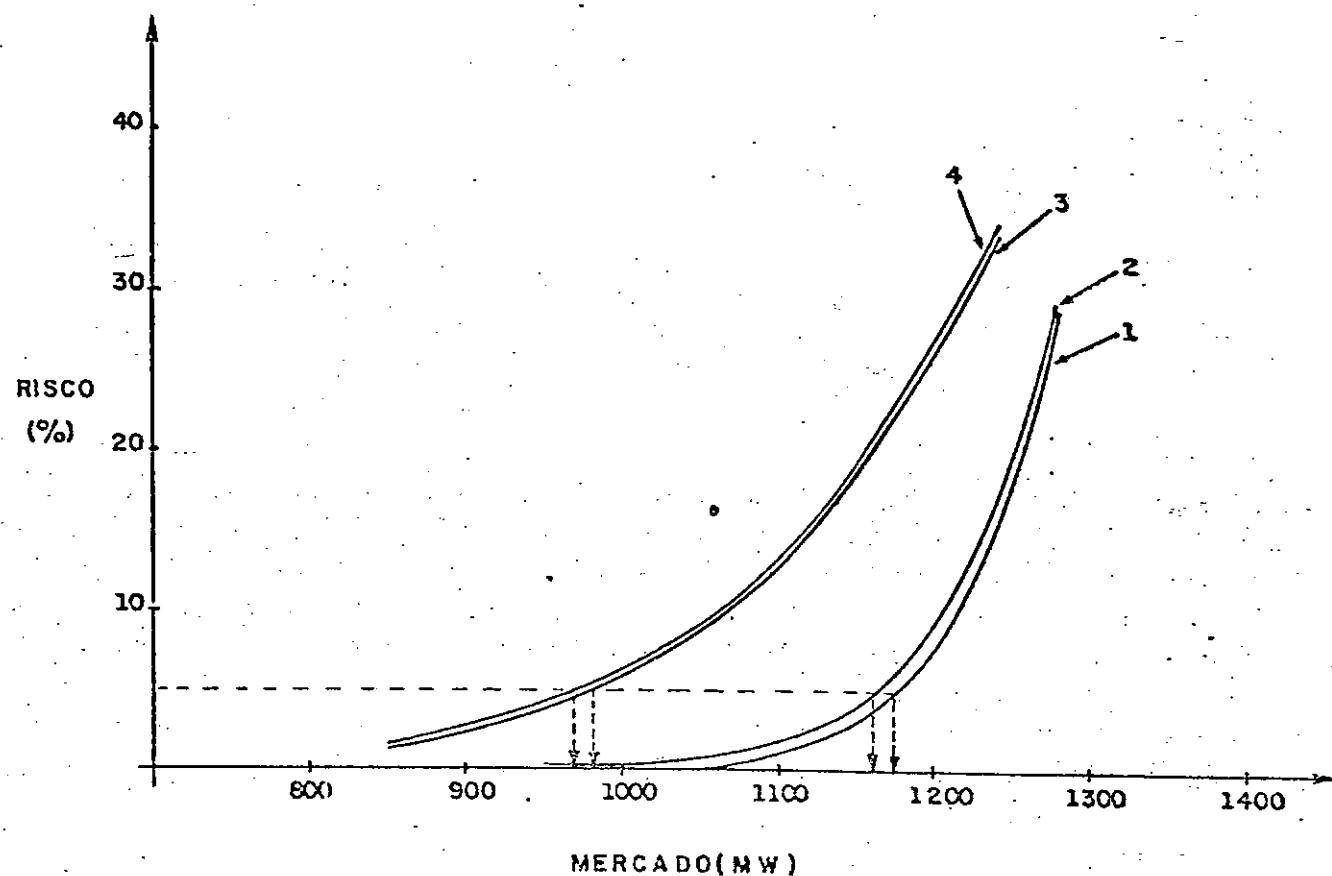
A curva 2 da figura (5.2) mostra a relação entre o mercado e o risco fornecida pela equação (5.7) baseando-se apenas nas informações amostrais. A capacidade de geração em Carolina é estimada na curva 2 em 1162 Mw.

### V.3.2 Uso das Informações Fornecidas pelo Modelo Conceitual do Balanço Hídrico

No capítulo 4 foi utilizada a relação entre chuva e vazão do modelo conceitual de Solomon para se obter uma amostra simulada de vazões anuais em Carolina. A tabela 5.1 compara os momentos de  $\mu_x$  e  $\sigma_x^2$  fornecidos pela amostra medida e pela amostra simulada. A combinação das duas fontes de informação é feita calculando-se novas estatísticas suficientes através da operação (2.8) definida em (2.32) a (2.35).

INFORMAÇÃO	$E(\mu_x)$	$V(\mu_x)$	$E(\sigma_x^2)$	$V(\sigma_x^2)$
AMOSTRA MEDIDA				
$n=11, v=10$	3537	33 037	363 409	$4,402 \times 10^{10}$
AMOSTRA SIMULADA				
$n=4,76, v=9,36$	3181	347 875	1 655 887	$102,312 \times 10^{10}$
AMOSTRA MEDIDA E SIMULADA				
$n=15,76, v=19,36$	3429	56 899	896 730	$10,470 \times 10^{10}$

TABELA 5.1 - MÔMENTOS DE  $\mu_x$  e  $\sigma_x^2$ , USO DO MODELO CONCEITUAL



**Figura 5.2 - CÚRVAS MERCADO-RISCO PARA CAROLINA.  
USO DO MODELO CONCEITUAL.**

Informação	Distribuição	Curva
Amostra Medida $n = 11, v = 10$	Normal	1
	Bayesiana	2
Amostra Medida E Simulada $n' = 15,76, v = 19,36$	Normal	3
	Bayesiana	4

Os momentos da distribuição bayesiana, dado as informações amostrais e conceituais, são:

$$E [qf|z, Co] = 3429 \quad (5.18)$$

$$V [qf|z, Co] = 953\,629 \quad (5.19)$$

As curvas 3 e 4 na figura (5.2) mostram as relações entre o mercado e o risco baseados na combinação das informações amostrais e conceituais levando-se em conta a incerteza dos parâmetros (curva 4) ou não (curva 3). As estimativas de capacidade de geração são: curva 3 - 980 Mw e curva 4 - 970 Mw.

### V.3.3 Uso das Informações Fornecidas Pelos Modelos de Regressão

No capítulo 3 informações na região do alto e médio Tocantins foram processados através de modelos de regressão com objetivo de se ter uma distribuição "a priori" para  $\mu_x$  e  $\sigma_x^2$ . A tabela 5.2 compara os momentos de  $\mu_x$  e  $\sigma_x^2$  fornecidos pelas amostras e pelos modelos de regressão e pela combinação das três fontes de informação através do teorema de Bayes (capítulo 2).

