

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE

CENTRO DE CIÊNCIAS E TECNOLOGIA

CURSO DE PÓS – GRADUAÇÃO EM ENGENHARIA CIVIL E AMBIENTAL

**ESTIMATIVA DE VAZÕES MÉDIAS SAZONAIS EM BACIAS SEM
DADOS HIDROMÉTRICOS PARA USO EM SISTEMAS DE
PREVISÃO HIDROMETEOROLÓGICA**

ROBERTA NÓBREGA TORREÃO DE MELO

CAMPINA GRANDE – PB

MAIO – 2003

ROBERTA NÓBREGA TORREÃO DE MELO

**ESTIMATIVA DE VAZÕES MÉDIAS SAZONAIS EM BACIAS SEM DADOS
HIDROMÉTRICOS PARA USO EM SISTEMAS DE PREVISÃO
HIDROMETEOROLÓGICA**

Dissertação apresentada ao Curso de Pós –
Graduação em Engenharia Civil, na área de
Engenharia de Recursos Hídricos, em
cumprimento às exigências para obtenção
do Grau de Mestre.

ÁREA DE CONCENTRAÇÃO: RECURSOS HÍDRICOS

ORIENTADORES: CARLOS OLIVEIRA GALVÃO

RAIMUNDO SÉRGIO S. GÓIS

Campina Grande – PB

Maio – 2003



M528e
2003

Melo, Roberta Nóbrega Torreão de
Estimativa de vazões médias sazonais em bacias sem dados
hidrométricos para uso em sistemas de previsão hidrometeorológica /
Roberta Nóbrega Torreão de Melo – Campina Grande: UFCG, 2003.
122p. : il.

Dissertação (Mestrado). UFCG/CCT
Inclui bibliografia

1. Previsão Hidrometeorológica 2. Regionalização de vazão 3.
Geoprocessamento I. Título

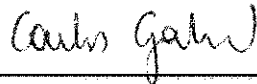
CDU: 556.06"32"

ROBERTA NÓBREGA TORREÃO DE MELO

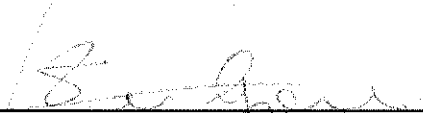
**ESTIMATIVA DE VAZÕES MÉDIAS SAZONAIS EM BACIAS SEM
DADOS HIDROMÉTRICOS PARA USO EM SISTEMAS DE
PREVISÃO HIDROMETEOROLÓGICA**

Aprovada em 30 de Maio de 2003

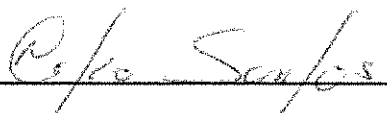
COMISSÃO EXAMINADORA



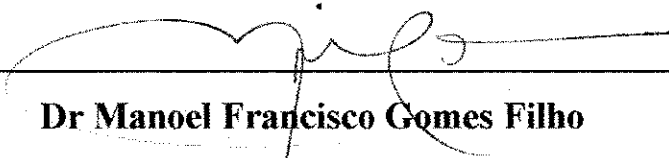
Prof. Dr. Carlos de Oliveira Galvão
Orientador



Prof. Raimundo Sérgio Santos Góis
Orientador



Dr Celso Augusto Guimarães Santos
Examinador



Dr Manoel Francisco Gomes Filho
Examinador

A
Marlene,
Romero
e Roberto

(Por serem a razão de
tudo em minha vida)

AGRADECIMENTOS

Ao professor Carlos de Oliveira Galvão, pela sua competência profissional essencial para a minha formação, pois além de grande orientador soube ser um amigo paciente, mostrando-me a verdadeira dimensão de um trabalho científico.

Ao professor Raimundo Sérgio Santos Góis, pela confiança, por tantos anos de trabalho, pela amizade e por me mostrar a área de Recursos Hídricos, incentivar-me ao mestrado e com quem muito aprendi.

Ao professor José Welington dos Santos, estatístico da EMBRAPA que sempre prestativo e atencioso, contribuiu para a minha pesquisa com o pacote estatístico e informações importantes de uso.

A todos os funcionários, professores e alunos da área de Recursos Hídricos da UFCG em Campina Grande, que ajudaram direta ou indiretamente no meu trajeto, e pelos quais adquiri amizade.

A meu esposo Roberto Torreão Viana de Melo pela compreensão, amor e incentivo nas muitas horas difíceis que enfrentei. Sem o qual não chegaria a lugar algum. E a seus pais (José Viana e Socorro) e irmãos (Reinaldo, Ricardo, Renata, Risele e Rossana) pelo apoio constante.

A meus pais (Romero de Sousa Nóbrega e Marlene Nóbrega de Sousa) e meus irmãos (Misael, Renata e Rossana), por confiarem, incentivarem e me fazerem acreditar que o mais importante de tudo é ser uma boa pessoa e sempre fazer o bem, que tudo mais é consequência.

E, em especial, a Deus por ser sempre meu guia e ter me ensinado que toda tarefa, por mais nobre que seja, está destinada a enfrentar obstáculos e por isso, devemos ser pacientes e acima de tudo, nunca perder a esperança. Ele disse: “Aquele, porém, que perseverar até o fim, esse será salvo” (Jesus).

RESUMO

Previsões de precipitação sazonal estão disponíveis para o nordeste brasileiro. Estas previsões de precipitações têm sido utilizadas para se prever vazões médias sazonais nas bacias com séries de vazão e precipitação observadas. Houve então a motivação para o desenvolvimento de um estudo para estimativas de vazões médias diárias sazonais em bacias onde não existem dados hidrométricos observados. Utilizaram-se para tal, métodos estatísticos de correlação e regressão linear múltipla, além da aplicação da técnica de regionalização hidrológica, adaptada à estimativa de valores anuais da vazão média, em contraposição à estimativa, usual em regionalização, da vazão média de longo período. É proposto, neste estudo, um modelo baseado nas características físicas e hidrológicas de bacias hidrográficas, e em seguida, este é aplicado a 13 bacias de postos fluviométricos, inseridos na bacia hidrográfica dos rios Piranhas-Açu-Apodi, no semi-árido nordestino. Foram obtidas boas correlações entre os dados observados e os valores obtidos na modelagem em testes de validação, indicando que o método proposto é uma técnica que pode ser utilizada em sistemas operacionais de previsão hidrometeorológica.

ABSTRACT

Forecasts of seasonal precipitation are available for the Brazilian northeast. These forecasts of precipitation have been used to forecast seasonal average discharge in the basins with discharge and precipitation observations. Then, there was the motivation for the development of a study for estimation of daily seasonal average discharge in basins without hydrometric observation. Statistical methods of correlation and multiple linear regressions were utilised, as well as the application of the technique of hydrologic regionalization, adapted to the estimation of annual values of the average discharge, in opposition to the estimation, usual in regionalization, of the long-term average discharge. It is proposed, in this study, a model based on the physical and hydrological characteristics of hydrographic basins, and soon after, this is applied to 13 basins inserted in the hydrographic basins of the rivers Piranha-Açu-Apodi, located in the semi-arid Northeastern Brazil. Good correlations between the observed data and the modelled values were obtained in validation tests, indicating that the proposed method is a technique that can be used in operational hydrometeorological forecasting systems.

SUMÁRIO

DEDICATÓRIA.....	i
AGRADECIMENTOS.....	ii
RESUMO.....	iii
ABSTRACT.....	iv
SUMÁRIO.....	v
LISTA DE TABELAS.....	vii
LISTA DE FIGURAS.....	viii
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA.....	3
2.1. REGIONALIZAÇÃO HIDROLÓGICA.....	5
2.2. PREVISÃO HIDROMETEOROLÓGICA.....	8
3. METODOLOGIA E DADOS.....	17
3.1. DESCRIÇÃO GERAL DA METODOLOGIA.....	17
3.2. SELEÇÃO, COLETA E PROCESSAMENTO DOS DADOS.....	22
3.2.1. Seleção da região.....	22
3.2.2. Dados descritivos da região.....	22
3.2.3. Dados hidrológicos (pluviométricos e fluviométricos).....	23
3.2.4. Características Físicas.....	24
3.2.4.1. Geoprocessamento.....	24
3.3. CONSTRUÇÃO DO MODELO.....	28
3.3.1. Avaliação preliminar.....	28
3.3.2. Montagem da série regional para o ajuste do modelo.....	31
3.3.3. Transformação das variáveis.....	33
3.4. AJUSTE DO MODELO.....	34
3.4.1. Análise de correlação.....	34
3.4.2. Análise de regressão.....	37
3.4.3. Análise do modelo linear com três variáveis.....	40
3.4.4. Montagem da equação.....	41

3.4.5 <i>Análise do intervalo de confiança das estimativas</i>	45
3.5. VALIDAÇÃO DO MODELO	46
3.5.1 <i>Método de validação cruzada</i>	47
3.6. ANÁLISE DOS RESÍDUOS	47
3.7. CASO DE ESTUDO	51
3.7.1. <i>Dados descritivos da região</i>	51
3.7.2. <i>Dados hidrológicos (pluviométricos e fluviométricos)</i>	54
3.7.3. <i>Características Físicas</i>	59
4. RESULTADOS	60
4.1. AVALIAÇÃO PRELIMINAR	60
4.1.1. <i>Divisão das séries de cada bacia</i>	60
4.1.2. <i>Análise do período chuvoso</i>	65
4.1.3. <i>Winsorização</i>	68
4.1.4. <i>Diagramas de dispesão bacia a bacia</i>	69
4.2. ANÁLISE DE REGRESSÃO	72
4.3. ANÁLISE DO AJUSTE E VALIDAÇÃO DO MODELO	76
4.4. ANÁLISE DOS RESÍDUOS	90
4.5 ANÁLISE DE REGRESSÃO SEM A BACIA DE SERRA NEGRA	100
5. CONCLUSÕES E RECOMENDAÇÕES	102
5.1 CONCLUSÕES	102
5.2 RECOMENDAÇÕES	103
REFERÊNCIA BIBLIOGRÁFICA	104
A. ANEXO	111
A.1. DADOS DA CALIBRAÇÃO DO MODELO	111
A.2. DADOS DO INTERVALO DE CONFIANÇA	117

LISTA DE TABELAS

Tabela 3.1 – Montagem da série de precipitação.....	32
Tabela 3.2 – Montagem da série de vazão.....	32
Tabela 3.3 - Quadro típico do SOC para uma regressão linear simples.....	42
Tabela 3.4 - Quadro típico do SOC para análise de variância em uma regressão linear múltipla....	43
Tabela 3.5 –Descrição das bacias dos postos fluviométricos.....	56
Tabela 3.6 - Séries sazonais de vazão e precipitação utilizadas.....	57
Tabela 3.7 – Características físicas das 13 bacias estudadas.....	59
Tabela 4.1 – Resultado da Winsorização.....	69
Tabela 4.2 – Matriz de correlação e nível descritivo do teste $H_0: \text{Correl}=0$, entre as variáveis do modelo.....	73
Tabela 4.3 – Estatística descritiva das variáveis para a construção do modelo.....	74
Tabela 4.4 - Análise de variância.....	74
Tabela 4.5 - Análise das estimativas para o modelo.....	74
Tabela 4.6 – Estatísticas para a análise do modelo.....	75
Tabela 4.7 - Resultados da validação cruzada.....	78
Tabela 4.8 – Matriz de correlação e nível descritivo do teste $H_0: \text{Correl}=0$, entre as variáveis do modelo.....	100
Tabela 4.9 – Estatística descritiva das variáveis para a construção do modelo.....	100
Tabela 4.10 - Análise de variância.....	101
Tabela 4.11 - Análise das estimativas para o modelo.....	101
Tabela 4.12 – Estatísticas para a análise do modelo.....	101
Tabela A.1 – Dados utilizados para a calibração do modelo.....	111
Tabela A.2 – Tabela do resíduo e intervalo de confiança.....	117

LISTA DE FIGURAS

Figura 3.1 – Esquema para a metodologia proposta.....	17
Figura 3.2 – Seleção, coleta e processamento dos dados para o modelo.....	19
Figura 3.3 – Construção do modelo.....	21
Figura 3.4 – Figura ilustrativa da aparência geral dos diagramas de dispersão associados com diversos valores de correlação (Kazmier, 1982).....	35
Figura 3.5 – Visualização do método dos mínimos quadrados plano em 3 dimensões numa regressão múltipla (Dunn & Clark, 1974).....	39
Figura 3.6 – Figura ilustrativa do intervalo de confiança.....	46
Figura 3.7 – Figura ilustrativa do resíduo.....	48
Figura 3.8 – Localização geográfica e bacias dos postos fluviométricos da bacia Piranhas-Açu-Apodi (UFPB/ANEEL, 2001).....	52
Figura 3.9 – Distribuição dos postos pluviométricos e fluviométricos selecionados (UFPB/ANEEL, 2001).....	55
Figura 4.1 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries com relação aos anos secos, intermediários e chuvosos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	61
Figura 4.2 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries analisando-se os períodos secos, intermediários e chuvosos e o número de dias com chuva das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	66
Figura 4.3 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda	

Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	70
Figura 4.4 – Vazão calculada versus vazão observada para todos os dados de todas as bacias na fase de ajuste do modelo.....	77
Figura 4.5 – Detalhe da Figura 4.4: vazão calculada versus vazão observada para todos os dados de todas as bacias na fase de ajuste do modelo.....	77
Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	79
Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	84
Figura 4.8 – Resultados da distribuição de frequência dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	91
Figura 4.9 – Resultados da distribuição cronológica dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	93
Figura 4.10 – Resultados da distribuição com valor calculado dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema e (m) Serra Negra.....	96
Figura 4.11 – Ajuste para todas as bacias estudadas, com exceção da bacia de Serra Negra.....	99
Figura 4.12 – Validação cruzada com a bacia de Serra Negra.....	99

1. INTRODUÇÃO

Recentemente muitos pesquisadores estão tentando explicar e prever o comportamento da precipitação, desenvolvendo sistemas de previsão meteorológica com uma semana a alguns meses de antecedência. Métodos estatísticos e modelos de circulação atmosférica permitem avaliar e prever o comportamento da precipitação acumulada na estação chuvosa do ano (“precipitação sazonal”) a partir dos meses que antecedem essa estação. Devido às substanciais variações temporal e espacial da precipitação, causando muitos problemas, a região nordeste do Brasil tem recebido considerável atenção de pesquisadores da área de previsão meteorológica e as informações obtidas estão disponíveis ao usuário. Para essa região, a maioria dos métodos produz previsões de um a três meses antes da estação chuvosa e se referem a todo o período chuvoso (3 ou 4 meses) e a toda a região. A previsão é anunciada pela primeira vez em novembro ou dezembro para o período de fevereiro – maio, a partir da situação dos preditores naquele momento, e a seguir atualizada mensalmente até o início de fevereiro.

Em função disso, há a possibilidade de se prever, com alguma antecedência, a vazão média diária sazonal em uma bacia hidrográfica. Se a bacia possui dados medidos de precipitação e vazão, pode-se construir uma relação (ou um modelo) entre essas variáveis para o desenvolvimento de tal estudo. Caso não haja dados de vazão medidos, o que é um fato em muitas bacias de açudes do nordeste brasileiro, há ainda a necessidade de transferência de informações de uma região onde se têm estes dados para bacias onde eles não estão disponíveis, a chamada regionalização hidrológica.

Com as estimativas de vazões na bacia, a partir das precipitações previstas pela meteorologia, seria permitido à sociedade fazer uma preparação de medidas preventivas contra eventos extremos, como cheias e secas, minimizando os efeitos negativos destes.

Um dos grandes problemas em recursos hídricos é a falta de dados hidrométricos e de conhecimento adequado da distribuição espacial e temporal da água, da variabilidade de usos e da disponibilidade hídrica. Há necessidade de permanente quantificação de

descargas líquidas, principalmente em locais sem monitoramento ou em locais onde existem dados constituindo séries de curta duração ou com períodos longos de falhas de observação. A implantação e a operação de postos hidrométricos numa bacia hidrográfica requerem alto investimento, técnicos capacitados para operacionalizá-los, além de vários anos de observação para se obter boa representatividade estatística.

Na falta dessas informações em determinados locais de interesse, a regionalização de uma variável, como neste caso a vazão, resolveria tal problema, já que a mesma envolve a determinação da variável desejada em diferentes locais de uma região hidrometricamente homogênea. O modelo de regionalização é formado pelas variáveis dependentes, por exemplo, a vazão, que são regionalizadas com base nas variáveis independentes, as características físicas da bacia: área, densidade de drenagem, declividade, comprimento do rio principal, fator de forma, dentre outras, e a característica hidrológica, a precipitação. O melhor conjunto de tais variáveis explicativas ou independentes será identificado para o modelo, de forma a melhor explicar a variável dependente. O melhor modelo será o que o for capaz de estimar a vazão de forma mais confiável e com o menor número de variáveis independentes.