INFORMAÇÃO	$E(\mu_x)$	$V(\mu_x)$	$E(\sigma_x^2)$	$V(\sigma_x^2)$
AMOSTRAS MEDIDA E SIMULADA				
$n = 15,76, v = 19,36$	3429	56 899	896 730	$10,470 \times 10^{10}$
MODELOS DE REGRESSÃO				
$n' = 14,6, v' = 27,8$	3291	104 703	1 528 697	$19,638 \times 10^{10}$
AMOSTRAS MEDIDA E SIMULADA E MODELOS DE REGRESSÃO				
$n'' = 30,36, v'' = 47,16$	3363	40 173	1 219 613	$6,893 \times 10^{10}$

TABELA 5.2: MOMENTOS DE  $\mu_x$  e  $\sigma_x^2$  - USO DOS MODELOS CONCEITUAL E DE REGRESSÃO

Informação	Distribuição	Curva
Amostra Medida	Normal	1
$n = 11, v = 10$	Bayesiana	2
Todas as Informações Normal		5
$n = 30, 36, v = 47, 16$	Bayesiana	6

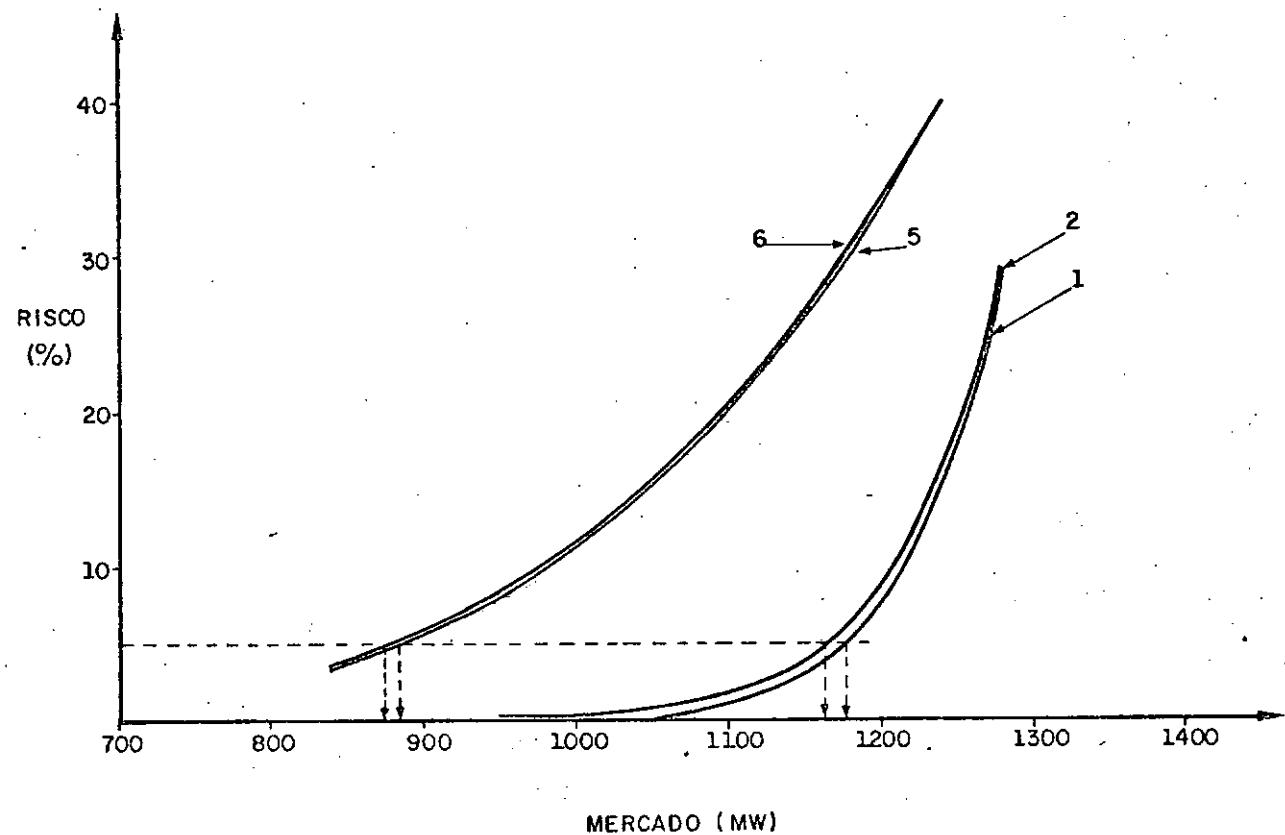


Figura 5.3 - CURVAS MERCADO-RISCO PARA CAROLINA.  
USO DOS MODELOS DE REGRESSÃO

Os momentos da distribuição bayesiana, dado as informações amostrais, conceituais e regionais, são:

$$E [qf | z, Ir, Co] = 3363 \quad (5.20)$$

$$V [qf | z, Ir, Co] = 1259784 \quad (5.21)$$

As curvas 5 e 6 da figura (5.3) mostram as relações entre o mercado e o risco baseadas na combinação de todos os tipos de informação disponível, levando-se em conta (curva 6) ou não (curva 5) a incerteza dos parâmetros. As curvas também mostram as estimativas da capacidade de geração (curva 5 - 895Mw e curva 6 - 885 Mw)

#### V.3.4 Aplicação do Método de Monte-Carlo

O uso da distribuição bayesiana em estudos de simulação de sistemas de recursos hidricos foi discutida por Davis<sup>17</sup>, Mcleod<sup>18</sup> e Vicens<sup>19</sup>. Conforme Davis<sup>17</sup> e Mcleod<sup>18</sup> as séries obtidas com o método não reproduzem as informações existentes. Por exemplo, se as vazões anuais possuem distribuição normal, as vazões geradas pelo método são distribuídas conforme a distribuição de student. O método correto para se levar em conta a incerteza de parâmetros é gerar séries de vazões anuais a partir de vetores de parâmetros gerados conforme a distribuição posterior.

Nos itens V.3.1, V.3.2 e V.3.3 a distribuição bayesiana foi usada para se obter uma aproximação para o primeiro momento do risco associado a um mercado. A seguir, compara-se estas aproximações com as soluções obtidas pelo método de Monte-Carlo.

Este método consiste em gerar aleatoriamente um número grande de valores ( $\mu_x$ ,  $\sigma_x^2$ ), de forma tal que os valores gerados se distribuam conforme  $f''(\theta)$ . Para cada par  $\theta_i = (\mu_x, \sigma_x^2)_i$  resolve-se a equação (5.4), (usando  $\theta_i$  como estimativas de  $\theta$ ) e obtem-se o risco  $r_i$ . A estimativa dos momentos de ordem k da distribuição do risco é dada por:

$$\hat{\gamma}_k = \frac{\sum (r_i)^k}{N} \quad (5.22)$$

onde  $N$  é o número de pares utilizados. Logo o valor médio da distribuição do risco é estimada usando a média  $\bar{r}$  dos  $r_i$ . O erro médio desta estimativa é dado por:

$$\sigma(\bar{r}) = \frac{sd}{\sqrt{N}} \quad (5.23)$$

onde,  $sd$  é o desvio padrão da distribuição do risco, cuja estimativa é dada por:

$$\hat{sd}^2 = \frac{\sum (r_i - \bar{r})^2}{N} \quad (5.24)$$

A figura (5.4) compara as curvas mercado-valor esperado do risco, para o aproveitamento de Carolina, obtidas pelo método de Monte-Carlo e pelo uso da distribuição bayesiana, quando apenas se considera as informações amostrais. Na mesma figura está indicado o intervalo de confiança de 95%, para o risco assumindo-se a normalidade da distribuição do risco e portanto estimado por  $\bar{r} \pm 1.65 \hat{sd}$ .