Este trabalho tem por objetivo, avaliar uma metodologia para ajuste de equações para previsão de vazão média diária sazonal em bacias hidrográficas que não possuem dados hidrométricos, mas dispõem de informações pluviométricas e cartográficas. Estas equações poderão ser empregadas para prever a vazão média diária em uma bacia, para a próxima estação chuvosa, usando as previsões de precipitações disponíveis associadas às características físicas da bacia, que são obtidas com o apoio das técnicas de geoprocessamento. Assim sendo, há as seguintes metas a cumprir:

- i- Desenvolver um modelo de previsão da vazão média diária sazonal, para ser usado em bacias onde não se têm dados hidrométricos observados, e a este aplicar variáveis, de preferência de fácil obtenção e que estejam disponíveis, como a precipitação e as características físicas das bacias.
- ii- Avaliar o comportamento desse modelo através de uma aplicação em uma determinada região do semi-árido nordestino.

2. REVISÃO DE LITERATURA

O semi-árido brasileiro, com área de cerca de 1 milhão de km², é caracterizado, dentre outros aspectos, por uma acentuada variabilidade espaço-temporal de sua precipitação (400 a 1800 mm/a) e uma elevada taxa de evaporação (acima de 2000 mm/a), associada a condições geológicas restritivas (subsolo cristalino de reduzida potencialidade hídrica), tendo como consequência a apresentação de intermitência de seus cursos d'água (Freitas, sd).

Como na maioria das bacias e sub-bacias hidrográficas no Brasil, e ainda mais especificamente na região Nordeste, as séries históricas de vazões disponíveis são curtas, pouco representativas da distribuição espacial, e muitas vezes, perturbadas por açudes ou obras a montante da seção fluviométrica no curso da água em que a vazão é medida, ou em sub-bacia influente, faz-se necessário metodologias para determinação das séries pseudo-históricas através de duas maneiras:

1. Quando se tem dado de vazão: Se há registros limimétricos, ou linigráficos, e curva-chuva em determinadas seções de um curso d'água. Neste caso o estabelecimento de séries cronológicas de vazão, resulta simplesmente da conversão dos dados cronológicos de nível d'água em vazão pela curva-chave. Para isso, trabalha-se com uma análise estatística da série de vazão que pode implicar ou não no ajuste de uma distribuição de probabilidade estatística teórica (distribuição de Gumbel, por exemplo) aos dados. Outras vezes o uso da distribuição empírica satisfaz. (Silveira e Silveira, 2001).
2. Quando não se tem dado de vazão: Neste caso as alternativas são: (a) medidas isoladas de vazão de estiagem no local de interesse e uso de metodologias críticas que as aproveitem. Este método original privilegia a síntese da curva de permanência, mas, seu núcleo básico é a geração com um modelo chuva-vazão simplificado da série cronológica de vazões, aproveitando a informação de algumas

vazões medidas em seqüência na estiagem. A depleção de estiagem numa bacia hidrográfica é um retrato do seu comportamento após a ocorrência das precipitações e essa metodologia não exige instalações onerosas tradicionais de funcionamento contínuo (linígrafos), devido as amostras de vazão serem reduzidas, se resumindo a um ajuste dos parâmetros por tentativa e erro ou por otimização, gerando séries de vazão e o cálculo da curva de permanência; (b) Geração de séries cronológicas de vazão por modelos chuva-vazão, com ajuste dos parâmetros extrapolados regionalmente; e (c) estudos de regionalização de variáveis ou funções hidrológicas referentes às vazões desejadas.

Até alguns anos atrás os projetos de novas obras hidráulicas se constituíam a principal motivação para a regionalização hidrológica, mas, com os enfoques mais recentes de valorização do gerenciamento da água, destacam-se os subsistemas de outorga e cobrança pelo uso da água bruta, o licenciamento ambiental e a elaboração dos planos diretores de bacias, os quais, igualmente, necessitam estimar curvas de permanência de vazões, na maioria das vezes não disponíveis, este assumiu relevante importância no cenário da hidrologia no Brasil (Diniz e Clarke, 2001).

E hoje, nas regiões tropicais, os mecanismos oceânicos e atmosféricos que determinam a maior parte da variabilidade climática interanual já estão razoavelmente caracterizados, dando origem a diversos métodos de previsão de *longo prazo* da precipitação. A precipitação total acumulada na estação chuvosa (“precipitação sazonal”) têm-se atribuído influência de fatores atmosféricos e oceânicos de larga escala, como variações na temperatura na superfície dos oceanos. Métodos estatísticos ou outros tipos como, por exemplo, a regressão, os métodos multivariados, a análise de periodicidades, as redes neurais, entre outros, têm sido os mais empregados nesta tarefa (Galvão e Clarke, 1999).

Este trabalho se propõe a apresentar a aplicação de uma metodologia de regionalizar vazões através de previsões de precipitações, ou seja, realizar previsões de vazões para bacias sem dados hidrométricos a partir de previsões de precipitações, devido ao fato de já existirem estudos de previsões de precipitações sazonais e já existirem estudos onde se prevêem vazões sazonais a partir de previsões de precipitações sazonais para bacias com dados. A proposta é juntar a técnica de regionalização e usar as previsões para se prever vazões agora nas bacias onde não se têm dados. Para tanto, foi feita uma busca

do que está sendo divulgado dentro do assunto proposto, os conceitos gerais e algumas aplicações que estão sendo desenvolvidas, as ferramentas utilizadas e a qualidade dos resultados encontrados para as duas técnicas.

2.1. REGIONALIZAÇÃO HIDROLÓGICA

De acordo com Tucci (1993) a regionalização consiste num conjunto de ferramentas que exploram ao máximo as informações existentes, visando a estimativa das variáveis hidrológicas em locais sem dados ou com dados insuficientes. Esta estimativa pode ser na forma de uma variável, função ou parâmetro. E ainda, os métodos de regionalização são classificados em:

- i- Métodos que regionalizam os parâmetros da distribuição de probabilidades;
- ii- Métodos que regionalizam o evento com um determinado risco; e
- iii- Métodos que regionalizam uma curva adimensional de probabilidades, genericamente denominado de método da cheia-índice.

Uma regionalização envolve algumas fases tais como:

- i- Determinação das curvas de probabilidade individuais, parâmetros da distribuição, valores característicos, valor médio;
- ii- Curva adimensional regional: adimensionalização das curvas individuais com base na vazão média de enchente e metodologia para obtenção da curva média;
- iii- Equação de regressão;
- iv- Estimativa dos parâmetros para determinar a variância da estimativa;
- v- Critérios para a definição das regiões homogêneas;

Em qualquer dos métodos de regionalização é necessário a delimitação de regiões hidrológica e estatisticamente homogêneas, cujas estações tenham séries oriundas de populações regidas pela mesma distribuição de probabilidades e apenas seus parâmetros variando entre as estações. A metodologia de regionalização do 3º tipo mencionada acima, onde se regionaliza a curva adimensional de probabilidades e estabelece, para cada região homogênea, uma equação de regressão para o fator de adimensionalização (variável dependente), que é uma variável hidrológica de longo termo, com características fisiográficas e meteorológicas (variáveis explicativas ou independentes), facilmente

quantificáveis, como a área de drenagem e a precipitação total anual média de cada estação (Silva e Araújo, 2002), está sendo muito utilizada atualmente em virtude da facilidade de obtenção através da utilização dos softwares de SIG (Sistemas de Informação Geográfica ou GIS – Geographical Information System), do qual faz uso das técnicas de geoprocessamento, tornando mais precisas e ágeis as obtenções de tais variáveis explicativas.

Segundo Tucci e Clarke (2001) a regionalização envolve as seguintes etapas:

- i- Definição dos limites da área a ser estudada;
- ii- Definição das variáveis dependentes e explicativas da regionalização;
- iii- Seleção dos dados das variáveis;
- iv- Funções regionais: relações regionais e definição de regiões homogêneas.

A regionalização de parâmetros para a estimativa de vazões é um recurso que vem sendo utilizado com grande aceitação, pois permite inferir sobre a vazão de determinadas bacias que não tem registros ou os possuem de forma insuficiente. Além dessa forma de regionalizar, muitas outras estão sendo estudadas e vem sendo desenvolvidos ao longo dos anos como já citou Elesbon (2002): Wright (1970), que apresentou uma equação de avaliação das médias das vazões mínimas usando a declividade e a área da bacia como variáveis independentes; Riggs (1973) que sugere que sejam realizadas algumas medições na seção sem dados em um período de vazões mínimas; o *Institute of Hydrology* (1980) que desenvolveu um estudo regional de vazão mínima para a Inglaterra; Lanna et al. (1983) que estabeleceram curvas adimensionais de frequência para três regiões da bacia do rio Jacuí. E muitos outros estudos como o de Góes e Campana (2001), citado por Freitas et al. (2002), que incorporaram variáveis que retratam a dinâmica da paisagem em modelos de regionalização de vazões para o Distrito Federal e observaram um melhor ajuste dos modelos com o uso dessas variáveis, quando a bacia estudada tiver sofrido uma alteração na sua paisagem natural. O *NRC - U.S. National Research Council*, sugere modelos de regionalização que propiciam a substituição de tempo por espaço, com o objetivo de melhorar as estimativas de eventos extremos (Silva e Araújo, 2002).

Portanto, a regionalização de vazões é uma técnica utilizada para transferir informações espacialmente, buscando explorar ao máximo os dados disponíveis numa determinada área geográfica. Os estudos de regionalização que têm sido desenvolvidos, geralmente, utilizam as séries de vazões existentes. Então, quando estas informações são

representativas, os resultados são bons; no entanto, quando os dados são deficientes, a regionalização fica comprometida. Geralmente existem séries de precipitações mais longas que as de vazão, desta forma, em regiões com dados deficientes é possível estender as séries de vazões através de modelos hidrológicos precipitação-vazão e obter séries de vazões mais representativas para a regionalização. É o que estudou Obregon et al. (sd), onde utilizando um modelo mensal de transformação de precipitação em vazão e realizando a regionalização da curva de regularização e de permanência das vazões mensais aplicou, o resultado nas bacias afluentes à Lagoa Mirim, no Rio Grande do Sul. A metodologia de regionalização normalmente se baseia na obtenção de duas funções de caráter regional, a curva de frequência regional e a função regional para a estimativa de valores médios da variável hidrológica de interesse. Essa metodologia foi preconizada pela ELETROBRAS (1985) em seu manual "*Metodologia para Regionalização de Vazões*".

Eventualmente quando se fala em estudo de regionalização se pensa em trabalhar com regionalização de vazões máximas, médias e mínimas na qual, o estudo de vazão média de longo período permite estimar a potencialidade hídrica da bacia. A determinação das vazões máximas é de interesse no dimensionamento de obras hidráulicas, tais como, vertedouros, diques de proteção contra inundações, galerias pluviais e pontes. E quanto à vazão mínima é utilizada para avaliar a disponibilidade natural dos cursos d'água e a necessidade de regularização artificial.

Os modelos chuva-vazão visam, de uma maneira geral, a obtenção de séries de vazão mais extensas em locais onde se dispõem de grandes períodos de observações de dados pluviométricos e pequenos períodos de dados fluviométricos. Então muitos estudos procuram formas de se conseguir resolver esse problema com a busca de técnicas integradas como é o caso do trabalho de Diniz e Clarke (2001) que focaliza a regionalização dos parâmetros de um modelo chuva-vazão (SMAP) para 14 bacias hidrográficas localizadas na região semi-árida do nordeste brasileiro com o uso do algoritmo SCE-UA (*Shuffled Complex Evolution – University of Arizona*) desenvolvido por Duan et al. em 1992, onde o conjunto de parâmetros obtidos na calibração é associado às características das bacias através do uso de uma rede neural artificial do tipo perceptron de múltiplas camadas, treinadas com algoritmo de retropropagação e para cada bacia omitida do processo de calibração, a rede neural é usada para estimar os parâmetros do modelo para esta bacia omitida.

O uso de redes neurais para a previsão de dados hidrológicos, vazão e precipitação, nos seus mais diferentes usos (Galvão e Trigo, 1999), é hoje uma área bastante promissora. De modo geral, as redes neurais têm permitido obter resultados satisfatórios e melhores que os modelos estatísticos tradicionais. Ainda assim, essa é uma técnica relativamente pouco empregada neste contexto. Pois se observa que, um caso difere de outro, e não há uma referência segura. Nada garante que, uma rede que funciona para um local, será eficiente para outro com características semelhantes (Favoreto et al. 2001).

Também muito se tem estudado sobre a regionalização pluvial (ou regionalização da precipitação) que é um processo pelo qual se identifica regiões homogêneas com relação à determinada(s) característica(s) pluviométrica(s). Podem ser levados em conta os comportamentos das séries de cada localidade envolvida, como também, totais médios de períodos nunca inferiores a 30 anos (Jacon, 1982).

Logo, a regionalização da precipitação, quando realizada de forma integrada, isto é, considerando as suas variações interanuais, tanto pode revelar a forma de atuação dos sistemas geradores da precipitação na região, assim como oferecer subsídios no preenchimento de falhas e densidade da rede em cada sub-região. E neste contexto, as investigações do comportamento temporal e espacial da precipitação pode ser concebida através de técnicas de análise multivariada tais como: análise fatorial, componente principal e análises de agrupamento (Braga, 1992).

2.2. PREVISÃO HIDROMETEOROLÓGICA

A previsão de vazão é um dos principais desafios relacionados com o conhecimento integrado de climatologia e hidrologia. E a previsão de inundações, da umidade do solo para agricultura, os níveis de navegação de uma via, a disponibilidade hídrica para abastecimento de água, irrigação e produção de energia são os usos conhecidos para a previsão antecipada da vazão numa bacia hidrográfica (Tucci et al. 2002).

A previsão climática é peça importante em qualquer planejamento inteligente das atividades humanas nas escalas sazonais e interanual, e suas aplicações são as mais variadas, de acordo com cada país e dentro do mesmo país em cada região, onde estas devem ser úteis ao processo de tomada de decisão nas áreas da agricultura, geração de energia, armazenagem e utilização de água, como por exemplo: para um melhor

planejamento da produção de energia hidrelétrica e sua distribuição em sistema interligado; ou para um melhor planejamento contra as secas e enchentes, incluindo o controle do nível de reservatórios, racionamento na distribuição de água potável às populações e alocação de sementes apropriadas nos casos de secas (Moura, 1994).

Os resultados da previsão de chuva também são empregados nos modelos de previsão de inundação no qual a previsão de chuva geralmente é feita a cada 30 minutos e o horizonte de previsão para uma região, do volume esperado de precipitação nas próximas três horas. Ou seja, cobre três horas à frente, empregando-se para isso as imagens de satélite ou radar das nuvens coletadas (precipitações observadas) a cada cinco minutos. A previsão de vazão é feita em intervalos de tempo que dependem do tamanho da bacia e a previsão é revista sempre que a previsão de chuva é refeita, ou seja, a cada 30 minutos. Para o caso de inundação o modelo estabelece para cada região três tipos de estados hidrológicos: atenção, alerta e emergência. O estado de atenção indica a possibilidade de extravasamento nas próximas três horas, o estado de alerta indica que a inundação é eminente nas próximas três horas e o estado de emergência indica que a inundação já está ocorrendo (Barros, 2003). E os erros nas previsões quantitativas de chuva aumentam com o horizonte de previsão, e em consequência, também aumentam os erros na previsão de vazão à medida que se afasta das condições iniciais observadas.

Secas, geadas ou chuvas de granizo têm impacto significativo na produtividade agrícola, entretanto com as previsões na área agrícola, levará a emissão de alertas contra essas ocorrências com antecedência, para geadas de pelo menos 12 horas. Fornecimento de previsões meteorológicas e aconselhamento agrícola para as principais culturas econômicas melhorará as atividades agrícolas de preparo do solo, plantio, aplicação de produtos agroquímicos, irrigação, controle e disseminação de pragas e doenças, estimativas da produtividade, ponto de colheita, calendário agrícola e zoneamento ecológico das principais culturas além da avaliação de prejuízos causados às culturas (Costa, 2003).

Então para se planejar Recursos Hídricos em uma bacia hidrográfica é necessário conhecer a distribuição espacial e temporal da água. As variáveis hidrológicas mais importantes são a precipitação e a vazão, mas há muitas dificuldades em se contar com um período longo de observação, a obtenção da informação hidrológica suficiente e a estimativa do valor da vazão dos cursos d'água. Mas, tradicionalmente, com a aplicação de métodos estatísticos regionais, são aproveitados registros para a estimativa de frequência de locais com pouco ou nenhum registro hidrológico, permitindo uma estimativa mais

confiável da vazão (Miranda, 1997). Além do que, esta estimativa pode ser fruto de modelos de previsão climática a *curto prazo* (previsão em tempo real), *médio prazo* (alguns meses) e de *longo prazo* (geralmente anos).

De acordo com Tucci et al. (2002) a previsão de *curto prazo* é utilizada para estimar a vazão numa seção da bacia com antecedência de horas ou poucos dias, acompanhando a ocorrência do fenômeno que está para atingir ou já atingiu a bacia. Do ponto de vista hidrológico, a previsão de *médio prazo* depende fortemente das condições conhecidas de clima e dos oceanos que podem influenciar nas futuras vazões. A previsão de *longo prazo* é a estimativa, geralmente estatística, dos riscos de ocorrência de vazões num determinado local, como por exemplo: o risco de inundação numa seção de um rio, a probabilidade de períodos úmidos e secos, entre outros.

Têm-se discutido muito a prática da previsão climática de *curto prazo*, as previsões das variações nas escalas sazonal e interanual, e graças ao avanço no entendimento físico sobre o fenômeno El Niño-Oscilação Sul (ENSO) com suas fases quente (El Niño) ou fria (La Niña) e sua modelagem, através de modelos conceituais, simplificados e até globais por meio de CGCMs (*Coupled Ocean-Atmosphere Global Circulation Models*). Muitos países já reconhecem o valor e a necessidade destas e desenvolvem, em maior ou menor grau, arranjos institucionais para a utilização criteriosa das previsões efetuadas, ainda em caráter experimental e dentro desta perspectiva, a Fundação Cearense de Meteorologia e Recursos Hídricos (FUNCEME) tem produzido um dos melhores exemplos da combinação de ações a partir das previsões interanuais, atualizadas com destreza ao nível sazonal e mensal, e até a utilização das previsões de tempo da ordem de 5 dias e este parece produzir resultados muito animadores (Moura, 1994).

Segundo Rodrigues et al. (1996), alguns pesquisadores acreditam que o estabelecimento de um modelo operacional de prognóstico da *Estação Chuvosa*, fundamentado em base física e com previsões fornecidas com antecedência de alguns meses em relação àquela de maior precipitação, serviria ao propósito de se prevenir e adequar medidas emergenciais destinadas a minorar os efeitos da seca e enchentes tantas vezes devastadoras sobre a população e economia regional, principalmente, se este modelo for estabelecido por microrregiões pluviometricamente homogêneas, o que certamente, diminuiria a margem de erro dos prognósticos, uma vez que a dispersão dos dados seria mínima.

O próprio Rodrigues et al. (1996), sugere um trabalho nessa linha, onde para a determinação das microrregiões pluviometricamente homogêneas, ele utilizou a técnica da Classificação Hierárquica, também conhecida por Análise de Agrupamento e que tem por objetivo, evidenciar a existência de grupos homogêneos e heterogêneos dentro de uma população de dados.

Os núcleos de meteorologia e recursos hídricos do Nordeste do Brasil têm se reunido para interpretar as previsões feitas por órgãos internacionais e estes têm emitido a previsão de precipitação a partir da pré-estação chuvosa para a informação do abastecimento humano, com o monitoramento dos reservatórios, vendo o risco de colapso no armazenamento e suprimento hídrico destes, e para a agricultura tem simulado a umidade disponível no solo para as culturas no *curto prazo* e sugerido o momento de semeadura. Atualmente, estas informações estão sendo produzidas principalmente pelo Centro de Previsão de Tempo e Estudos Climáticos do Instituto Nacional de Pesquisas Espaciais – CPTEC/INPE, pelo *Hadley Centre for Climate Prediction and Research*, do serviço meteorológico britânico e pela Universidade de Wisconsin, nos Estados Unidos (Galvão, 1999).

Segundo Ramos e Galvão (2000), os modelos de circulação global (MCGs) apresentam a particularidade de fornecer seus resultados em diferentes escalas temporais, podendo ser sazonal, mensal e até mesma diária, apesar da redução da escala temporal de interpretação implicar num aumento do erro associado a essas previsões. Os MCGs são modelos computacionais complexos que descrevem as condições climáticas da Terra, baseando-se em códigos numéricos de formas aproximadas das equações que regem as leis físicas dos movimentos na atmosfera e nos oceanos, e as interações entre a atmosfera e a superfície. Esses modelos geralmente apresentam uma malha horizontal com resolução de 2 a 4 graus de latitude-longitude, e são divididos verticalmente em 10 a 30 camadas. Devido à baixa resolução horizontal da malha (equivalente a 2 a 5 centenas de quilômetros em regiões tropicais), variáveis relevantes à escala local (até alguns milhares de m^2) ou à escala da bacia hidrográfica (até cerca de 10 mil km^2) são mal representadas, ou mesmo ignoradas.

Outros pesquisadores como é o caso de Rocha et al. (2002), fazem previsão das vazões a partir da precipitação observada avaliando o método estatístico da correlação linear simples e múltipla para a previsão de vazão em rios e para isso, fazendo uso do balanço hídrico e aplicando o método de THIESSEN para estimar as precipitações médias

em pequenas bacias e analisar o regime pluviométrico e hidrométrico da bacia do rio estudado.

Já Silva (1988), desenvolveu um modelo probabilístico que permite estimar o total das chuvas de outono a partir do total precipitado durante o verão, com uma probabilidade de êxito de 80%. Também foi definido como estação chuvosa o período do verão e outono. E objetivando dar maior operacionalidade ao modelo de Silva (1988), Lopes et al. (1996) estudaram as microrregiões de precipitações homogêneas determinadas através da técnica de análise de agrupamento para a região do Rio Grande do Norte.

Ainda para o Rio Grande do Norte, Braga (1992) estudou a identificação de sub-regiões homogêneas a partir de valores médios decendiais de séries temporais da precipitação de no mínimo 30 anos de dados para 59 localidades deste Estado, também através do método da análise multivariada, em particular o de agrupamento hierárquico e análise de componentes principais.

Silva e Rao (1994), desenvolveram um estudo propondo a adoção de uma expressão algébrica para caracterização da forma de distribuição intra-anual da pluviosidade, como também a regionalização pluvial baseada no total anual médio de cada localidade e sua forma de distribuição intra-anual (de janeiro a dezembro mês a mês), aplicando-se tal metodologia a dados de 72 estações com mais de 30 anos de registros no Estado da Paraíba.

E pensando-se no problema da produção agrícola do semi-árido nordestino, também afetado pelas variações climáticas da região, que possui uma agricultura de subsistência, ainda rudimentar, Paixão et al. (1996) tentaram fornecer uma contribuição mais racional ao planejamento agrícola, em especial para o Estado do Ceará, usando técnicas baseadas em distribuições probabilísticas de chuva e da demanda hídrica que serviram de base para um possível programa de irrigação, sendo evidenciado neste caso, o cultivo do feijão macassar.

Guetter e Cunha (1998) desenvolveram um sistema de modelagem para prever o escoamento superficial para sub-bacias de cabeceira em função da previsão de chuvas para a bacia do Rio Iguaçu, produzindo previsões com resolução de meses a horas e os erros nas previsões quantitativas de chuva aumentando com o horizonte de previsão, e em consequência, também aumentando os erros na previsão de vazão à medida que se afasta das condições iniciais observadas, sendo que este atende aos critérios operacionais para

suporte ao gerenciamento de reservatórios. Estes afirmam que os princípios fundamentais para o desenvolvimento desses sistemas de previsões hidrológicas, são:

1. *Divisão em sub-bacias*: A subdivisão da bacia em sub-bacias menores e razoavelmente homogêneas em acordo com o relevo, disponibilidade de estações hidrométricas, e localização dos aproveitamentos hidrelétricos;
2. *Criação da base de dados*: Coleta de todos os dados disponíveis para a modelagem, sendo estes primeiramente consistidos;
3. *Organização da estrutura do modelo*: Aplicação da modelagem chuva-vazão para cada sub-bacia e para os segmentos que conectam as sub-bacias;
4. *Períodos para teste*: Os operadores do sistema definem a resolução temporal e os períodos para teste do sistema;
5. *Estimativa dos parâmetros*: Alguns parâmetros do modelo chuva-vazão-propagação têm grande variabilidade para uma grande bacia. Requerem-se testes iniciais para se estimar os parâmetros do modelo, antecedendo a operação do modelo de previsão em tempo real. Tanto métodos manuais quanto automáticos subsidiam a calibração dos parâmetros do modelo;
6. *Verificação do modelo*: Verificam-se como os resultados da simulação se comparam com as observações para diferentes opções, resolução do modelo, e períodos de teste (e. g., período úmido e período seco).

A disponibilidade hídrica efetiva numa bacia hidrográfica é constituída pela disponibilidade nos açudes e quando os rios são perenes pela disponibilidade a fio de água. A disponibilidade a fio d'água é estimada, primeiramente, a partir de séries de vazões medidas no curso d'água. Na falta destes, parcial ou total, utiliza-se, por exemplo, métodos alternativos como: geração sintética de séries de vazão a partir de modelos chuva-vazão (Silva e Lira, 2001), onde a previsão da vazão, com antecedência pode ser realizada a partir da precipitação observada, ou ocorrida na bacia ou nos níveis ou vazões observadas a montante do local de interesse, utilizando um modelo hidrológico que transforma precipitação em vazão, sendo realizada com antecedência de poucas horas ou dias.

Segundo Brum e Tucci (2001), os modelos chuva-vazão podem ser dos seguintes tipos: conceitual, empírico ou combinado. Os modelos empíricos utilizam formulações matemáticas sem nenhuma relação com o processo físico desta parte do ciclo hidrológico. A principal vantagem deste tipo de modelo reside na facilidade de atualização dos seus

parâmetros em tempo real. Contudo, a sua principal limitação decorre do empirismo da sua formulação, produzindo fortes incertezas quando é utilizado fora do intervalo dos dados em que foi ajustado. Os modelos conceituais, ao contrário dos anteriores, apresentam como vantagem a capacidade de reproduzir os processos hidrológicos de forma mais próxima dos processos físicos, permitindo melhor extrapolação de cenários diferentes do ajuste. A grande limitação dos modelos conceituais tem sido a dificuldade de atualização em tempo real de parâmetros e de variáveis estado.

Tucci e Clarke (1980) utilizaram o modelo IPH II com otimização automática dos parâmetros para atualização em tempo real destes, procurando evitar a limitação do modelo conceitual, mas os resultados não apresentaram ganhos significativos devido, principalmente, a qualidade das informações utilizadas. Cruz (1987) utilizou o modelo conceitual com parâmetros fixos e realizou a atualização no resíduo do modelo através do uso de um modelo empírico, portanto, uma solução híbrida. Os resultados foram promissores, mas pode introduzir distorções na simulação na medida que persiste o empirismo da atualização dos parâmetros. Bertoni et al. (1992) mostraram que uma análise cuidadosa dos parâmetros do modelo permite reduzir as limitações dos modelos conceituais na previsão em tempo real. Os autores definiram faixas de valores para os parâmetros e também testaram a previsão estatística da precipitação futura.

Tucci et al. (2002) se concentraram na comparação de modelos estatísticos utilizados atualmente na prática, com a combinação de modelos climáticos e hidrológicos determinísticos para a previsão de *médio prazo* (de até seis meses para as vazões afluentes), desenvolvidos para a bacia do rio Uruguai. E para o desenvolvimento deste utilizaram:

- i- *Modelo estatístico de longo prazo* onde este modelo analisa as vazões observadas no passado e utiliza como previsão as estatísticas das mesmas;
- ii- *Modelo estocástico*, sendo que este tipo de modelo utiliza as características temporais das séries hidrológicas para a previsão de *médio prazo*;
- iii- *Modelo empírico* o qual utiliza relações empíricas entre variáveis dos oceanos para previsão de *médio prazo* das vazões afluentes;
- iv- *Modelo climático-hidrológico* que é a utilização combinada de um modelo climático e hidrológico para a previsão das vazões de *médio prazo*.
- v- E ainda, as previsões com o *modelo hidrológico*, baseado na chuva prevista pelos modelos climáticos, os quais ocorreram em três etapas: A primeira etapa foi a

previsão baseada no uso das previsões climáticas diretamente, sem correção. A segunda etapa foi a previsão de vazões baseada em previsões de chuva corrigidas. E, finalmente, a terceira etapa foi a previsão de vazões com base nas previsões de chuva do modelo regional (*RAMS*), ou seja, do *downscaling* estatístico.

Este modelo de previsão de vazão, usado por Tucci (2002), assim como grande parte dos modelos de previsão de vazão, é um modelo hidrológico distribuído, que realiza a transformação de precipitação em vazão e, portanto, as previsões de vazão hidroclimáticas estão baseadas na aplicação de um modelo climático para a previsão de chuva e um modelo hidrológico para a previsão de vazão, a partir desta chuva. Estes modelos meteorológicos podem estimar previsões aceitáveis com antecedência de 72 horas (3 dias).

A formulação de um modelo estocástico-dinâmico de previsão hidrológica em desenvolvimento para a *COPEL* para suporte ao gerenciamento de reservatórios possui um sistema de equações diferenciais que produzem previsões com resolução de meses a horas. Além de vazões, o modelo estima a umidade média do solo na escala da sub-bacia (Guetter e Georgakakos, 1996), que é um elemento fundamental para o entendimento do efeito continental sobre a variabilidade climática e o componente estocástico minimiza os erros na previsão de vazões associados à incerteza nos dados de entrada passados e à imprecisão na estimativa dos parâmetros.

Até recentemente, a previsão de vazão em um horizonte de *curto prazo* (de algumas horas a alguns dias) era realizada através de modelos determinísticos, acompanhando os fenômenos climático-hidrológicos. Já em um horizonte de *longo prazo*, as previsões eram feitas através da estatística dos valores históricos, com a qual é possível prever os valores, mas não as datas de ocorrência. Com a evolução dos modelos climáticos e o uso de relações empíricas entre variáveis como a temperatura da superfície do mar, precipitação e escoamento, introduziu-se na literatura a chamada previsão de *médio prazo* (até seis meses) através de relações determinísticas, conceituais ou empíricas (Tucci et al. 2002).

Portanto, os sistemas operacionais de previsão hidrológica agrupam diversos componentes (precipitação, escoamento superficial e propagação de cheias) e requerem a descrição da topologia da rede natural de drenagem, tipos de solo, tipos de cobertura vegetal, e características dos canais, sendo representado de forma semi-distribuída, usando uma representação concentrada da transformação chuva-vazão na escala da sub-bacia, e um modelo hidrodinâmico para a propagação de vazões afluentes e originadas da sub-bacia. O sistema de modelagem prevê o escoamento superficial para cada sub-bacia de

cabeceira em função da previsão de chuvas, então o componente hidrodinâmico transforma o escoamento superficial em vazão efluente. As vazões previstas das sub-bacias de cabeceira alimentam as sub-bacias internas e soma-se ao escoamento superficial local para posterior propagação das vazões e determinação das vazões efluentes das sub-bacias internas, seguindo a topologia da rede de drenagem natural (Guetter e Cunha, 2003).

E a previsão de cheias em tempo real é um instrumento de suporte importante para gerenciar os recursos hídricos a *curto prazo* e para proteger vidas e bens materiais, durante a ocorrência de inundações. Enquanto que a previsão de *médio prazo* traz como benefício direto à previsão de afluência em reservatórios hidrelétricos e, como consequência prever a sua melhor operação e a formação de preço, principalmente no Brasil, em que a matriz energética é, predominantemente, baseada em energia hidráulica. Sendo a previsão de *longo prazo* muito importante para o planejamento dos recursos hídricos por considerar as variações sazonais e interanuais da sua disponibilidade. O que mostra a grande utilidade de tais estudos, onde especialmente no caso de secas e cheias excepcionais, dispor destas informações pode levar a preparação antecipada de medidas preventivas contra a calamidade prevista (Wilhite, 1991).

3. METODOLOGIA E DADOS

3.1. DESCRIÇÃO GERAL DA METODOLOGIA

A metodologia proposta neste trabalho para estimativa de vazão média diária sazonal para bacias sem dados hidrométricos a partir de previsões de precipitações possui duas fases (Figura 3.1): (a) Seleção, coleta e processamento dos dados e (b) Construção do modelo. Em cada fase, há várias etapas, descritas brevemente a seguir, e em maior detalhe nas próximas seções.