A figura (5.5) apresenta os resultados obtidos combinando as informações amostrais, regionais e conceituais. As figuras (5.4) e (5.5) indicam que os dois métodos fornecem valores praticamente iguais para o potencial de Carolina. O fato de que os riscos obtidos pelo uso da distribuição bayesiana são sempre os menores não tem grande importância, já que o método de Monte-Carlo demonstrou que o risco pode estar situado numa faixa bastante ampla. Por exemplo, o risco associado ao mercado de 952 Mw (energia firme) é menor que 6%, segundo a amos tra (figura 5.4), e menor que 18%, segundo a combinação das informações amostrais, conceituais e regionais (figura 5.5). A tabela (5.3) mostra que o método de Monte-Carlo exige um esforço computacional significativamente maior que o cálculo pela distribuição bayesiana.

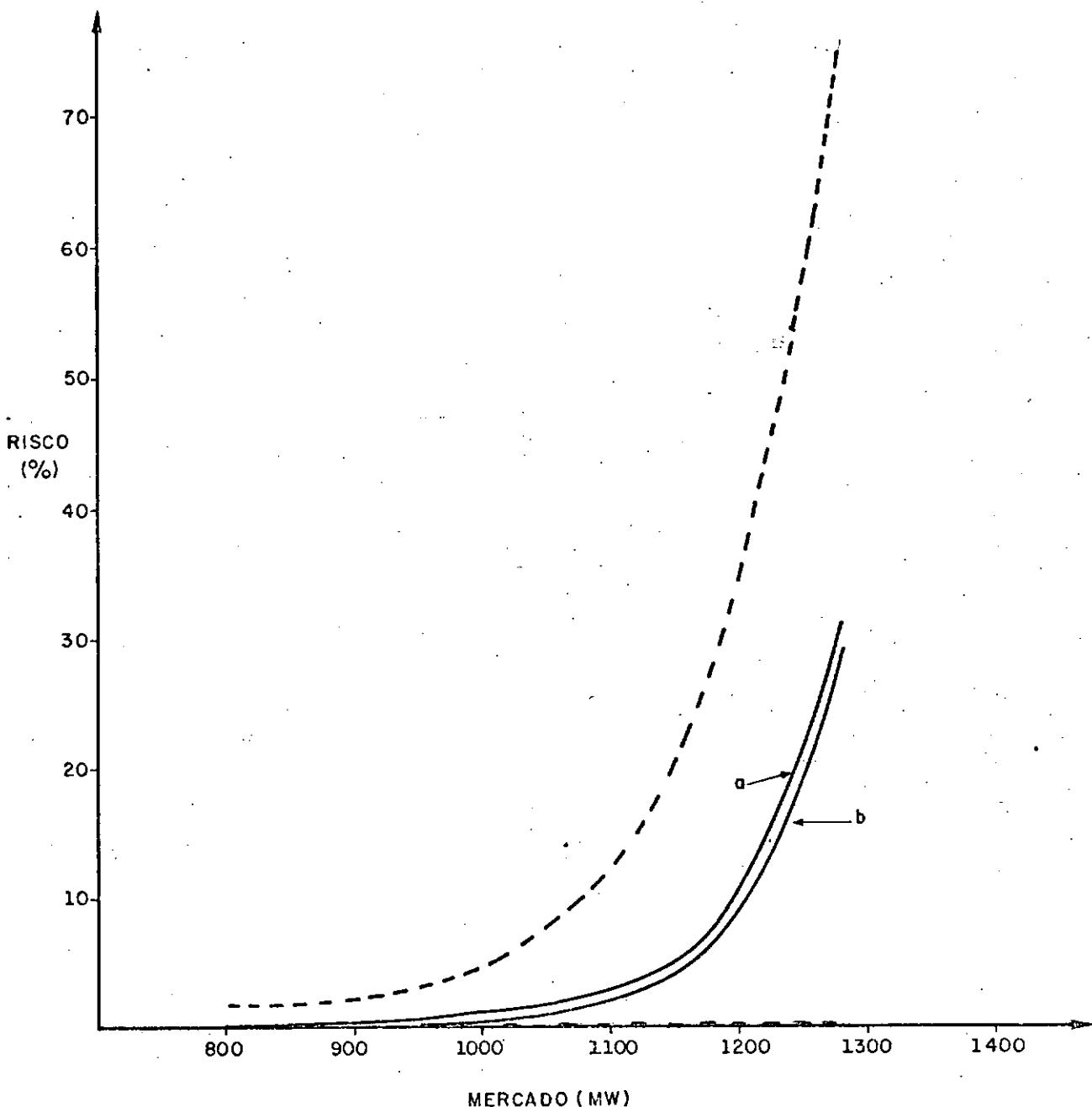


Figura 5.4 - VALOR ESPERADO DO RISCO E INTERVALO DE CONFIANÇA DE 95%. APENAS INFORMAÇÃO AMOSTRAL:

(a - método de Monte Carlo com N=1000,  
b - aproximação de 1ª ordem)