A primeira fase – Seleção, coleta e processamento de dados – consta das seguintes etapas (Figura 3.2):

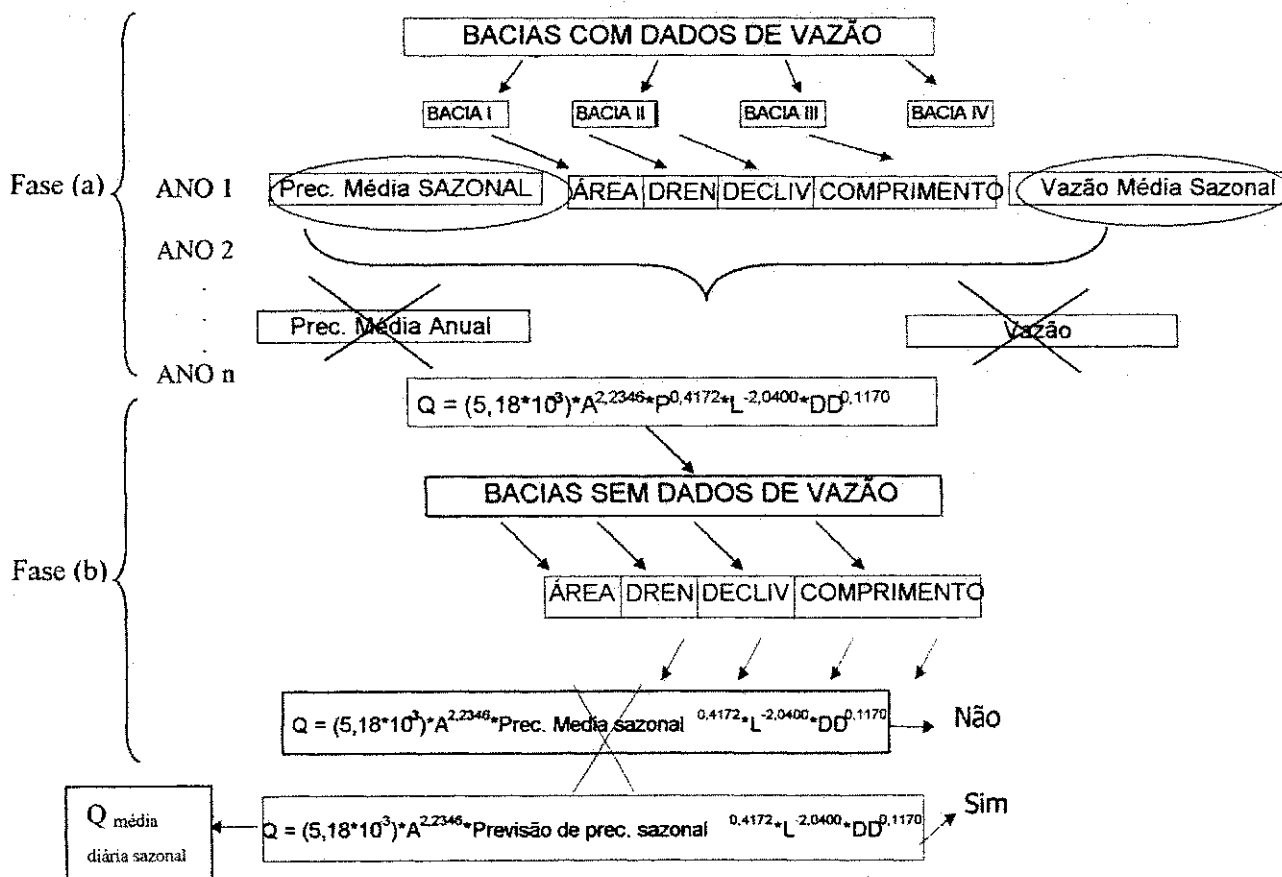


Figura 3.1 – Esquema para a metodologia proposta.

1. **Seleção da região** a ser estudada; que deve ser hidrologicamente homogênea;
2. **Coleta dos dados descritivos da região** (cobertura vegetal, relevo, formação geológica, pedologia, localização geográfica e política da região, rios principais e seus afluentes, declividade média dos trechos principais, caracterização climática, principais aproveitamentos hídricos e rede hidrométrica); para que se adquira familiaridade com a região em estudo, ou seja, se fazer uma **análise regional**;
3. **Coleta dos dados hidrológicos das bacias** (precipitação e vazão), após a identificação dos postos pluviométricos e fluviométricos localizados na região;
4. **Identificação das falhas** nos dados hidrológicos, para um preenchimento, se necessário, utilizando postos vizinhos com características climáticas semelhantes (método de ponderação regional, método de regressão linear, etc.);
5. **Análise de consistência** nos dados hidrológicos, procedimento posterior ao de identificação das falhas para se comprovar sua homogeneidade em relação a outros postos vizinhos (método da dupla massa, por exemplo);
6. **Construção das séries sazonais** dos dados hidrológicos, ou seja, a montagem da série de precipitação e vazão construídas para cada ano do total precipitado e das vazões médias diárias do período correspondente à estação chuvosa (ver item 3.3.2 deste – Tabelas 3.1 e 3.2);
7. **Período único**: Verificação da possibilidade de uso de um único período para as séries, o que é desejável. Caso as séries não apresentem período coincidente com a extensão adequada, serão utilizadas as séries completas de cada um dos postos;
8. **Seleção da série final**: devem ser identificados os postos de acordo com a qualidade da informação e sua distribuição geográfica;
9. **Coleta das características físicas das bacias com dados hidrometeorológicos disponíveis localizados na região** (área das bacias, densidade de drenagem, comprimento do rio principal, declividade do rio principal), extraídas de mapas em papel ou em meio digital. Caso estes mapas estejam em meio analógico (papel), estas podem ser digitalizadas e através da ferramenta de geoprocessamento extraídas as informações necessárias de cada bacia;
10. **Dados para construção/ajuste do modelo**: tabela com as séries de todas as bacias juntas, ano a ano, e suas características físicas para a **construção do modelo** de regressão linear múltipla apropriado.

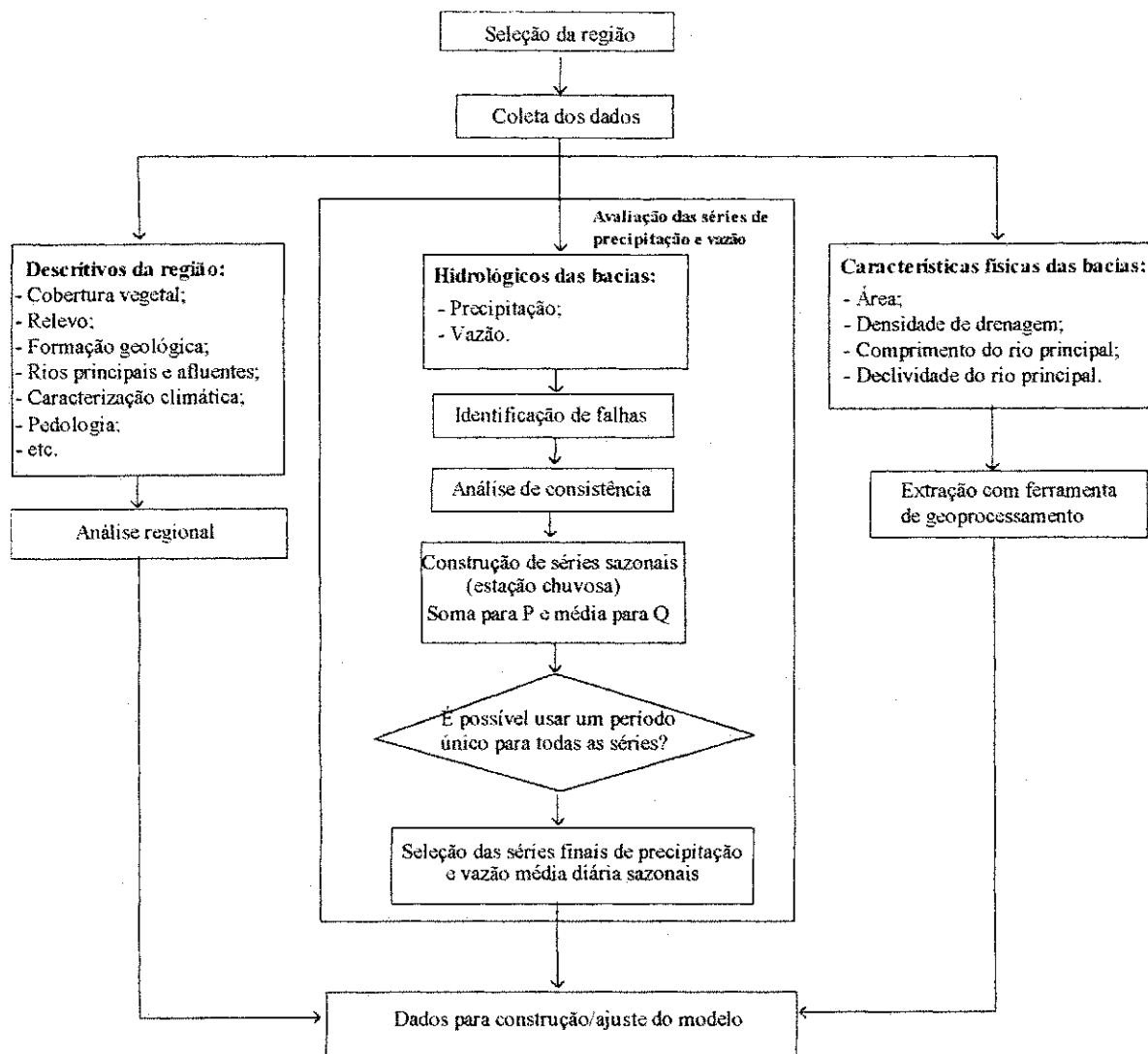


Figura 3.2 – Seleção, coleta e processamento dos dados para o modelo.

Na segunda etapa de **construção do modelo**, têm-se (Figura 3.3):

1. **Avaliação preliminar:** o diagnóstico bacia a bacia das relações anuais entre a precipitação e a vazão através de diagramas de dispersão para se obter uma melhor sensibilidade com os dados e ter uma prévia do melhor modelo (com maior coeficiente de correlação);
2. **Modelo único:** seria o melhor modelo encontrado na etapa anterior desde que este se ajuste bem a todas as bacias sem que haja perdas significativas com uma grande diminuição do coeficiente de correlação. Este modelo é desenvolvido com todas as bacias juntas, para ser aplicado em toda a região. Caso não haja um modelo único, terá

que ser ajustado um modelo para cada bacia diferentemente, correspondente ao modelo melhor ajustado por ela(s) na fase anterior;

3. **Montagem da série para o ajuste do modelo:** tabela contendo os dados de vazão, precipitação, área, densidade de drenagem, comprimento e declividade do rio principal e demais características físicas das bacias de todos os anos de todas as bacias que será a série utilizada para o ajuste do modelo;
4. **Transformação das variáveis (ANAMORFOSE),** para eventual ajuste de modelos não lineares utilizando o algoritmo de regressão linear;
5. **Montagem da equação, através da análise de regressão:** a equação terá uma variável dependente, a vazão média diária, e as variáveis independentes que são a precipitação sazonal e as características físicas das bacias dos postos fluviométricos selecionados;
6. Cálculo do **intervalo de confiança** da previsão utilizada e a equação ajustada;
7. **Volta das variáveis ao seu formato original,** não linear;
8. **Análise dos resíduos,** para verificar a adequação do modelo, analisando o comportamento dos resíduos entre cada valor observado (y_i) e o respectivo valor previsto pelo modelo (\hat{y}_i), pela distribuição de frequência, cronológica e com relação ao valor previsto (\hat{y}_i);
9. **Validação cruzada:** partindo do princípio que se quer futuras estimativas da vazão baseados na equação proposta e em valores medidos, há a necessidade de se avaliar qual a confiança deste resultado. É um método bastante simples onde para um conjunto de N observações é necessário que se divida tal amostra em k subconjuntos diferentes com tamanhos (aproximadamente) iguais, sendo em seguida aplicado o modelo sempre omitindo um dos subconjuntos, onde este subconjunto omitido será utilizado para a validação do modelo encontrado. Este processo é feito para todas as bacias.

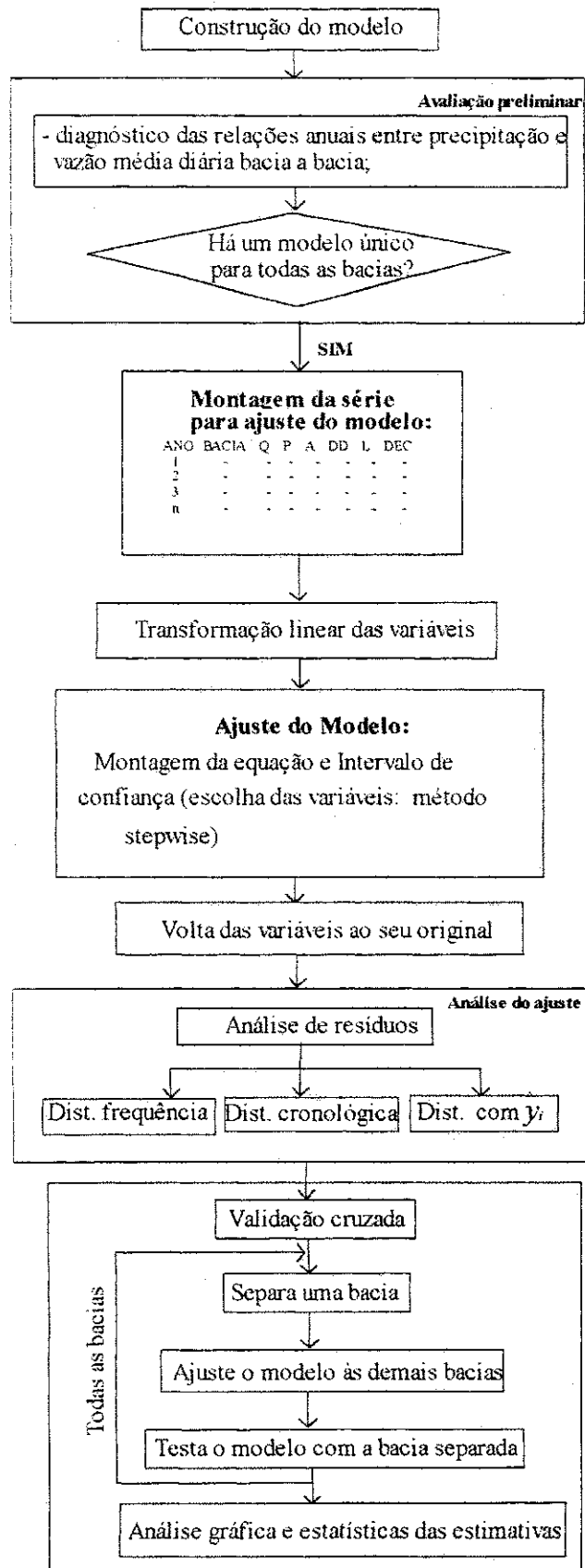


Figura 3.3 – Construção do modelo

3.2. SELEÇÃO, COLETA E PROCESSAMENTO DOS DADOS

3.2.1. SELEÇÃO DA REGIÃO

A escolha da região para a aplicação da metodologia, bacias hidrográficas dos rios Piranhas-Açu-Apodi, se deu principalmente por:

1. Ser uma região homogênea e típica do semi-árido do Nordeste;
2. Possuir algumas bacias hidrográficas de postos fluviométricos inseridas nesta, (Antenor Navarro (NA), Aparecida (AP), Augusto Severo (AS), Emas (EM), Fazenda Angicos (FA), Mossoró (MO), Pau dos Ferros (PF), Pedra de Abelhas (PA), Piancó (PI), São Fernando (SF), Sitio Volta (SV), Upanema (UP) e Serra Negra (SN)), com dados observados e com problemas típicos de outras bacias no Nordeste;
3. Ser uma bacia já trabalhada com estudo de regionalização.

3.2.2. DADOS DESCRITIVOS DA REGIÃO

Para a obtenção de um modelo que representa as informações espaciais e temporais, ou seja, de um modelo espaço-temporal que trata dos processos ambientais, deve-se fazer uma análise geral das informações das características da região em estudo, haja vista a complexidade dos processos envolvidos.

Portanto, faz-se necessário o levantamento das características de cobertura vegetal, relevo, formação geológica, rios principais e afluentes, caracterização climática, pedologia, dentre outros, para a extração de conclusões ou para uma melhor compreensão do modelo.

A obtenção de tais dados pode ser realizada com o uso de ferramentas de geoprocessamento, como visto mais a seguir.

3.2.3. DADOS HIDROLÓGICOS (PLUVIOMÉTRICOS E FLUVIOMÉTRICOS)

Os dados pluviométricos e fluviométricos utilizados neste trabalho foram obtidos do banco de dados nacional da ANEEL, onde conta neste banco de dados, um total de 28 postos fluviométricos e 162 postos pluviométricos inseridos na bacia dos rios Piranhas-Açu-Apodi, região escolhida para a aplicação deste estudo.

Um estudo anterior (UFPB/ANEEL, 2001) já havia feito a análise das séries para a mesma região, identificando falhas e aplicando teste de estacionariedade (teste de Wilkinox) às séries dos postos que apresentavam açudes de maior porte localizado nas suas bacias. Então, para este estudo foi usado o resultado de tal análise.

Para a pluviometria, tal estudo obedeceu aos seguintes critérios para seleção dos postos: foram escolhidos aqueles que possuíam pelo menos dez anos completos de dados e/ou se localizavam geograficamente representando uma boa cobertura da região estudada. Mas para a utilização destes postos, neste estudo, deveria ainda ser feito o cálculo dessas precipitações ao nível sazonal, onde estas são obtidas com as somas dos valores diários dos meses de Fevereiro a Maio de cada ano.

Para a escolha final dos postos fluviométricos também no estudo UFPB/ANEEL (2001), foram analisados os seguintes critérios:

- 1 – “Exame da curva de descarga e análise de continuidade de volumes, com base no relatório de consistência da CPRM;
- 2 – Representatividade da série em relação ao regime hidrológico da região (seqüência de anos secos e chuvosos);
- 3 – Análises dos valores dos coeficientes de escoamento e vazões específicas;
- 4 – Análise dos períodos de falhas;
- 5 – Resultados dos testes de estacionariedade;
- 6 – Fichas de avaliação dos postos, com base no relatório de consistência da CPRM”.

A qualidade das séries, de um modo geral, deixa a desejar por possuir muitas falhas, muitas repetições de registros e interrupções, tornando as séries mais curtas que a duração desejada, que seria um mínimo de 10 anos completos de dados para os postos pluviométricos e 05 anos para os postos fluviométricos, idealizando-se uma série com

períodos iguais e sem interrupções de vazão e precipitação. O que não se mostrou diferente neste estudo.

Logo, não foi considerado um período comum para as séries das bacias estudadas, para evitar perda de dados em séries já relativamente curtas; não se preencheu falhas; nas lacunas nas séries de vazões diárias foram desprezados os dados não registrados, levando-se em consideração apenas os dados disponíveis naquele ano; e foram observados os anos de dados incompletos com os anos completos do mesmo posto e de postos vizinhos, para verificar se as lacunas existentes fazem parte de meses secos, caso dos dados restantes terem sido aproveitados.

3.2.4. CARACTERÍSTICAS FÍSICAS

Além da qualidade dos dados pluviométricos e fluviométricos, analisou-se a qualidade das demais fontes necessárias para se fazer uso em tal estudo, que são as características físicas das bacias.

A rede de drenagem e as curvas de nível, de onde são extraídas tais características, geralmente apresentam significativa variabilidade em densidade e resolução, de carta para carta, provavelmente em virtude das variações da qualidade dos dados básicos empregados na sua derivação, introduzindo incertezas quando da utilização destas variáveis na formulação das funções de regionalização e na sua posterior aplicação.

As extrações das características físicas neste estudo foram feitas em meio digital através de software de geoprocessamento. Nesta etapa, para a caracterização física da região foram utilizadas cartas na escala de 1:100.000, fornecidos pela SUDENE, a partir das quais foram digitalizadas para em seguida serem calculadas as características físicas das bacias, através das ferramentas do software.

3.2.4.1. GEOPROCESSAMENTO

Os softwares de geoprocessamento têm sido muito utilizados para a obtenção das características físicas das bacias na regionalização. Geoprocessamento nada mais é que o uso automatizado de informação que de alguma forma está vinculada a um determinado lugar no espaço, seja por meio de um simples endereço ou por coordenadas, capaz de gerar a informação mais rápida e precisa. Vários sistemas fazem parte do Geoprocessamento

dentre os quais o SIG - Sistemas de Informação Geográfica (ou GIS - *Geographical Information Systems*) é o sistema que reúne maior capacidade de processamento e análise de dados espaciais. A utilização destes sistemas produz informações que permitem tomar decisões para colocar em prática ações, a partir de bases cartográficas digitais.

As bases cartográficas digitais consistentes de qualquer mapeamento temático, segundo Cauvin e Rimbart (1976), torna-se fonte importante para qualquer tipo de planejamento ou mesmo para implementação de projetos de infra-estrutura urbana e regional. Tais bases podem nos dá informações relativas à morfologia da superfície do terreno, entre elas a declividade média da bacia ou de um rio, a rede de drenagem, os divisores de água da bacia ou de um conjunto de sub-bacias, o comprimento e grau de concavidade ou convexidade de uma vertente, o tempo de concentração e os parâmetros de forma utilizados em hidrologia.

O SIG também pode ser usado na obtenção de isolinhas através da interpolação de valores, que pode ser desenvolvido através de vários métodos, sendo muito utilizado a autocovariância espacial ou *Kriging*, do qual se permite criar mapas que pode facilitar a visualização de um atributo qualquer ao longo de uma superfície. Tal procedimento é muito usado para se mapear a interpolação da vazão em bacias hidrográficas.

Uma implementação em SIG é importante para a automatização dos procedimentos e ter como suporte na tomada de decisão. Numa bacia hidrográfica há uma infinidade de problemas ambientais resultado de inúmeros fatores como a agricultura, a indústria e mesmo o próprio homem, portanto, há uma necessidade de avaliar ou quantificar tais fatores, já que esse vai interferir nos componentes do ciclo hidrológico (equilíbrio ambiental) de uma bacia hidrográfica. O geoprocessamento permite hoje se ter uma visão detalhada em todas as posições no espaço dessa realidade e relacionar atributos tabulares, informações, com os mapas, já que estas são as principais características de um SIG: a aquisição ou armazenamento da informação, a manipulação e a utilização dessa informação, o que resulta em um extenso banco de dados referenciados geograficamente contendo as informações necessárias ao modelo que se está utilizando, como também para a análise e comparação dos vários resultados.

Portanto, de posse dessa ferramenta, o geoprocessamento, que lida com a informação espacial, e a utilização de um modelo matemático para a análise temporal, a hidrologia teria um ganho excepcional e, essa união é algo evidente pelos hidrólogos hoje. Como escreve Mendes e Cirilo (2001) que o geoprocessamento permite a variabilidade

espacial de forma bastante efetiva, mas existem dificuldades de acompanhamento da dinâmica temporal e a maioria dos modelos matemáticos utilizados em recursos hídricos considera a variabilidade temporal, logo a união entre as duas ferramentas surge naturalmente. Nada que uma linguagem de programação com a funcionalidade de um programa de geoprocessamento e a dinâmica do SIG não resolva.

Através de ferramentas de geoprocessamento foram calculadas neste estudo as características físicas das bacias dos postos fluviométricos selecionados: área de drenagem, comprimento do rio principal, densidade de drenagem, declividade do rio principal, etc., que são os dados físicos, de modo a serem usadas como variáveis explicativas ou independentes na análise de regressão da variável vazão.

Nesta etapa, para a caracterização física da região foram utilizadas cartas na escala de 1:100.000, fornecidas pela SUDENE, as quais foram digitalizadas, através de uma mesa digitalizadora, editoradas e tendo uma separação das diferentes entidades em camadas, de forma a facilitar a obtenção das informações; daí então, foram localizados os postos fluviométricos, através de suas coordenadas geográficas e delimitados manualmente suas bacias hidrográficas, utilizando como referência a rede de drenagem e a topografia do terreno; para em seguida serem calculadas as características físicas de tais bacias, através das ferramentas de software de análise que efetua medições como distâncias, comprimentos, áreas e volumes.

Vale ressaltar que, as variáveis independentes envolvidas nos modelos devem ser de fácil obtenção, para que os modelos possam ser utilizados pelos mais variados usuários, na tentativa de quantificação de vazões em cursos d'água de bacias hidrográficas dentro da região considerada hidrológicamente homogênea. Em função disso, apenas variáveis explicativas como o comprimento do rio principal, que nada mais é do que o comprimento total do caminho (canal principal) que a água percorre até atingir o exutório da bacia drenando a maior área no interior da bacia; a declividade de trechos de rios, que é a declividade média do curso d'água principal apresentando influencia direta na velocidade do fluxo (Paiva e Paiva, 2001); as áreas das sub-bacias contribuintes, que são obtidas fazendo-se o traçado do rio principal, iniciando-se da foz até a nascente e em seguida seus afluentes, de forma que ao final tem-se a rede de drenagem da bacia e contornando-se estes, através da ligação dos pontos externos de maiores cotas e dos divisores topográficos da bacia, tem-se a delimitação desta; a densidade de drenagem, que refletem controles topográficos, litológicos, pedológicos e de vegetação, além da influencia antrópica (Paiva e

Paiva, 2001), sendo ela a soma de todos os comprimentos dos rios da bacia, divididos por sua área total; e a precipitação, que é toda a água proveniente do meio atmosférico que atinge a superfície terrestre (Bertoni e Tucci, 2000), foram utilizadas neste trabalho.

No entanto, as variáveis associadas às características e ao uso e ocupação do solo, que são os dados descritivos das bacias podem melhorar a regionalização pretendida, já que a quantidade e qualidade dos recursos hídricos estão intimamente ligadas ao uso do solo na bacia. Também essas características são mais facilmente obtidas através das ferramentas de geoprocessamento.

Com a série definida e as características físicas das bacias extraídas do geoprocessamento, passou-se à etapa de cálculos estatísticos e a construção do modelo.

3.3. CONSTRUÇÃO DO MODELO

3.3.1. AVALIAÇÃO PRELIMINAR

A bacia hidrográfica é o elemento fundamental de análise de um estudo hidrológico e é a "área de captação natural da água da precipitação que faz convergir os escoamentos para um único ponto de saída, seu exutório", ponto de medida da vazão de saída ou volume escoado de uma bacia hidrográfica. Ou seja, a bacia hidrográfica pode ser considerada um sistema físico capaz de transformar uma entrada de volume concentrada no tempo (precipitação) em uma saída de água (escoamento ou volume de água escoado - vazão) de forma mais distribuída no tempo, considerando-se como perdas intermediárias os volumes evaporados e transpirados e também os infiltrados profundamente. Para um evento isolado pode-se desconsiderar estas perdas e analisar apenas a transformação de chuva em vazão feita pela bacia (Silveira, 2000).

Por isso a importância de se conhecer bem o comportamento da vazão e da precipitação em estudos hidrológicos. No caso deste estudo, muitas análises iniciais foram feitas, bacia a bacia, com a finalidade de se obter uma certa sensibilidade com os dados de entrada (precipitação) e saída (vazão) e com a região estudada, já que estes seriam o embasamento teórico para geração do modelo proposto.

As séries de vazão e precipitação utilizadas devem ser independentes, representativas e homogêneas (Tucci e Clarke, 2001). São independentes quando não existe correlação serial entre os mesmos, representativas quando a amostra tem poucas incertezas e as estatísticas da amostra pouco difere das estatísticas da população e são homogêneas quando as estatísticas não mudam com o tempo, ou seja, não ocorreram alterações na bacia ou no clima que alterem a tendência aleatória dos valores.

Deve-se fazer ainda uma análise com relação à escolha das séries, na qual duas abordagens podem ser utilizadas: usar um período comum, ou seja, series coincidente, com

a desvantagem da perda de informações, já que parte da série será descartada para se obter um período comum; ou não utilizar período comum algum e sim toda a série disponível sem falhas. Essa segunda opção foi a adotada por este estudo.

Foram considerando apenas os meses de fevereiro a junho para a série de vazão e de fevereiro a maio para a série de precipitação, meses estes considerados chuvosos para a região estudada.

Inicialmente foram calculados a média das médias anuais, o desvio padrão e as anomalias das séries, numa tentativa de se trabalhar com a média dos dados como normalmente se faz e verificar se os dados normalizados fariam alguma diferença. Estas anomalias (método usado em meteorologia para testar se a chuva está acima ou abaixo da média) são calculadas subtraindo-se de cada valor individual da série anual a média de todos os anos juntos para cada posto e o resultado dividindo-se pelo desvio padrão da mesma série.

Uma outra análise desenvolvida neste estudo foi a divisão da série em três períodos: anos chuvosos, anos intermediários e anos secos. Estabelecendo-se como anos chuvosos, os anos que superassem $(P+\sigma/2)$, anos secos aqueles totais que fossem inferior a $(P-\sigma/2)$ e anos intermediários os totais compreendidos no intervalo $(P-\sigma/2)$ e $(P+\sigma/2)$, sendo P a precipitação média sazonal anual e σ o desvio padrão sobre toda a amostra de dados para cada posto.

De acordo com Matos (1995), algumas outras alternativas poderiam ser usadas, como: variáveis centradas (*centered*), uma transformação simples que consiste em centrar cada variável em relação à sua média, onde a variável transformada m_k é obtida através de:

$$m_{ik} = x_{ik} - \bar{X}_k$$

outra forma seria a standardização (*standardized*) que corresponde a uma transformação para média nula e desvio padrão unitário de cada variável original X_k , onde

$$z_{ik} = \frac{x_{ik} - \bar{X}_k}{s_k}$$

uma nova variável Z_k é obtida através de: ou com norma unitária (*unit lenght*) que substitui os valores de cada variável X_k por uma nova variável, obtida da

$$w_{ik} = \frac{x_{ik} - \bar{X}_k}{d_k}$$

seguinte maneira:

Mais uma tentativa foi testada neste estudo que é a *Winsorização* ou a eliminação de extremos das séries, aplicada principalmente quando se têm dados com valores muito acima ou muito abaixo dos demais. Neste caso, tais dados são sanados e é visto se há

mudança de comportamento nos resultados. Isso é feito porque tais valores podem dar uma tendência aos resultados diferentes da realidade.

Todos estes procedimentos visam compatibilizar, de algum modo, variáveis que podem ter escalas e dispersões muito diferentes, a fim de se verificar a comparação da influência relativa das diversas variáveis, com base nos parâmetros estimados.

Feito isso, partiu-se para um outro tipo de análise ainda em cada bacia separadamente, o comportamento da vazão com a precipitação através de diagramas de dispersão, que nada mais é do que a representação dos pares de valores num sistema cartesiano (Bussab e Morettin, 1987), onde este diagrama fornece uma idéia da forma da relação exibida através dos pares de valores. A intenção disso era encontrar uma resposta para o porquê da análise anterior não apresentar bons resultados e o que fazer então.

Num diagrama de dispersão, para cada pares de valores, faz-se a análise da reta melhor ajustante, e uma vez observada uma associação entre as variáveis quantitativas, é muito útil quantificar essa associabilidade onde o tipo mais simples é a linear. Era esperada, então, essa relação aproximadamente linear entre a vazão e a precipitação na escala anual, haja visto que em uma bacia hidrográfica, ao analisarmos seu balanço hídrico, desconsiderando que houve armazenamento e sabendo-se que as perdas são constantes numéricas, a vazão seria praticamente igual à precipitação e as bacias teriam comportamentos hidrológicos semelhantes. Então:

$\Delta S = -Q - \text{Perdas} + P$, sendo ΔS a variação do armazenamento; Q a saída da bacia (vazão); Perdas as evapotranspirações, evaporações, infiltrações, etc; e P a entrada na bacia (precipitação). Ou seja:

$$Q = P - \text{Perdas} \pm \Delta S$$

\nearrow cte
 \nearrow ≈ 0

Portanto se $Q_1 = a_1P + b_1$; $Q_2 = a_2P + b_2$; ...; $Q_n = a_nP + b_n$, então:

$$a_1 \cong a_2 \cong \dots \cong a_n$$

Mas caso isso não aconteça, as bacias podem apresentar uma função não *Linear*, apresentando assim uma função *Exponencial*, *Potência*, ou outro.

Um tipo *Linear*, para duas variáveis X e Y segundo o programa EXCEL[®], é o que calcula o ajuste por mínimos quadrados para uma linha representada pela seguinte equação: $y = ax + b$, onde m é a inclinação da reta do ajuste e b é a interseção com o

eixo dos Y . A linha de tendência linear geralmente mostra que algo está aumentando ou diminuindo com uma taxa fixa e seus dados serão lineares se o padrão nos pontos de dados se parecer com uma linha reta.