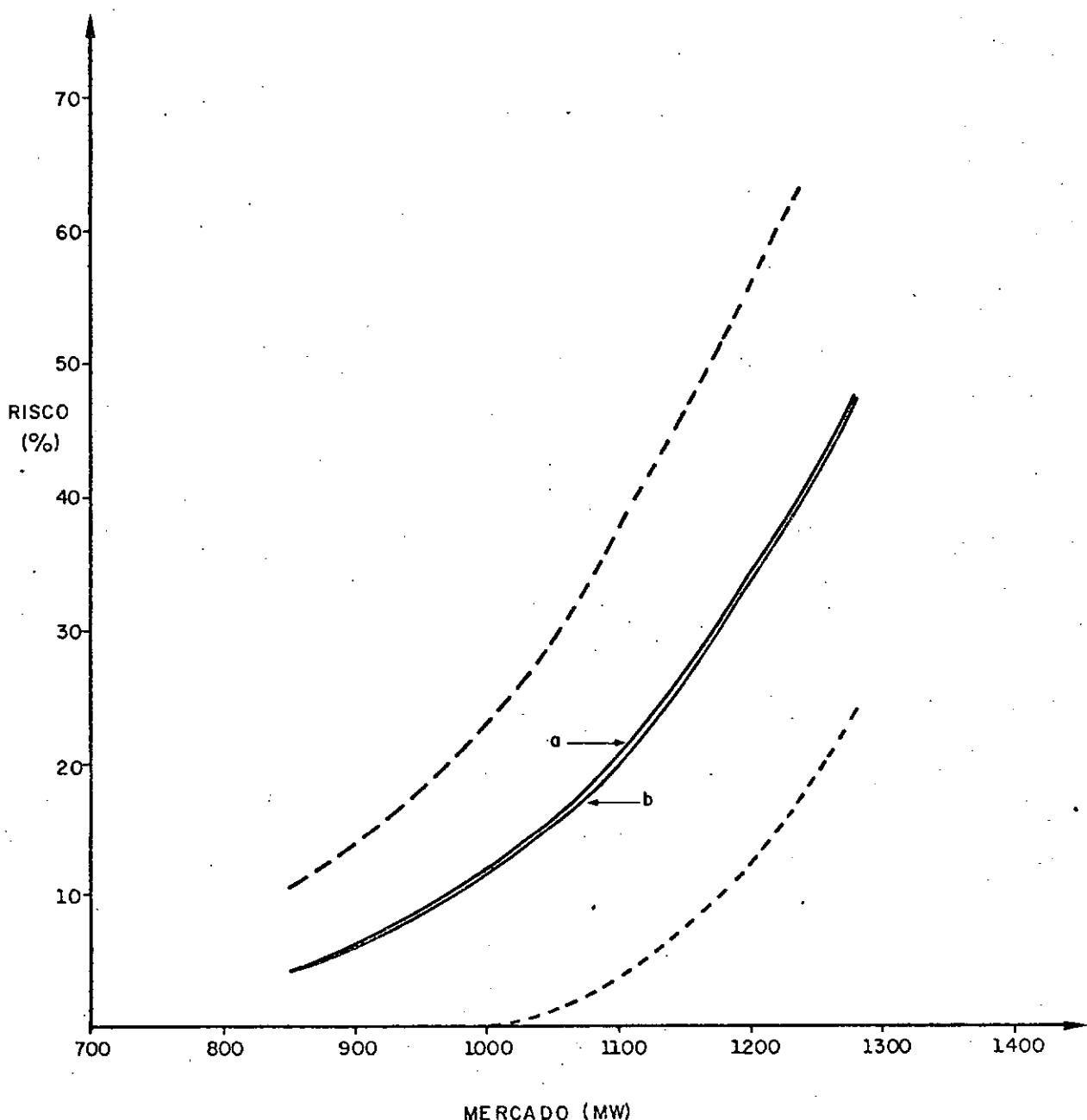


Figura 5.5 - VALOR ESPERADO DO RISCO E INTERVALO DE CONFIANÇA DE 95%. TODAS INFORMAÇÕES COMBINADAS.

(a - método de Monte-Carlo com  $N=1000$ ,  
 b - aproximação de 1<sup>a</sup> ordem)

MÉTODO	FIGURA 5.4	FIGURA 5.5
DIST. BAYESIANA	00:07	00:07
MONTE CARLO	18:01	25:27

TABELA 5.3 - TEMPO DE PROCESSAMENTO EM C.P.U. NO COMPUTADOR PDP-1170 DOS DOIS MÉTODOS.

#### V.4

#### SUMÁRIO

Este capítulo apresenta uma metodologia para a avaliação da capacidade de geração de um aproveitamento hidroelétrico baseado na curva mercado-risco. Mostra-se como o risco do não atendimento de um certo mercado compõe a solução de um sistema de equações lineares, cujos coeficientes são obtidos da distribuição de probabilidades da vazão afluente ao reservatório,  $f_x(x)$ . Quando existem incerteza adicionais, referentes ao valor dos parâmetros de  $f_x(x)$ , o risco associado a um certo mercado é visto como uma variável aleatória, cujo primeiro momento pode ser obtido por uma aproximação de primeira ordem usando a distribuição bayesiana de uma vazão futura. Alternativamente, o uso do método de Monte Carlo permite obter estimativas de momentos de qualquer ordem e portanto definir intervalos de confiança para o risco.

Na aplicação da metodologia ao aproveitamento de Carolina, no rio Tocantins, a consideração das incertezas referentes ao valor da média e variância das afluências tem um efeito insignificante na estimativa da capacidade de geração do aproveitamento. Por outro lado, o uso de informações regionais e conceituais diminui em  $100 \times \frac{1.175-885}{1.175} = 25\%$  a avaliação do potencial de Carolina.

CAPÍTULO VICONCLUSÕES

Nesta tese a metodologia bayesiana é usada com dois objetivos: estimar o efeito das incertezas envolvidas nas estimativas de parâmetros hidrológicos e avaliar a importância do uso da informações regionais e conceituais em projetos de reservatórios.

A metodologia foi aplicada a apenas um caso específico: definir qual o mercado energético que o aproveitamento de Carolina, no rio Tocantins, pode atender com um risco de 5%. O comprimento reduzido dos registros fluviométricos na bacia do alto e médio Tocantins indicam a necessidade do uso de modelos de transferência de informação em qualquer estudo hidrológico para esta região. A existência de apenas um posto com registro longo (Porto Nacional) reduz as chances de se estender de forma confiável os registros curtos através do uso da correlação espacial. A pequena densidade de postos fluviométricos torna mais importante o uso de modelos de regressão relacionando parâmetros hidrológicos a características físicas da bacia de forma tal que se possa obter estimativas para locais sem postos de medição. Devido às dimensões da região, a aplicação do método através de arquivos em malha quadrada facilita a execução dos estudos necessários tais como: cálculo e escolha variáveis fisiográficas, obtenção das equações e definição de regiões homogêneas (PHCA<sup>12</sup>).

Para a construção dos modelos de regressão a região estudada foi considerada como uma unidade homogênea baseando-se em PHCA<sup>12</sup>, que analisa basicamente os mesmos dados sem encontrar evidência em contrário. Eventualmente, à medida que novos dados sejam incorporados, pode-se tornar necessário a definição de sub-regiões. O modelo ajustado à variância de vazões anuais obteve um ótimo rendimento. Ou seja, a quantidade de informações transferidas é equivalente ao registro mais longo disponível na região. Pode-se concluir que a precisão das estimativas de variância na região só pode ser aumentada à medida que estejam

disponíveis registros mais longos. Já o modelo de previsão para a média pode ter o seu rendimento melhorado por exemplo através do uso racional das informações geológicas, pedológicas e de vegetação que não foram exploradas neste estudo. Convém citar que tanto as equações de predição resultantes, quanto às duas últimas conclusões, foram baseadas nas observações disponíveis de vazões anuais na região. A medida que maior volume de dados esteja à disposição, novos estudos fornecerão estimativas e avaliações da situação com confiabilidades crescentes.

O modelo conceitual usado no Capítulo IV para aproveitar a existência de registros longos de precipitações na região tem um conteúdo de informação reduzido porque se comporta mal fora do período de calibração (ver anexo 3). Isto pode ser causado por erros de media de precipitação (Johnson<sup>20</sup>) mas em geral este é o principal defeito do uso de modelos conceituais em transferência de informação (Clarke<sup>21</sup> ou Sokolov<sup>22</sup>). Eventualmente o uso de outro modelo conceitual onde o intervalo de tempo do impulso (precipitação) seja menor que o intervalo de tempo da resposta (vazão) pode fornecer melhores resultados (Johnson<sup>20</sup>). Outros desenvolvimentos seriam a divisão em sub-bacias e o maior detalhamento do ciclo hidrológico a partir do conhecimento do solo e vegetação da região. É interessante observar que o potencial fornecido pela combinação da amostra com a série simulada pelo modelo conceitual, 980 Mw, é praticamente igual ao valor obtido por Engevix<sup>1</sup>, 952 Mw, onde também estendeu-se os dados usando um modelo físico (SSARR).

Os resultados encontrados no Capítulo V permitem estabelecer as seguintes conclusões de caráter geral, válidas não apenas para o caso estudado.

- Em situações semelhantes (quantidade de informações, variabilidade de vazões e escala de projeto similares) o uso de informações regionais e conceituais em projetos de reservatórios pode ter um efeito significante.
- A consideração das incertezas quanto aos parâmetros teve pouca influência no resultado final. Pode-se prever que situações de maior escassez de dados e/ou maior variabilidade de vazões provoquem o aumento desta influência.

ANEXO 1

DESCRÍÇÃO DOS ARQUIVOS DA MALHA QUADRADA DESENVOLVIDOS PELO PHCA NA BACIA DO RIO TOCANTINS.

A malha adotada usa como sistema de referência os paralelos e meridianos e tem como limite inferior esquerdo o ponto de coordenadas longitude  $52^{\circ}\text{W}$  e latitude  $17^{\circ}\text{S}$ . Cada grau de latitude e longitude correspondem respectivamente a uma linha e uma coluna de elementos.