Já a função *Exponencial*, o EXCEL® considera quando este calcula o ajuste também por mínimos quadrados através de pontos usando a seguinte equação: $y = ce^{bx}$, onde c e b são constantes, e e é a base do logaritmo natural. A linha de tendência exponencial é uma linha curva muito útil quando os valores de dados estiverem crescendo ou caindo com taxas cada vez mais altas.

Enquanto que a função conhecida pelo EXCEL® como *Potência* calcula o ajuste por mínimos quadrados através de pontos usando a seguinte equação: $y = cx_k^b$, onde c e b são constantes. A linha de tendência de potência é uma linha curva melhor usada com conjuntos de dados que comparam medidas que aumentam a taxas específicas — por exemplo, a aceleração de um carro de corrida em intervalos de 1 segundo. No caso deste estudo, o melhor comportamento foi esse tipo *Potência*.

Todo este estudo preliminar serve para que se tenha um parâmetro inicial que possibilite julgar as diferenças entre os níveis de ajustes nas bacias e identificar possíveis problemas com os dados. Com esse diagnóstico bacia a bacia, veio a decisão de se adotar como modelo único o modelo *Potência*, em virtude de este ter apresentado um melhor resultado nessa etapa tão fundamental do estudo.

3.3.2. MONTAGEM DA SÉRIE REGIONAL PARA O AJUSTE DO MODELO

Normalmente tem-se conseguido relacionar Q e P de todos os postos em uma única equação, incluindo as características físicas das bacias, quando P e Q são as médias de longo período, isso suaviza muito o resultado. No caso deste estudo, houve uma preocupação de se trabalhar ano a ano, todas as bacias juntas, num regime de sazonalidade onde foram escolhidos apenas os anos que não apresentaram falhas agravantes.

Foram selecionadas, todas as estações fluviométricas com séries de vazão gerada, superiores a pelo menos cinco anos completos de dados. E para as estações pluviométricas foram selecionados os postos com pelo menos dez anos completos de dados e analisou-se ainda a localização geográfica dos postos, para se obter boa cobertura sobre a região.

Analisando os dados, pôde-se observar que além das interrupções nas séries, havia também uma grande heterogeneidade; ou seja, dados em áreas pequenas possuindo valores de vazão menores do que numa outra bacia com área maior. Por isso, optou-se por juntar as séries de todas as bacias ano a ano quando na montagem da série de ajuste para o modelo. O procedimento usado para isso, ou seja, a escolha dos dados de precipitação e vazão para o caso deste estudo, pode ser melhor esclarecido ao observar as Tabelas 3.1 e 3.2 abaixo, onde foram usados os totais para as precipitações e as médias para as vazões. Estas tabelas mostram bem a forma de separação das séries de precipitação e vazão, além destes dados, para compor a tabela para o ajuste do modelo.

Tabela 3.1 – Montagem da série de precipitação.

<i>Postos Pluv.</i>	<i>Data (mês/ano)</i>	<i>Chuva dia 01</i>	<i>Chuva dia 02</i>	<i>.....</i>	<i>chuva dia 31</i>	<i>Soma</i>
B	fev/X	1	2	3	-
B	mar/X	4	5	6	-
B	abr/X	7	8	9	-
B	mai/X	10	11	12	-
						Total B no ano X
C	fev/X	1	2	3	-
C	mar/X	4	5	6	-
C	abr/X	7	8	9	-
C	mai/X	10	11	12	-
						Total C no ano X

Tabela 3.2 – Montagem da série de vazão.

<i>Postos Fluv.</i>	<i>Data (mês/ano)</i>	<i>Vazão dia 01</i>	<i>Vazão dia 02</i>	<i>.....</i>	<i>Vazão dia 31</i>	<i>Média</i>
B	fev/X	1	2	3	-
B	mar/X	4	5	6	-
B	abr/X	7	8	9	-
B	mai/X	10	11	12	-
B	jun/X	13	14	15	-
						Média B no ano X
B	fev/Y	1	2	3	-
B	mar/Y	4	5	6	-
B	abr/Y	7	8	9	-
B	mai/Y	10	11	12	-
B	jun/Y	13	14	15	-
						Média B no ano Y

Todo o processamento e a análise desses dados foram desenvolvidos com o auxílio dos programas ACCESS 2000[®] - software da Microsoft[®] para criação e manipulação de banco de dados, inicialmente com a seleção das séries, depois com o EXCEL 2000[®] - software também da Microsoft[®], no diagnóstico das relações de precipitação e vazão e uma escolha preliminar do modelo através do melhor ajuste, e por fim com o SOC - software de análise estatística desenvolvido pela EMBRAPA, num tratamento mais refinado da análise estatística e escolha da melhor equação.

Além das séries de precipitação e vazão foram usadas ainda para o ajuste do modelo, as características físicas das bacias dos postos fluviométricos, tais como: área da bacia, declividade do rio principal, densidade de drenagem e comprimento do rio principal. Características estas obtidas através das técnicas de geoprocessamento.

3.3.3. TRANSFORMAÇÃO DAS VARIÁVEIS

Se a relação entre as variáveis puder ser expressa por uma equação de primeiro grau, temos uma regressão linear, mas há numerosos casos em que as equações de regressão não são lineares e segundo Gomes (1978), estas podem ser consideradas lineares por *ANAMORFOSE*, isto é, por substituição dos valores de uma variável por uma função e evidentemente, pode-se dizer o mesmo para com duas ou mais variáveis, se necessário. Por exemplo: a equação desenvolvida neste estudo foi baseada numa equação do tipo $Z = a b^x$ onde por aplicação de logaritmo nos dá $\ln Z = \ln a + x \ln b$, se $y = \ln Z$, $M = \ln a$ e $N = \ln b$, teremos a equação linear $y = M + N X$, cujos parâmetros podem ser estimados pelo método da regressão linear simples. Esta função $Z = a b^x$ é conhecida na economia como função *COBB-DOUGLAS* e neste estudo chamamos de função potência por ser assim chamada pelo programa Excel[®], de uso muito comum na área de Engenharia.

Como o modelo escolhido inicialmente foi o Potência e o programa SOC trabalha bem com a regressão linear simples, houve a necessidade de uma logaritmização dos dados, o que os tornou “normalizados”, cuja variação dos dados se tornaram bem mais homogênea e com isso os resíduos passaram necessariamente a ter uma distribuição

normal, como será mostrado mais a seguir. Este é um exemplo de uma transformação *Box-Cox*.

Após o ajuste do modelo, os dados voltam ao seu formato original, não linear, ou seja, é feito o caminho inverso ao logaritmo para que os dados, no seu formato original, possam ser usados para o cálculo da análise de ajuste e validação do modelo. Ou seja:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_kx_k$$

$$\text{Log } \hat{y} = \text{log}b_0 + b_1 \text{log}X_1 + b_2 \text{log}X_2 + b_3 \text{log}X_3 + \dots + b_k \text{log}X_k$$

$$\text{Log } Q = \text{log}b_0 + b_1 \text{log}P + b_2 \text{log}A + b_3 \text{log}DD + \dots + b_k \text{log}L$$

$$b \text{ log } P = \text{log } P^b \quad (\text{Propriedade de logaritmo})$$

$$\text{Log } Q = \text{log}b_0 + \text{log } P^{b_1} + \text{log}A^{b_2} + \text{log}DD^{b_3} + \dots + \text{log}L^{b_k}$$

$$X(\text{Log } Q) = X(\text{log}b_0 + \text{log } P^{b_1} + \text{log}A^{b_2} + \text{log}DD^{b_3} + \dots + \text{log}L^{b_k}) \quad (\text{inverso do log})$$

$$X^{\text{log } Q} = Q \quad \text{e} \quad X^{a+b} = X^a \cdot X^b$$

$$Q = X^{\text{log}b_0} \cdot P^{b_1} \cdot A^{b_2} \cdot DD^{b_3} \cdot \dots \cdot L^{b_k}$$

cte

3.4. AJUSTE DO MODELO

3.4.1. ANÁLISE DE CORRELAÇÃO

Num estudo de dispersão, onde se têm gráficos nos quais a cada ponto plotado representa um par observado de valores para as variáveis dependentes e independentes (Kazmier, 1982), podemos ter, como neste caso, a variável dependente sendo a vazão *Q* e a variável independente a precipitação *P* e a esses dados plotados serem designadas linhas que sejam melhores ajustantes e sua localização precisa. O procedimento para isso é o uso do método dos mínimos quadrados e a resposta disso é a idéia da correlação, fornecida através de um coeficiente denominado coeficiente de correlação linear – *r*, entre as duas variáveis e seu comportamento visual.

O coeficiente de correlação linear indica o grau de intensidade da correlação ou relação entre as variáveis *X* e *Y*. O sentido dessa correlação é indicado através de seu sinal

(+ ou -) e seu valor absoluto indica a extensão da relação, sendo este pertencente ao intervalo $[-1, +1]$, com forte correlação positiva ou negativa, e calculado através da seguinte fórmula:

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma_x} \right) \left(\frac{y_i - \bar{Y}}{\sigma_y} \right) \quad , \text{ sendo } \sigma_x \text{ o desvio padrão com relação à } X \text{ e } \sigma_y \text{ o}$$

desvio padrão com relação à Y , ou:

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \quad . \text{ Assim sendo, a fórmula do coeficiente}$$

de correlação é utilizada quando o propósito da análise é determinar a extensão e o tipo do relacionamento entre duas variáveis. Com ela o sinal do coeficiente é determinado automaticamente, sem necessidade de observar ou calcular a declividade da linha de regressão. Assim sendo podemos dizer que este coeficiente mede o quanto os dados aproximam-se de uma reta. Ver Figura 3.4 que ilustra melhor tal afirmação.

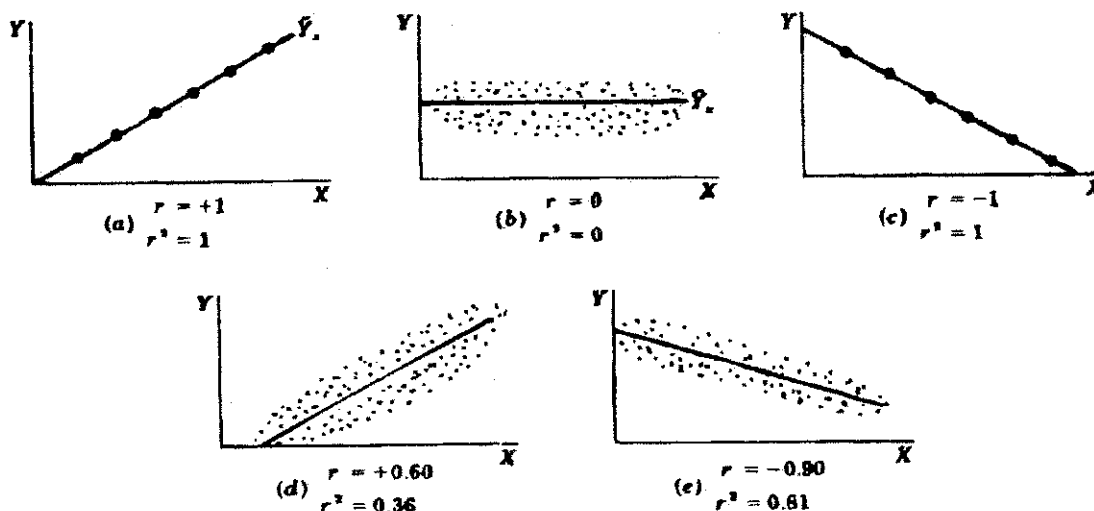


Figura 3.4 – Figura ilustrativa da aparência geral dos diagramas de dispersão associados com diversos valores de correlação (Kazmier, 1982).

O coeficiente de correlação é também chamado coeficiente de correlação de Pearson. Sendo imprescindível que uma relação descrita pelo coeficiente de correlação de Pearson se aproxime de uma função linear (Crespo, 1994), equação essa do tipo: $Y = aX + b$, onde a é a declividade da linha de regressão (inclinação da reta) e expressa por:

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

e b o ponto de interseção da linha com o eixo Y (onde $X = 0$) e expresso por:

$$b = \bar{Y} - b\bar{X} \quad \text{onde, } \bar{X} = \frac{\sum x_i}{n} \text{ e } \bar{Y} = \frac{\sum y_i}{n}, \text{ sendo } n \text{ o número de observações.}$$

De acordo com Crespo (1994), dependendo do tamanho da amostra, para podermos tirar algumas conclusões significativas sobre o comportamento simultâneo das variáveis analisadas, é necessário que:

$0,6 \leq |r| \leq 1$ para que haja uma boa correlação entre as variáveis.

Se $0,3 \leq |r| < 0,6$, há uma correlação relativamente fraca entre as variáveis.

Se $0 < |r| < 0,3$, a correlação é muito fraca e, praticamente, nada podemos concluir sobre a relação entre as variáveis em estudo.

Como consequência, pode-se analisar também o coeficiente de determinação – R^2 que mede a proporção da variação de Y em relação à média que é explicada pela regressão. Um resultado a reter é que $R^2 = \sigma_{\hat{y}y}^2$ (quadrado do coeficiente de correlação entre Y valor observado e \hat{y} valor calculado). Em princípio, a qualidade do ajuste será tanto maior quanto mais R^2 se aproximar da unidade e sendo expresso por:

$$R^2 = \frac{b \sum y + a \sum xy - n\bar{y}^2}{\sum y^2 - n\bar{y}^2} \quad (\text{para uma correlação simples) ou:}$$

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum e_i^2}{\sum y_i^2} = \frac{\hat{b}_1 \sum yx_1 + \hat{b}_2 \sum yx_2}{\sum y^2} \quad (\text{para uma correlação com três variáveis}).$$

Muito embora o R^2 seja relativamente fácil de interpretar, segundo Kazmier (1982), ele não pode ser testado estatisticamente, contudo, a raiz quadrada deste $r = \sqrt{R^2}$ (para dados amostrais), pode ser testado estatisticamente, pois está incluída em uma estatística de teste que é distribuída segundo uma distribuição t , quando a relação populacional ρ , é igual a 0. Além disso, o sinal aritmético associado com o coeficiente de correlação, que é sempre

o mesmo sinal associado com β na equação de regressão, indica a direção da relação entre x e y (positivo = direta; negativo = inversa). Por tais razões, o coeficiente de correlação em lugar do coeficiente de determinação para dados é mais freqüentemente apresentado como medida da relação.

Para o coeficiente de determinação múltipla quando se tem duas ou mais variáveis independentes e considerando o grau de liberdade, quando se adiciona variáveis independentes ou explicativas, pode-se ter como referência do coeficiente de determinação ajustado (R^2_{ajust} ou \bar{R}^2), onde este pode ser calculado como:

$$\bar{R}^2 = 1 - \left(1 - R^2\right) \frac{n-1}{n-k}, \text{ sendo } n \text{ o número de observações ou quantidade de}$$

dados observados; e k o número de variáveis estimadas.

Então, com o coeficiente de determinação sabe-se a proporção da variância na variável dependente que é estatisticamente explicada pela equação de regressão.

Logo, pelas razões citadas acima, o coeficiente de determinação é adotado neste estudo como indicativo da correlação, haja vista que este nos leva intuitivamente ao coeficiente de correlação da relação.

3.4.2 ANÁLISE DE REGRESSÃO

Considerando-se que o melhor modelo de regressão seria aquele que estimasse a variável dependente com pequeno erro e com base no menor número possível de variáveis independentes. E sabendo-se ainda, que este tem como finalidade reduzir o trabalho no processo de predição das vazões, foram utilizadas neste estudo, as variáveis independentes de mais fácil obtenção: área, precipitação, densidade de drenagem, declividade do rio principal e comprimento do rio principal.

E então, com o objetivo preditivo, o de obter uma relação que nos permita, perante futuras observações das variáveis X_k , prever o correspondente valor de Y , sem necessidade de o medir, optou-se por fazer a Análise de Regressão Linear.

Segundo Matos (1995), a Regressão nasce da tentativa de relacionar um conjunto de observações de certas variáveis, designadas genericamente por X_k ($k = 1 \dots p$), com as leituras de uma certa grandeza Y . No caso da regressão linear, está subjacente uma relação

do tipo: $Y = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + e_p$, onde a, b_1, b_2, \dots, b_p seriam os parâmetros da relação linear procurada. Independentemente dos objetivos, as variáveis X_k são muitas vezes designadas por variáveis explicativas, uma vez que tentam explicar as razões da variação de Y . Esta consiste em estimar os valores dos parâmetros a, b_1, b_2, \dots, b_p , através da minimização da soma dos quadrados dos desvios que às vezes se utiliza, nomeadamente Regressão Simples quando $p = 1$ ou Regressão Múltipla que é usado para explicitar o caso $p > 1$.