O acesso das informações dos arquivos fisiográficos (PHYSIO) e de drenagem (DRAIN) para um certo quadrado é feito consultando previamente um arquivo auxiliar denominado INDEX.

ARQUIVO INDEX

- Arquivo auxiliar que permite rápido acesso às informações.
- Possui um registro para cada elemento da malha.
- Cada registro contém uma palavra.

Se o elemento correspondente pertence a área do estudo a palavra contém a posição do seu registro no arquivo fisiográfico. No caso contrário é zero.

- A posição no arquivo INDEX do registro do elemento da coluna NOCOL e a linha NOROW é dada por:

$$N = 113 \text{ (NOCOL - 1)} + NOROW$$

onde 113 é o número máximo de linhas.

ARQUIVO PHYSIO

Arquivo contendo as informações topográficas e de vegetação, solo e geologia superficial para cada quadrado da área estudada. É formado por um registro inicial seguido de 2485 registros de dados, cada um correspondendo a um quadrado da área estudada. Cada registro contém 30 palavras descritas a seguir, ao lado da simbologia adotada nesta tese.

1. NOCOL - Número da coluna do quadrado
2. NOROW - Número da linha do quadrado
3. LONG - Código de longitude
4. LAT - Código de latitude
5. A - Área do quadrado ( $\text{km}^2$ )
6. ELN - Elevação do canto nordeste (m)
7. ALT - Altitude média (m)
8. S - Declividade média ( $100 \times \%$ )
9. AZ - Azimute da declividade (graus)
10. Dn - Distância até o oceano na direção norte (km)
11. Dne - Distância até o oceano na direção nordeste (km)
12. De - Distância até o oceano na direção este (km)
13. Dse - Distância até o oceano na direção sudeste (km)
14. Ds - Distância até o oceano na direção sul (km)
15. Bn - Barreira na direção norte (km)
 

Barreira é definida pela diferença entre a altitude média do quadrado e a maior altitude entre o quadrado e o oceano na direção considerada. Se não existe altitude maior que a do quadrado entre o quadrado e o oceano a barreira é zero.
16. Bne - Barreira na direção nordeste (km)
17. Be - Barreira na direção este (km)
18. Bse - Barreira na direção sudeste (km)
19. Bs - Barreira na direção sul (km)
20. En - Escudo na direção norte (km)
 

Escudo é definido como a soma dos trechos ascendentes do oceano até o quadrado (km)
21. Ene - Escudo na direção nordeste (km)
22. Ee - Escudo na direção este (km)
23. Ese - Escudo na direção sudeste (km)
24. Es - Escudo na direção sul (km)
25. GEO - Código de geologia superficial
26. VEG - Código de vegetação
27. SOIL - Código de solo
28. OUT - Código da direção da drenagem no quadrado  
(Norte: 1; Este: 2; Sul: 3 e Oeste: 4)
29. - - Não usada
30. NREC - Número do próximo registro

ARQUIVO DRAIN

Arquivo contendo informações do sistema de drenagem. Tem seus registros correspondendo um a um com o arquivo fisiográfico. Cada registro contém 8 palavras descritas a seguir:

- 1 - Número da coluna do quadrado
- 2 - Número da linha do quadrado
- 3 - Posição do registro do quadrado de jusante
- 4 - Área do quadrado
- 5 - Posição do registro de quadrado de montante
- 6 - Posição do registro de quadrado de montante
- 7 - Posição do registro de quadrado de montante
- 8 - Número do próximo registro

A figura A.1 mostra a representação do sistema de drenagem na área do estudo.

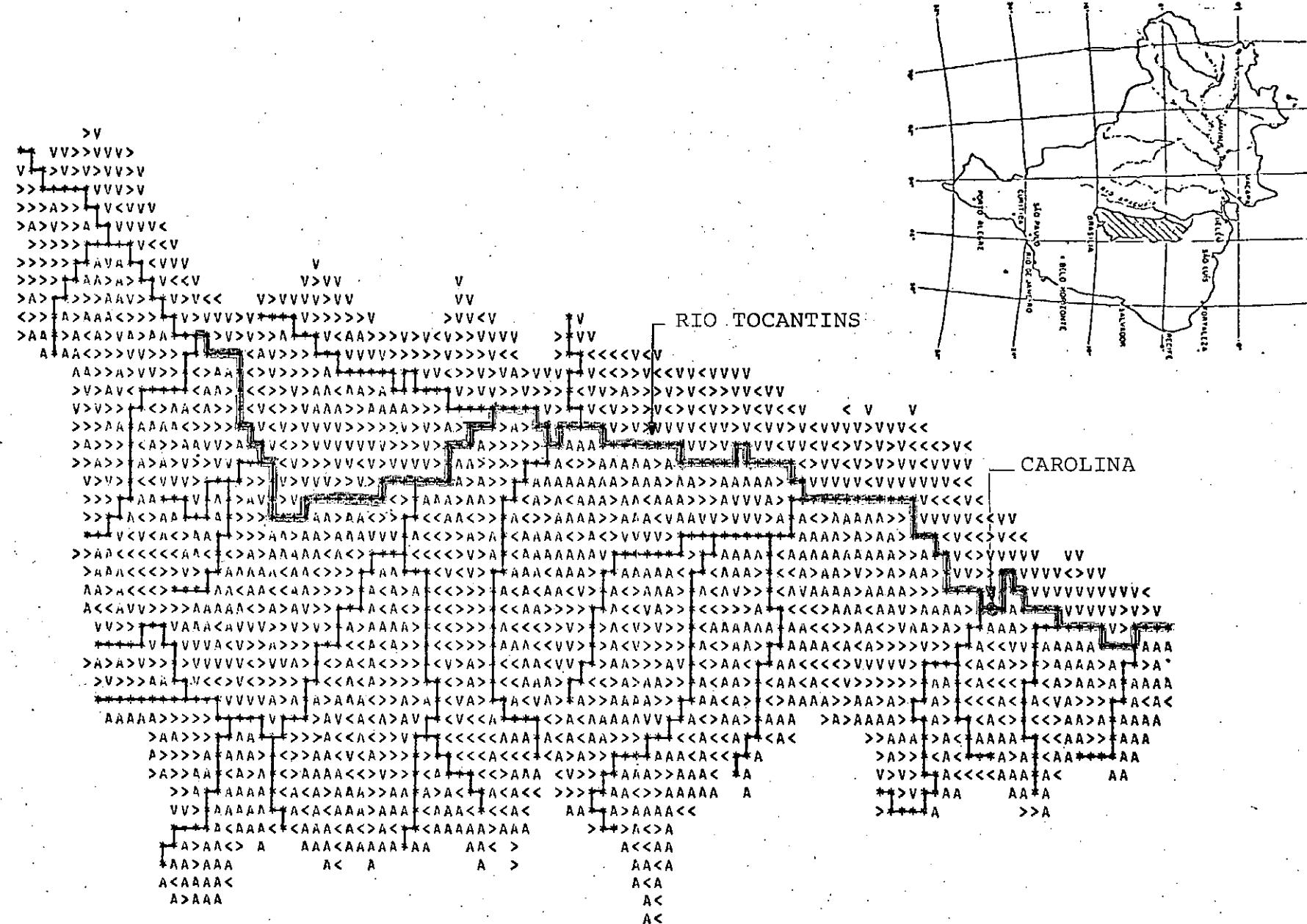


FIGURA A.1 - REPRESENTAÇÃO DO SISTEMA DE DRENAGEM DA BACIA DO ALTO-MÉDIO TOCANTINS

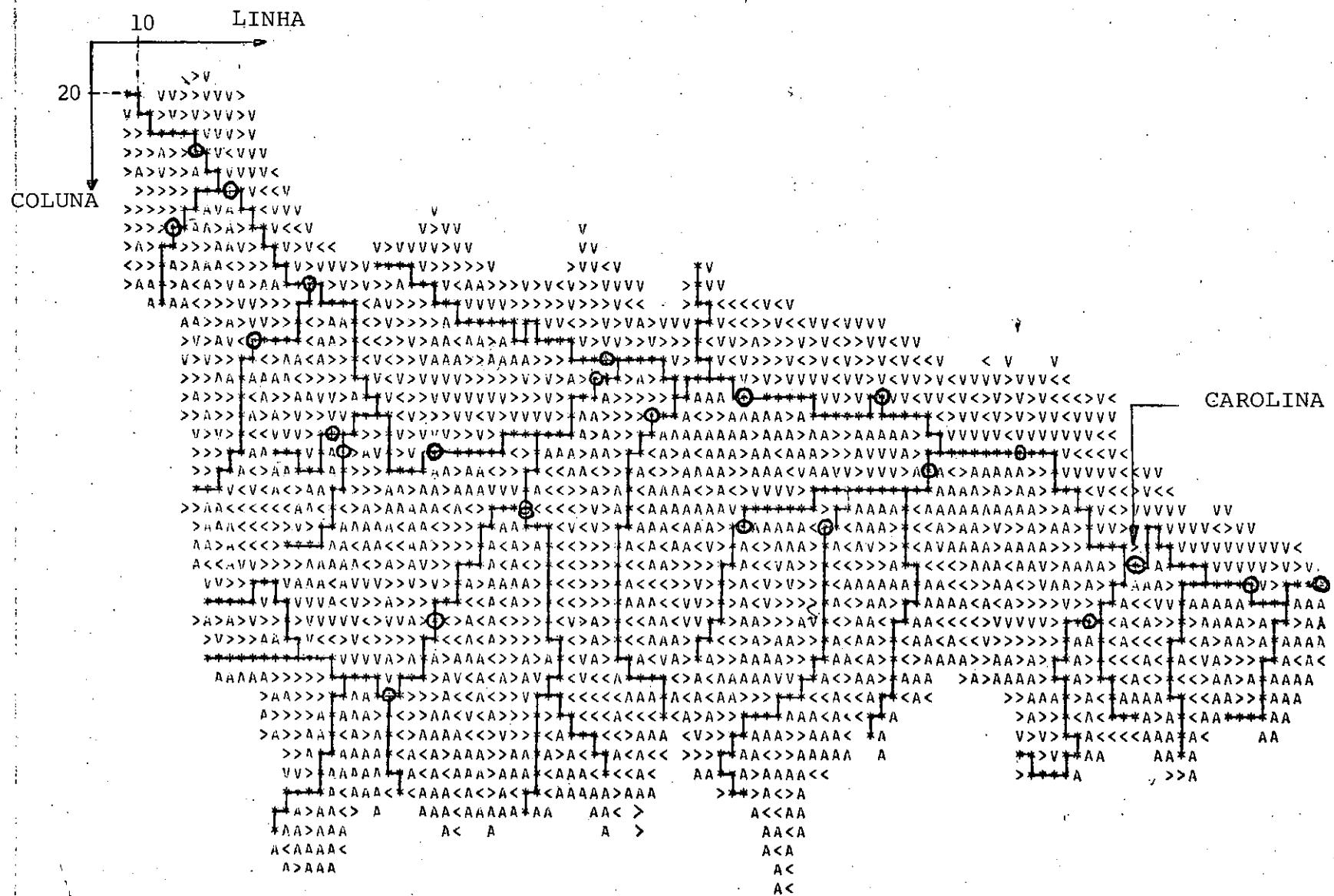


FIGURA A.2 - REDE DE ESTAÇÕES FLUVIOMÉTRICAS

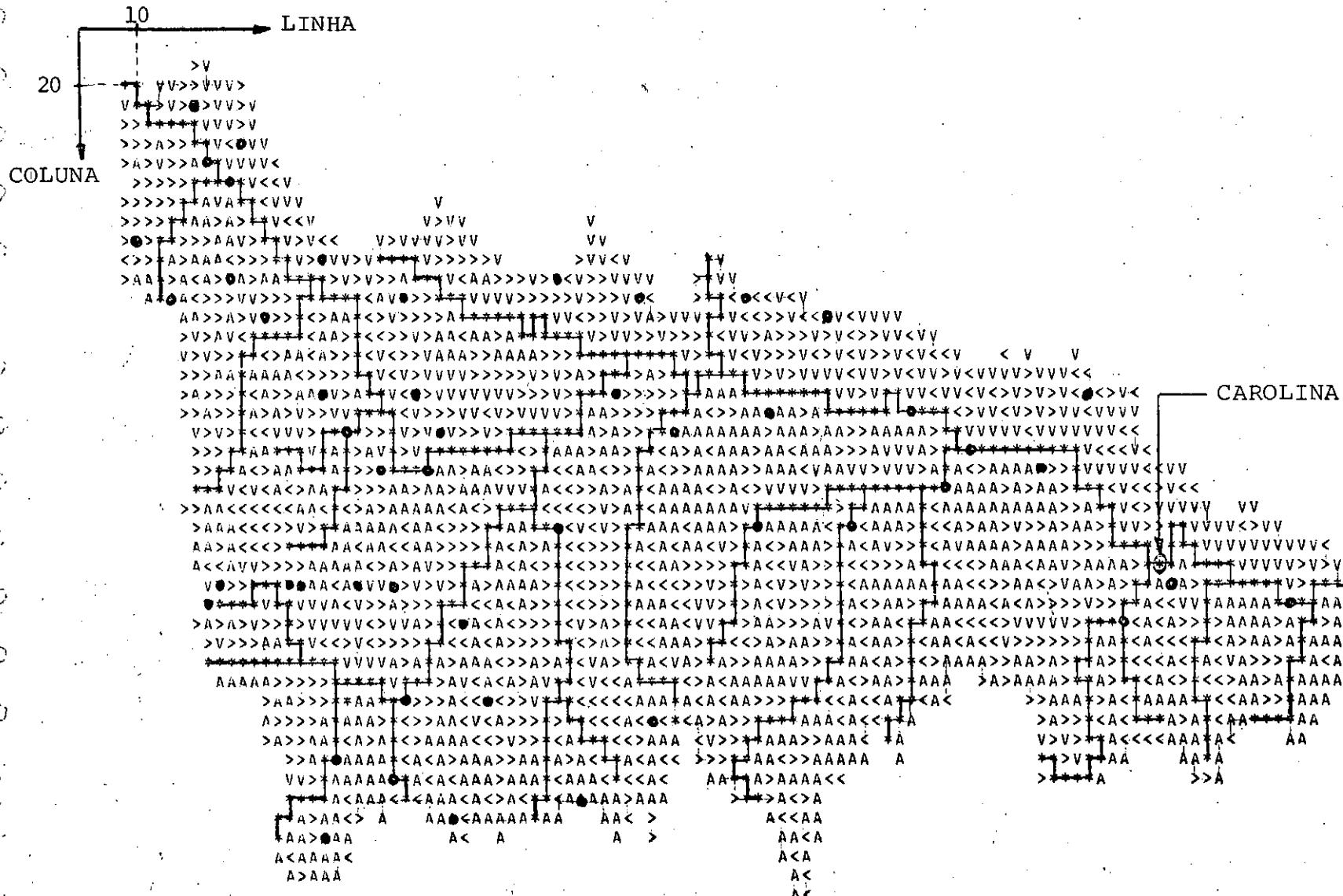


FIGURA A.3 - REDE DE ESTAÇÕES PLUVIOMÉTRICAS

ANEXO 2APROXIMAÇÃO PARA A CORRELAÇÃO ENTRE  
ESTIMADORES DE MÉDIA E VARIÂNCIA

Sejam duas variáveis aleatórias X e Y. Admita-se que:

$$E(X) = E(Y) = 0 \quad (A2.1)$$

$$E(X^2) = E(Y^2) = 1 \quad (A2.2)$$

$$E(XY) = \rho \quad (A2.3)$$

$$E(XY^2) = E(X^2Y) = 0 \quad (A2.4)$$

$$E(X^2Y^2) = 1 + 2\rho^2 \quad (A2.5)$$

Existem N + N1 observações independentes de X e N + N2 observações independentes de Y arrumadas nas amostras AM1 = {x<sub>1</sub>, x<sub>2</sub>, ..., x<sub>N</sub>, x<sub>1</sub><sup>\*</sup>, x<sub>2</sub><sup>\*</sup>, ..., x<sub>N1</sub><sup>\*</sup>} e AM2 = {y<sub>1</sub>, y<sub>2</sub>, ..., y<sub>N</sub>, y<sub>1</sub><sup>\*</sup>, y<sub>2</sub><sup>\*</sup>, ..., y<sub>N2</sub><sup>\*</sup>}, sendo N o número de observações concomitantes, N1 o número de observações de X sozinho e N2 de Y sozinho. Sejam as estatísticas:

$$\hat{\mu}_x = \frac{\sum x + \sum x^*}{N + N1} \quad (A2.6)$$

$$\hat{\mu}_y = \frac{\sum y + \sum y^*}{N + N2} \quad (A2.7)$$

$$\hat{\alpha}_x = \frac{\sum x^2 + \sum (x^*)^2}{N + N1} \quad (A2.8)$$

$$\hat{\alpha}_y = \frac{\sum y^2 + \sum (y^*)^2}{N + N2} \quad (A2.9)$$

$$\hat{\sigma}_x^2 = \hat{\alpha}_x - \hat{\mu}_x^2 \quad (A2.10)$$

$$\hat{\sigma}_y^2 = \hat{\alpha}_y - \hat{\mu}_y^2 \quad (A2.11)$$

(A2.1) e (A2.2) fornecem;

$$E(\hat{\mu}_x) = E(\hat{\mu}_y) = 0 \quad (A2.12)$$

$$VAR(\hat{\mu}_x) = \frac{1}{N + N_1} \quad (A2.13)$$

$$VAR(\hat{\mu}_y) = \frac{1}{N + N_2} \quad (A2.14)$$

Supondo N grande ou N1 e N2 grandes, tem-se:

$$\hat{\sigma}_x^2 = \hat{\alpha}_x \quad (A2.15)$$

$$\hat{\sigma}_y^2 = \hat{\alpha}_y \quad (A2.16)$$

Logo, (A2.2) e (A2.5) fornecem;

$$E(\hat{\sigma}_x^2) = E(\hat{\sigma}_y^2) = 1 \quad (A2.17)$$

$$VAR(\hat{\sigma}_x^2) = \frac{2}{N + N_1} \quad (A2.18)$$

$$VAR(\hat{\sigma}_y^2) = \frac{2}{N + N_2} \quad (A2.19)$$

Inicialmente deseja-se obter o coeficiente de correlação entre  $\hat{\mu}_x$  e  $\hat{\mu}_y$ , denotado por  $\rho_{\mu}$ . Por (A2.12) tem-se:

$$\text{cov}(\hat{\mu}_y, \hat{\mu}_x) = E(\hat{\mu}_x \cdot \hat{\mu}_y)$$

Expandindo  $E(\hat{\mu}_x \cdot \hat{\mu}_y)$ , obtém-se:

$$\text{cov}(\hat{\mu}_y, \hat{\mu}_x) = E\left[\frac{(\Sigma x + \Sigma x^*) (\Sigma y + \Sigma y^*)}{(N+N_1)(N+N_2)}\right] =$$

$$= E\left[\frac{\Sigma x \Sigma y}{(N+N_1)(N+N_2)}\right] = \frac{N E(XY)}{(N+N_1)(N+N_2)} ;$$

por (A2.3) :

$$\text{cov}(\hat{\mu}_x, \hat{\mu}_y) = \frac{N}{(N+N_1)(N+N_2)} \cdot \rho$$

usando (A2.13) e (A2.14) :

$$\rho_M = \frac{N}{(N+N_1)^{0.5} (N+N_2)^{0.5}} \cdot \rho \quad (\text{A2.20})$$

Em seguida, deseja-se conhecer o coeficiente de correlação entre  $\hat{\sigma}_x^2$  e  $\hat{\sigma}_y^2$ , denotando por  $\rho_V$ . (A2.15) e (A2.16), fornecem:

$$E(\hat{\sigma}_x^2 \cdot \hat{\sigma}_y^2) = E(\hat{\alpha}_x \cdot \hat{\alpha}_y)$$

Expandindo  $E(\hat{\alpha}_x \cdot \hat{\alpha}_y)$ , obtem-se:

$$\begin{aligned} E(\hat{\sigma}_x^2 \cdot \hat{\sigma}_y^2) &= E\left[\frac{(\sum x^2 + \sum (x^*)^2)(\sum y^2 + \sum (y^*)^2)}{(N+N_1)(N+N_2)}\right] = \\ &= \frac{N E(X^2 Y^2) + N^2 - N + NN_1 + N2N + N2N_1}{(N+N_1)(N+N_2)} = \\ &= 1 + \frac{N [E(X^2 Y^2) - 1]}{(N+N_1)(N+N_2)}, \end{aligned}$$

por (A2.5) :

$$E(\hat{\sigma}_x^2 \cdot \hat{\sigma}_y^2) = 1 + \frac{2N}{(N+N_1)(N+N_2)} \cdot \rho^2 \quad (\text{A2.21})$$

usando (A2.17), (A2.18) e (A2.19) :

$$\rho_V = \frac{N}{(N+N_1)^{0.5} (N+N_2)^{0.5}} \cdot \rho^2 \quad (\text{A2.22})$$

ANEXO 3NÚMERO DE ANOS EQUIVALENTES DE INFORMAÇÃO DO MODELO CONCEITUAL

O modelo conceitual apresentado no capítulo IV é aplicado a bacia de Porto Nacional para se obter empiricamente coeficientes de redução do número de anos equivalentes de informações fornecidas pelo modelo.

Sejam  $q_i$  e  $qs_i$ ,  $i = 1, \dots, n$ , respectivamente, vazão anual observada e vazão anual simulada pelo modelo. Obviamente se  $q_i = qs_i$ , as informações da série observada e da série simulada são equivalentes. Por outro lado se as estatísticas:

$$\bar{q} = \frac{\sum q_i}{n} \quad (A3.1)$$

$$\bar{q}_S = \frac{\sum qs_i}{n} \quad (A3.2)$$

são iguais, as informações referentes a média de vazões anuais fornecidas pela série observada e pela série simulada são equivalentes e neste caso os números de anos equivalentes de informação para a média de vazões anuais das duas séries são iguais a  $n$ .