A análise de regressão linear tem por objetivo descrever, através de um modelo matemático, a relação entre duas variáveis, partindo de n observações das mesmas (Crespo, 1994). Ou seja, tal análise prediz o valor de uma variável (a variável dependente), dado que seja conhecido o valor de uma variável associada (a variável independente). Na qual, a expressão análise de regressão simples indica que a predição da variável dependente é feita em uma variável independente, enquanto a análise de regressão múltipla diz respeito à predição da variável dependente com base em duas ou mais variáveis independentes (Kazmier, 1982). O procedimento de análise consistiu na busca do melhor resultado em termos de melhoria do coeficiente de determinação (R^2), na diminuição dos resíduos e na análise dos demais coeficientes apresentados em cada resultado.

Portanto, dado um problema onde temos duas ou mais variáveis, procura-se verificar se existe alguma relação entre elas e qual o grau dessa relação que é a extensão da dispersão dos pontos com relação à linha de regressão. E sendo essa relação entre as variáveis de natureza quantitativa, a correlação é o instrumento adequado para descobrir e medir essa relação que, entretanto, é descrita através de uma função matemática. Logo, temos a regressão como instrumento adequado para a determinação dos parâmetros dessa equação (Crespo, 1994).

Neste estudo, usou-se a regressão linear múltipla, uma extensão da análise linear simples, que envolve duas ou mais variáveis independentes (X_1, \dots, X_k) para se prever o comportamento da variável dependente Y . Um processo mais complicado, mas que hoje, com o avanço dos computadores, se tornou uma ferramenta mais usual. Pois não é mais necessário se resolver tal problema na mão, basta entender o programa, desenvolver o algoritmo sobre o qual seus dados serão analisados e saber interpretar os resultados fornecidos por eles (Dumm e Clarck, 1974).

Entretanto, os cálculos necessários para determinar os valores das constantes numa equação de regressão múltipla, bem como para determinar os valores dos erros padrões associados, são cálculos bastantes complexos e que geralmente envolvem álgebras de matrizes (Kazmier, 1982). Contudo, para realizar tais cálculos, encontra-se à disposição vários programas de computadores como é o caso do SOC, citado anteriormente.

No caso de se trabalhar com duas variáveis independentes, denotadas por X_1 e X_2 , a equação de regressão múltipla é $y_c = a + b_1X_1 + b_2X_2$, onde y_c é o “valor calculado de Y ”, ou “ Y calculado” e essa equação, assim como no caso da regressão simples, identifica a linha ajustante baseado no método dos mínimos quadrados (Figura 3.5), onde este tem como objetivo tornar mínima a soma dos quadrados dos desvios. E para esse caso, a melhor linha ajustante é uma linha através de um espaço n -dimensional (no caso de duas variáveis independentes, o espaço é tridimensional).

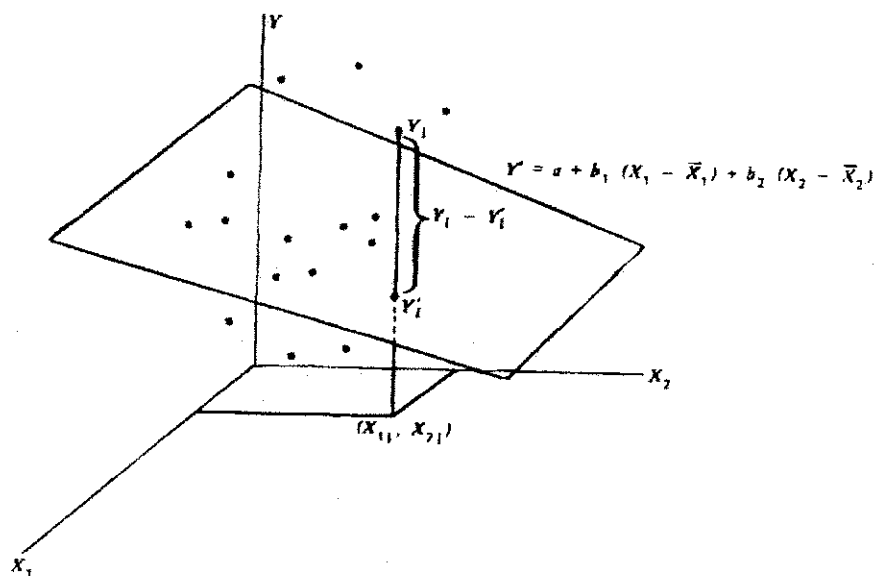


Figura 3.5 – Visualização do método dos mínimos quadrados plano em 3 dimensões numa regressão múltipla (Dunn & Clark, 1974)

3.4.3. ANÁLISE DO MODELO LINEAR COM TRÊS VARIÁVEIS

Como foi visto anteriormente, a análise de regressão múltipla é usada para testar hipóteses sobre a relação de uma variável dependente, y , e duas ou mais variáveis independentes, X , e para previsões.

O modelo linear de três variáveis pode ser escrito como foi desenvolvido por Salvatore (1982) e mostrado a seguir:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + u_i$$

Havendo relação linear exata entre os X , a estimativa dos parâmetros de mínimos quadrados simples (MMQ) para essa equação pode ser obtida pela minimização da soma dos quadrados dos resíduos, ou seja:

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1x_{1i} - b_2x_{2i})^2, \text{ resultando na estimativa dos}$$

coeficientes parciais da regressão, quando expressados na forma de desvios, como sendo:

$$b_1 = \frac{(\sum x_1y)(\sum x_2^2) - (\sum x_2y)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

$$b_2 = \frac{(\sum x_2y)(\sum x_1^2) - (\sum x_1y)(\sum x_1x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1x_2)^2}$$

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

Logo, para se testar a significância estatística desses parâmetros estimados, precisa-se das variâncias das estimativas:

$$\text{var } b_1 = \sigma_u^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2}$$

$$\text{var } b_2 = \sigma_u^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2}$$

b_0 normalmente apresenta pouco interesse.

Sendo σ_u^2 conhecido, a variância residual, S^2 , é usada como uma estatística não viesada de σ_u^2 e dada por:

$$S^2 = \hat{\sigma}_u^2 = \frac{\sum e_i^2}{n-k} \text{ onde } k \text{ é o número de parâmetros estimados.}$$

Então as estimativas não viesadas de b_1 e b_2 são dadas por:

$$S_{b_1}^2 = \frac{\sum e_i^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 \sum x_2)^2}$$

$$S_{b_2}^2 = \frac{\sum e_i^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 \sum x_2)^2}, \text{ onde } S_{b_1}^2 \text{ e } S_{b_2}^2 \text{ são os erros padrões}$$

das estimativas.

Portanto b_1 e b_2 são também normalmente distribuído, podendo ser usada a distribuição t com n-k graus de liberdade para testar hipóteses e intervalos de confiança para b_0 , b_1 e b_2 . Enquanto que a significância global da regressão pode ser testada pela razão da variância explicada com a variância não explicada. Logo esta apresenta uma distribuição F com k-1 e n-k graus de liberdade, onde:

$$F_{k-1, n-k} = \frac{\sum \hat{y}_i^2 / (k-1)}{\sum e_i^2 / (n-k)} = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

Se a razão $F_{k-1, n-k}$ calculada excede o valor de F (tabelado) a um dado nível de significância e graus de liberdade, aceita-se a hipótese de que os parâmetros da regressão não são todos iguais a zero e que R^2 é significativamente diferente de zero.

3.4.4 MONTAGEM DA EQUAÇÃO

Na seleção da melhor equação são usadas álgebras de matrizes, método dos mínimos quadrados além dos testes de independência serial (coeficiente de autocorrelação), os testes de estacionariedade (t – Student para as médias e F – Fisher para as variâncias) das séries para verificação da adequabilidade do tratamento probabilístico dessas séries e o melhor conjunto de parâmetros associados na análise de regressão múltipla.

O teste-F (Fisher) é um teste de hipótese utilizado para comparar as variâncias de duas populações a um nível de significância α , relacionado à probabilidade de ocorrência da rejeição de uma hipótese verdadeira, ou ainda, para verificar qual das duas populações com distribuição normal têm mais variabilidade. Segundo Dunn e Clark (1974), o teste-F é um critério, usualmente, para adicionar ou remover termos no modelo, ou seja, quando se tem apenas uma variável na equação e acrescentamos outra, logo pode haver uma resposta diferente do modelo se esta variável estiver muito correlacionada com a outra variável que entrou no modelo. O teste-F mostra o grau dessa correlação, dizendo se a variável deve permanecer ou sair do modelo com o critério de rejeição/aceitação (ou nível descritivo) da ordem de:

$$(PR > F) \leq 0,01 - \text{significância a 1\% ;}$$

$$0,01 < (PR > F) \leq 0,05 - \text{significância a 5\% e;}$$

$$(PR > F) > 0,05 - \text{não significância. (P > 0,05).}$$

A equação desenvolvida neste estudo terá uma variável dependente, a vazão média diária, e as variáveis independentes que são a precipitação sazonal e as características físicas das bacias dos postos fluviométricos selecionados, arrumada na forma da função COBB-DOUGLAS, vista anteriormente no item 3.3.3 *Transformação das variáveis*, e cujos parâmetros foram fornecidos pelas tabelas de saída do programa SOC, na análise de regressão múltipla.

Muito comuns em diversos tipos de estudos estatísticos como este, são as tabelas de análise de variância, sendo freqüentemente incorporadas nos programas dedicados à regressão e nas folhas de cálculo que incluem este tipo de estudo. Estas são as tabelas de saída do programa SOC. A organização dos valores tem normalmente o aspecto como o mostrado nas Tabelas 3.3 para regressão linear simples e Tabela 3.4 para a regressão linear múltipla.

Tabela 3.3 - Quadro típico do SOC para uma regressão linear simples.

<i>Fonte de Variação</i>	<i>Gl</i>	<i>Soma de quadrados</i>	<i>Quadrado Médio</i>	<i>Valor F</i>	<i>R²</i>	<i>PR>F (Teste F)</i>
Regressão	P	SSR	MSR	MSR/MSE	SSR/(SSR+SSE)	*
Resíduo	n-p-1	SSE	MSE			
Total	n-1	TSS				

Tabela 3.4 - Quadro típico do SOC para análise de variância em uma regressão linear múltipla.

Fonte de Var.	Gl	Soma de quadrados	MS	EMS	F calc.	F tab.
Regres. em x_1, \dots, x_k	K	$\sum_{i=1}^n (y_i - \bar{y})^2 = b_1 \sum x_{1i} y_i + \dots + b_k \sum x_{ki} y_i$	$MS_{1..k}$	$\sigma_e^2 + \sum_{i=1}^k \sum_{j=1}^k (\beta_i \beta_j \sum x_i x_j)$	$MS_{1..k} / S_e^2$	F[1-k; n-k-1]
Residuo	n-k-1	$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum y_i^2 - b_1 \sum x_{1i} y_i - \dots - b_k \sum x_{ki} y_i$	S_e^2			
Total	n-1	$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum y_i^2$				

Tais iniciais referidas nas tabelas acima são abreviações da língua inglesa. Ou seja:

SSR = "Sum of Squares due to the Regression" (soma quadrática devida à regressão);

SSE = "Sum of Squares due to the Error" (soma quadrática devida ao erro);

SSM = "Sum of Squares due to the Mean" (soma quadrática devida à média);

TSS = "Total Sum of Squares" (soma quadrática total);

TSS (adj) = "Adjusted Total Sum of Squares" (soma quadrática total ajustada);

MSR = "Mean of Squares due to the Regression" (média quadrática devida à regressão);

MSE = "Mean of Squares due to the Error" (média quadrática devida ao erro).

E ainda, as médias dos quadrados são obtidas dividindo-se as somas dos quadrados pelos graus de liberdade correspondentes, como:

$$MSE = SSE / (n-p-1). \text{ Repare-se que } MSE = \hat{\sigma}^2$$

$$TSS = SSM + SSR + SSE, \text{ ou seja,}$$

$$\sum_i y_i^2 = n \bar{y}^2 + \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i r_i^2 \quad \text{Ou ainda,}$$

$$TSS (\text{adj}) = SSR + SSE$$

Gl = Grau de liberdade

Os valores de F apresentados na tabela permitem realizar testes de nulidade dos parâmetros, facilitando o cálculo do valor da probabilidade do teste-F, permitindo uma avaliação imediata da rejeição ou não da hipótese $H_0 = 0$. A rejeição dá-se quando o valor da probabilidade é pequeno, correspondendo a valores elevados de F. Os níveis descritivos associados a estes testes de nulidade para H_0 são discriminados nas tabelas de correlação na análise de regressão.

Para isso é feita uma análise de modelos de regressão linear univariada através do método de mínimos quadrados, selecionando o melhor modelo segundo o método STEPWISE. A construção o melhor modelo é realizada de acordo com o seguinte processo (SOC, 1995):

1. O primeiro modelo é construído com a variável independente mais correlacionada com a variável dependente. Uma estatística F, chamada "F para sair" é gerada para verificar se a variável selecionada deve permanecer no modelo. A variável permanece no modelo se o nível de significância da estatística calculada, for menor do que um valor pré-especificado. Caso contrário nenhuma variável é selecionada e o processo é interrompido.

2. O processo de construção continua com o cálculo de uma estatística F, chamada "F para entrar", para cada uma das variáveis independentes não incluídas no modelo. A próxima variável a ser incluída no modelo é aquela com o menor nível de significância associado à estatística "F para entrar", desde que este valor seja menor que um valor pré-estabelecido.

3. Selecionada uma nova variável, um novo modelo é construído com todas as variáveis selecionadas e a estatística "F para sair" é calculada para cada uma das variáveis. A variável com o maior nível de significância associado à estatística "F para sair" é retirada do modelo, desde que este valor seja maior que um valor pré-estabelecido. Um novo modelo é construído com as variáveis que permaneceram e o processo de verificação para a retirada de variáveis é repetido.

Os itens 2 e 3 são repetidos até que não haja mais nenhuma variável para entrar ou sair do modelo, ou seja, são realizadas todas as iterações possíveis. Ou estes são repetidos até que o número de iterações pré-especificado seja atingido, já que há a possibilidade de se fixar um número de vezes em que o processo se repetirá. Para cada iteração todo o resultado estatístico é mostrado para que se possa escolher a melhor combinação de resultados.

O teste a realizar usando a estatística F, com γ_0 típico da ordem de 0,05 é calculado como:

$$H_0: b_2 = 0 \quad H_a: b_2 \neq 0$$

$$F = \frac{SSR - SSR_1}{rMSE} \quad \text{ou} \quad F = \frac{MSR}{MSE}$$

Rejeição de H_0 se $\gamma < \gamma_0$, sendo $F_{1-\gamma_0}(r, n-p-1) \leq F$

Ou seja, valores elevados de F conduzem à rejeição da hipótese $b_2=0$, e as variáveis em causa são mantidas no modelo. Quanto maior é o valor calculado de F, mais fácil é aquela rejeição, por ser menor o valor de γ_0 para a qual o valor da tabela é menor ou igual ao valor calculado de F. O valor limite γ_0 a utilizar depende das circunstâncias (0,05 ou 0,10), mas pode ir até 0,25, numa opção cautelosa (no sentido de manter o modelo) que torna mais difícil não rejeitar a hipótese nula.

Uma outra alternativa ao uso de F é o cálculo de C_k , onde:

$$C_k = \frac{SSE_1}{MSE} - (n - 2k), \quad \text{Sendo } k \text{ o número de variáveis do modelo reduzido.}$$

Então se o valor de c_k for muito superior a k , deve suspeitar-se que algumas variáveis importantes serão rejeitadas, caso se opte pelo modelo reduzido. Devem procurar-se, portanto, subconjuntos de variáveis que conduzam a valores de c_k próximos de k . Menos formalmente, podem comparar-se os valores de MSE e R^2 para os modelos "completo" e reduzido. Se não diferirem muito, será mais econômico usar o modelo reduzido.

3.4.5 ANÁLISE DO INTERVALO DE CONFIANÇA DAS ESTIMATIVAS

Uma vez que se quer futuras observações de Q baseados na equação proposta e em valores medidos, há uma necessidade de se avaliar qual a confiança deste resultado, se Q é realmente a vazão representativa desejada e qual o nível de confiança correspondente.