As estatísticas  $\bar{q}$  e  $\bar{q}_S$  são estimativas da vazão média anual com variância dadas por:

$$\text{var } (\bar{q}) = \frac{\sigma^2}{n} \quad (A3.3)$$

$$\text{var } (\bar{q}_S) = \frac{\sigma^2}{n} + E(\sigma_{\Delta}^2) \quad (A3.4)$$

$$\sigma_{\Delta}^2 = (\bar{q} - \bar{q}_S)^2 \quad (A3.5)$$

onde  $\sigma^2$  é a variância da vazão anual.

O termo  $E(\sigma_{\Delta}^2)$  representa os erros em média cometidos pelo modelo representado pela diferença entre  $\bar{q}$  e  $\bar{q}_S$  e eventualmente depende de  $n$ .

A razão entre os números de anos equivalentes de informação para a média, das séries simulada ( $n_S$ ) e observada ( $n$ ), é dada por:

$$\frac{n_S}{n} = \frac{\text{var}(\bar{Q})}{\text{var}(\bar{Q}_S)} \quad (\text{A3.6})$$

Para o parâmetro, variância das vazões anuais, as estimativas são dadas por:

$$s^2 = \frac{\sum (q_i - \bar{q})^2}{n - 1} \quad (\text{A3.7})$$

$$ss^2 = \frac{\sum (qs_i - \bar{q}_S)^2}{n - 1} \quad (\text{A3.8})$$

cujas variâncias são dadas por:

$$\text{var}(s^2) = \frac{2\sigma^4}{(n-1)} \quad (\text{A3.9})$$

$$\text{var}(ss^2) = \frac{2\sigma^4}{(n-1)} + E(\sigma_{\Delta}^2) \quad (\text{A3.10})$$

$$\sigma_{\Delta}^2 = (s^2 - ss^2)^2 \quad (\text{A3.11})$$

A razão entre os números de anos equivalentes de informação para a variância das séries simuladas ( $v_S$ ) e observada ( $v = n - 1$ ) é dada por:

$$\frac{v_S}{v} = \frac{\text{var}(s^2)}{\text{var}(ss^2)} \quad (\text{A3.12})$$

A amostra com 24 anos de precipitação e vazão na bacia de Porto Nacional pode ser dividido em duas sub-amostras. Uma para o ajuste do parâmetro  $w$  do modelo pela fórmula (4.8) e outra para o cálculo de  $\sigma_{\Delta}^2$  e  $\sigma_{\Delta'}^2$ . Usando 14 amostras de ajustes (todas com 11 anos), os 14 valores de  $\sigma_{\Delta}^2$  e  $\sigma_{\Delta'}^2$ , fornecem:

$$E(\sigma_{\Delta}^2) = 2631 \quad (\text{A3.13})$$

$$E(\sigma_{\Delta'}^2) = 12978160 \quad (\text{A3.14})$$

A fórmula (3.53), aplicada em Porto Nacional, fornece:

$$\sigma^2 = 14764 \quad (\text{A3.15})$$

Finalmente, para  $n = 14$  tem-se:

$$\frac{n_S}{n} = 0.34 \quad (\text{A3.16})$$

$$\frac{v_S}{v} = 0.72 \quad (\text{A3.17})$$

BIBLIOGRAFIA

- | 1 | ENGEVIX; ECOTEC, Estudos Hidroenergéticos da bacia do Rio Tocantins, Centrais Elétricas do Norte do Brasil-ELETRO NORTE, 1975.
- | 2 | SOLOMON, S.I.; DENOUVILLIEZ, J.P.; CHART, E.J.; WOOLLEY, J. A.; CADOU, C., The use of a Square Grid System for computer estimation of precipitation, temperature and runoff. Water Resources Research, 4(5), 1968.
- | 3 | BASSO, E., Apresentação. In: III Simpósio Brasileiro de Hidrologia, Brasília, 1979.
- | 4 | BASSO, E.; CADOU, C.; OSORIO, C.; SOLOMON, S.I.; SOULIS, E., Estudo piloto de aplicação de técnica da malha quadrada na bacia do rio Amazonas, Brasil. III Simpósio Brasileiro de Hidrologia, Anais, Vol. 1, 1979.
- | 5 | LLOYD, E.H., Wet and dry water, Bulletin Institute of Mathematica and its applications, 1974a
- | 6 | GOMIDE, F.S., Range and deficit analysis using markov chains, Colorado State University, HP79, 1975.
- | 7 | KLEMES, V., The unreliability of reliability estimates of storage reservoir performance based on short stream flow records, Reliability in water resources management Water Resources publications, 1979.
- | 8 | VICENS, J.G., RODRIGUEZ-ITURBE, I.; SCHAAKE, J.C., A bayesian framework for the use of regional information in hydrology, Water Resources Research, 11(3), 1975.
- | 9 | GOMIDE, F.S., Estimativa de valores médios de variáveis hidrológicas, Parte I e Parte II, III Simpósio Brasileiro de Hidrologia, Anais, Vol. 1, 1979.

- |10| CAREY, D.I.; HAAN, C.T., Using parametric models of runoff to improve parameter estimates for stochastic models, Water Resources Research, 11(6), 1975.
- |11| RAIFFA, H.; SCHLAIFER, R., Applied Statistical Decision Theory, M.I.T. Press, 1961
- |12| Projeto de Hidrologia e Climatologia da Amazonia, Síntese de informações hidrológica com o método da malha quadrada, Segunda parte, Aplicação do método à bacia do rio Tocantins, publicacão número 18, SUDAM, 1979.
- |13| SOLOMON, S., Relationship between precipitation, evaporation, and runoff in tropical equatorial regions, Water Resources Research, 3(1), 1967.
- |14| BOUCHET, R.J., Evapotranspiration réelle et potentielle, signification climatique. Gen. Assembly Berkeley, pub. 62, AIHS, 1963.
- |15| NUNES, G.S.S.; ANDRÉ R.G.B.; VIANELLO, R.L.; MARQUES, V. S., Estudo da distribuição de radiação solar incidente sobre o Brasil, Instituto de Pesquisas Espaciais, Relatório INPE - 1190 - NTE/110, 1978.
- |16| DORAN, D.G., An efficient transition definition for discrete state reservoir analysis, The divided interval technique, Water Resources Research, 11(6) , 1975.
- |17| DAVIS, D. R., Comment on "Bayesian Generation of synthetic streamflows", by G.J. Vicens, I.Rodrigues-Iturbe, and J.C. Schaake, in, Water Resources Research 13(5), 1977.
- |18| MCLEOD, A.I.; HIPEL, W.R., Reply, Water Resources Research, 15(6), 1979.

- | 19 | VICENS, J.G., Comment on "Simulation procedures for Box-Jenkins models" by A.I. Mcleod and K.M. Hipel, Water Resources Research, 15(6), 1979.
- | 20 | JOHNSON, P., Integrated network and the influence fo error in precipitation and evaporation data on streamflow prediction, Hydrological network design and information transfer, World Meteorological Organization, Report nº 8, 1976.
- | 21 | CLARKE, R.J., Comments and summary of discussions, Hidrological network design and information transfer World Meteorological Organization, Report nº8, 1976.
- | 22 | SOKOLOV, A.A.; CHAPMAN, Métodos para o cálculo do balanço hídrico, tradução do Estudos e Relatórios de Hidrologia, nº 17, UNESCO, Centrais Elétricas Brasileira, 1978.