Então, o cálculo do intervalo de confiança permite ter uma idéia indireta da qualidade da regressão, ou seja, no cálculo do intervalo de confiança é obtido para valores fixados de X, estimativas \hat{y} para os correspondentes valores de Y e essas estimativas têm um erro que convém ser estimado através do intervalo de confiança. Contudo, é evidente

que ao tomar os valores fixados de X , a nova estimativa será diferente da do verdadeiro valor observado representado pela reta na Figura 3.6 mostrada abaixo.

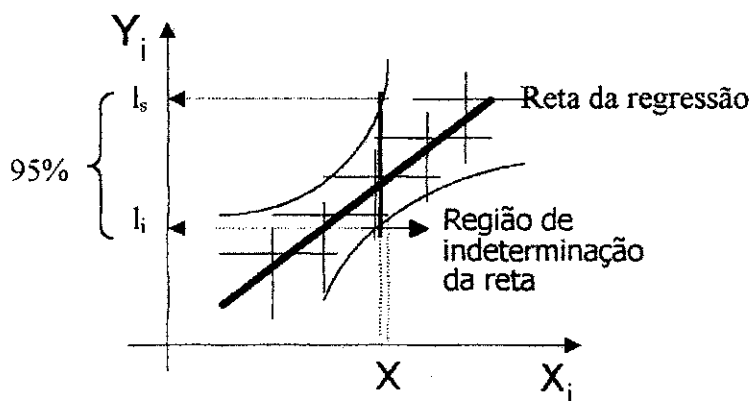


Figura 3.6 – Figura ilustrativa do intervalo de confiança

As versões habituais baseiam-se na distribuição normal e, em alguns casos, do χ^2 , aqui substituídas respectivamente pelas distribuições t Student e F, dado que a variância σ é sempre estimada.

Portanto, tem-se um alfa que é o nível de significância utilizado para calcular o nível de confiança. O nível de confiança é igual a $100(1 - \alpha)\%$ ou, em outras palavras, um alfa de 0,05 indica um nível de confiança de 95%. Se for considerado que α é igual a 0,05, precisa-se calcular a área sob a curva normal padrão que é igual a $(1 - \alpha)$ ou 95%. Com isso podemos mostrar que há 95% de chance do resultado está dentro do intervalo calculado.

3.5. VALIDAÇÃO DO MODELO

Para uma melhor análise dos resultados obtidos foi realizada a validação cruzada dos dados, sendo analisadas as 13 bacias hidrográficas dos postos fluviométricos que pertencem às bacias dos Rios Piranhas-Açu-Apodí, em escala sazonal.

O uso da validação cruzada na construção do modelo serve tanto para a seleção da melhor equação encontrada, quanto para a avaliação da capacidade preditiva do modelo proposto.

3.5.1 MÉTODO DE VALIDAÇÃO CRUZADA

No caso de se trabalhar com uma boa quantidade de dados, pode-se fazer uso da técnica de validação cruzada, uma estratégia de estimativa de erros para avaliar o modelo e saber se este explica bem os dados ou se ele permite predizer bem.

O método é bastante simples e desenvolvido da seguinte forma: para um conjunto de N observações é necessário que se divida tal amostra em k subconjuntos diferentes com tamanhos (aproximadamente) iguais e em seguida, que se ajuste o modelo sempre omitindo um dos subconjuntos e que se use só esse subconjunto omitido no modelo ajustado. Neste trabalho se usou como subconjunto as séries das bacias dos postos fluviométricos.

Este método difere do método de dividir a amostra, pois neste só um único subconjunto fixo é usado para a validação, enquanto que, na validação cruzada são usados k conjuntos diferentes que são testados um a um.

Neste trabalho foram analisadas 13 bacias: Antenor Navarro (NA), Aparecida (AP), Augusto Severo (AS), Emas (EM), Fazenda Angicos (FA), Mossoró (MO), Pau dos Ferros (PF), Pedra de Abelhas (PA), Piancó (PI), São Fernando (SF), Sítio Volta (SV), Upanema (UP) e Serra Negra (SN). Foi feita a calibração do modelo primeiramente excluindo (AN), ou seja, usou-se AP, AS, EM, FA, MO, PF, PA, PI, SF, SV, UP e SN e ajustou-se o modelo a estas, conseqüentemente foi encontrada uma equação, em seguida usou-se os dados da bacia excluída (AN) nesta equação ajustada, para verificar se os resultados encontrados correspondiam bem aos observados. Da mesma forma, se excluiu a próxima bacia (AP) e usou-se o mesmo procedimento anteriormente citado e obtendo-se um novo conjunto de resposta. Desta forma foram feitos os testes com todas as bacias uma a uma.

E então para todos os casos foram analisados os resultados dos coeficientes de correlação e sua forma gráfica.

3.6. ANÁLISE DOS RESÍDUOS

Segundo Barroso et al. (1987), uma maneira de verificar a adequação do modelo é comparar cada valor observado y_i com o respectivo valor predito pelo modelo \hat{y}_i . A diferença entre estes dois valores é o resíduo: $r_i = y_i - \hat{y}_i$.

Logo, os resíduos são a parte do valor observado que o ajuste não foi capaz de explicar. Ou seja, para um dado valor de X_0 , pode-se encontrar um valor \hat{Y} , com um erro, veja Figura 3.7 ilustrativa a seguir:

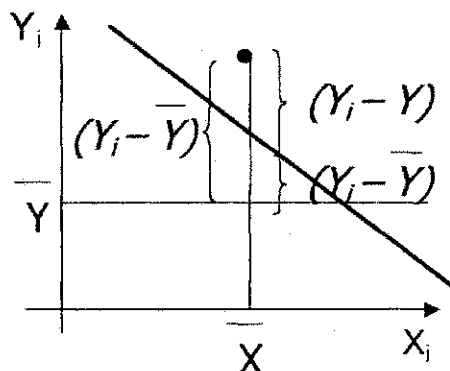


Figura 3.7 - Figura ilustrativa do resíduo

Na regressão múltipla usada, está implícito que o resíduo é multiplicativo, já que este deve voltar na forma transformada, linear, como um termo aditivo, tendo uma distribuição log-normal.

Segundo Tucci e Clarke (2001), para a verificação da análise de regressão múltipla, deve-se fazer a análise dos resíduos da regressão e esta deve satisfazer três condições:

1) Se for necessário fazer testes de significância sobre os coeficientes de regressão b_0, b_1, \dots, b_k ou se for necessário calcular intervalos de confiança para valores estimados da variável \hat{Y} , dado os valores de X_1, X_2, \dots, X_k é necessário verificar que os resíduos têm distribuição aproximadamente normal.

2) A variância dos resíduos deve ser constante, não variando com o tamanho das variáveis independentes;

3) Os resíduos não mostram qualquer evidência de relação curvilínea, ao plotá-los com as variáveis exploratórias (a presença de uma relação curvilínea sugere a necessidade de incluir outras variáveis exploratórias na regressão).

Essas análises foram também observadas por Draper e Smith (1981). Tais análises são definidas neste estudo, seguindo a denominação: (1) Distribuição de frequência; (2) Distribuição cronológica e (3) Distribuição com o valor ajustado.

A distribuição de frequência é a verificação visual da normalidade e segundo Matos (1995) ela também pode ser feita ordenando os resíduos de forma crescente desenhando-os

em papel de distribuição normal, então, se a presunção de normalidade se verificar, os resíduos deverão estar aproximadamente em linha reta.

A distribuição cronológica mostra se os resíduos são estacionários, ou seja, se elas são aleatórios. Caso não ocorra aleatoriedade deve-se fazer mudanças no modelo.

As distribuições com o valor calculado mostram as desigualdades de variância dos erros. Estas podem ser detectadas a partir de um gráfico dos resíduos r_i em função dos \hat{y}_i . Se o aspecto não for uma mancha de largura uniforme, por exemplo, alargando com o aumento de \hat{y}_i , poderá ser necessário transformar y , por exemplo, fazendo $\ln y$, $1/y$, etc ou alterar o modelo (Matos, 1995).

Essas análises foram feitas neste estudo e poderão ser vistas no capítulo 4 - *Resultados*.

Um outro gráfico, mas dos quadrados dos resíduos, pode confirmar suspeitas e ajudar a detectar erros isolados.

A utilização do modelo sugerido ou qualquer outro que seja, incluirá sempre uma parcela de resíduo, devendo-se, pois, fazer uma avaliação da qualidade do ajuste. Onde, utilizando o índice e_i ($i = 1 \dots n$) para indicar cada conjunto, temos então:

$$y_i = a + b_1 x_{i1} + b_2 x_{i2} + \dots + b_n x_{in} + e_i$$

Um pressuposto importante para todo o desenvolvimento é de que os resíduos do modelo têm média nula, não estão correlacionados e têm variância constante. Outra condição adicional para os resíduos do modelo é de que estejam normalmente distribuídos. Logo, apesar de não ser essencial para a derivação das expressões de cálculo das estimativas dos parâmetros, estes pressupostos são indispensáveis para toda a matéria referentes a testes de hipóteses e derivação de intervalos de confiança e, em geral, para toda a validação estatística dos resultados.

Para contemplar esse importante aspecto, há alguns procedimentos de verificação desta parcela, a análise de resíduos ou erro quadrático, onde o valor minimizado do quadrado dos erros pode ser calculado através de: $\sum e^2 = \sum (y - \hat{y})^2$.

Além dos resíduos correspondentes diretamente aos erros do modelo, $r = y - \hat{y}$, é usual calcular também os resíduos percentuais, ou seja, $r = \left(\frac{Q_o - Q_c}{Q_c} \right) * 100$. Este tipo de análise de resíduo é usado para o cálculo da distribuição cronológica neste estudo e

também para a análise visual nos gráficos, resultados da validação. No caso do cálculo da distribuição de frequência e da distribuição com o valor calculado usou-se o resíduo como sendo: $r = Q_o - Q_c$, sendo Q_o a vazão observada e Q_c a vazão calculada.

3.7. CASO DE ESTUDO

3.7.1. DADOS DESCRITIVOS DA REGIÃO

A região em estudo, ver Figura 3.8, é formada pelas bacias hidrográficas dos rios Apodi (rio estadual) e Piranhas-Açu (rio federal); cujos principais municípios desta bacia são Mossoró, Açu e Caicó, localizados no Estado do Rio Grande do Norte, e Patos, Sousa e Cajazeiras, no Estado da Paraíba.

A UFPB/ANEEL (2001) assim caracteriza a região estudada:

“É uma região com características bastante peculiares, com rios intermitentes (vazões nulas em grande parte do ano); possui altas taxas de evaporação; uma baixa densidade de postos fluviométricos; além de possuir impacto antrópico, pelo desmatamento da vegetação nativa e, principalmente, pela alta densidade de reservatórios superficiais, a maioria sem informações cadastrais e operacionais.

“A região estudada está inserida parte no Estado do Rio Grande do Norte e parte no Estado da Paraíba entre os meridianos $36^{\circ} 14' 15''$ e $38^{\circ} 46' 22''$ de longitude oeste e entre os paralelos $4^{\circ} 48' 04''$ e $36^{\circ} 14' 15''$ de latitude sul, situada na região semi-árida. Esta região é denominada pela ANEEL - Agência Nacional de Energia Elétrica como sub-bacia 37 e esta sub-bacia está limitada ao oeste pela sub-bacia 36, ao leste pela sub-bacia 38, ao norte pelo oceano Atlântico e ao sul pela sub-bacia 48. E os principais afluentes do rio Piranhas-Açu são os rios Espinharas, do Peixe, Piancó e Seridó e quanto ao Apodi é o rio do Carmo.

“Toda a cobertura vegetal encontrada pertence ao que se convencionou denominar de Caatinga, uma vegetação do tipo savana estépica, ‘estacional-decidual, portanto com os estratos arbóreo e gramíneo-lenhoso periódicos e com numerosas plantas suculentas, sobretudo cactáceas’ (IBGE, 1995).

“O clima apresenta, de modo geral, dois períodos definidos sendo um mais longo, seco, intercalado por um período pluvial curto. A estação seca, de um modo geral, se estende por oito meses do ano, ou às vezes mais. A variabilidade espacial da precipitação, quando considerada em maior escala, define apenas uma região com algumas variações, em geral definidas por acidentes topográficos. Quando a variabilidade é considerada em menor escala, mostra diferenças (sem padrão definido) entre a ocorrência de precipitação em áreas relativamente próximas espacialmente. Esta variabilidade em pequena escala é determinada pela natureza convectiva das precipitações. As chuvas se iniciam eventualmente a partir de outubro, na chamada ‘pré-estação’, culminando com uma estação chuvosa concentrada entre fevereiro e maio.

“A área estudada possui, devido à baixa latitude, uma temperatura elevada em quase toda a sua extensão, sendo a temperatura média do mês mais frio sempre superior a 18 °C. O que faz variar o clima da área é a irregularidade do relevo e também a distribuição desigual das chuvas, tanto de um local para outro como de um ano para outro, sendo também irregular sua distribuição durante a estação chuvosa. A evaporação é alta em toda a bacia com taxas anuais entre 2000 mm e 3000 mm, com baixa variabilidade espacial e interanual.

“De uma maneira geral, os sistemas aquíferos são hidrologicamente modestos, mesmo quando conseguem perenizar os cursos d’água regionais. Isto se deve, claro, não somente às características hidrogeológicas, geralmente sofríveis, dos sistemas aquíferos, mas esta reduzida contribuição hídrica está relacionada com as condições climáticas da região semi-árida onde se insere, com ênfase à ação da evapotranspiração real e com a cobertura vegetal, ao lado do solo, de baixo poder regulatório hidrológico onde ocorrem os seguintes sistemas aquíferos: Cristalino, Rio do Peixe, Apodi e Aluvial.

“Devido às características do regime pluviométrico e à formação geológica, com solos pouco profundos e não adequados ao armazenamento subterrâneo da água, os rios apresentam trechos intermitentes, tornando-se perenes apenas quando regularizados por reservatórios.

“O regime dos cursos d'água segue o do regime pluviométrico da região semi-árida, iniciando-se no mês de fevereiro e estendendo-se até o mês de maio ou junho, com picos entre março e maio, na maioria dos postos em abril.

“Uma importante característica da hidrologia regional da região é a altíssima densidade de reservatórios superficiais, com capacidade de armazenamento que varia de poucos milhares de m^3 até mais de dois bilhões de m^3 . Uma particularidade que deve ser observada é que a quase totalidade desses reservatórios não está inventariada nem possui informações cadastrais (época de construção, capacidade, curva cota-volume, etc.) ou séries temporais de registros de afluência, armazenamento, retirada de água para consumo e liberação para jusante. Sua presença tem modificado bastante o regime hidrológico dos rios, e a falta de informações dificulta a compreensão deste regime, comprometendo a representatividade das séries fluviométricas.”

3.7.2. DADOS HIDROLÓGICOS (PLUVIOMÉTRICOS E FLUVIOMÉTRICOS)

Dos 28 postos fluviométricos para a bacia Piranhas-Açu-Apodi, do banco de dados nacional da ANEEL, apenas 13 postos fluviométricos puderam ser empregados neste trabalho. Isso devido ao fato de no estudo anterior (UFPB/ANEEL, 2001), os demais postos apresentarem comportamentos distintos e possuírem açudes construídos a montante dos postos os quais interferiram no escoamento do rio na seção onde o posto está instalado, ou simplesmente não foram aprovados pelo teste não paramétrico, da estacionariedade das séries, que constatou a não homogeneidade da mesma. Com os 13 postos escolhidos haveria a possibilidade ainda de se comparar os resultados fornecidos pelo modelo com os valores observados nestes postos. As séries apresentaram muitas falhas, muitas repetições de registros e interrupções, tornando-as mais curtas que a duração desejada, mesmo assim não foram feitos o preenchimento dessas falhas, nem a análise de consistência para estas séries em tal estudo.

Foram escolhidos então, as séries que não apresentaram falhas agravantes e foi decidido adotar para este estudo, os mesmos postos e as séries sazonais do estudo anterior consideradas estacionárias, mas, não foi possível obter dados de precipitação correspondentes a toda a série de vazão para as bacias de Emas (nos anos de 1985 a 1999)

e de Pau dos Ferros (nos anos de 1984 a 1992), onde se resolveu retirar esses anos das séries destes postos.

A Figura 3.9 apresenta as bacias e a distribuição dos postos selecionados cujas características estão apresentadas na Tabela 3.5

A Tabela 3.6 apresenta os dados originais de precipitação (chuva) e vazão, já no nível diário sazonal (apenas para a estação chuvosa), para cada bacia estudada, mostrando a disposição das séries destas.

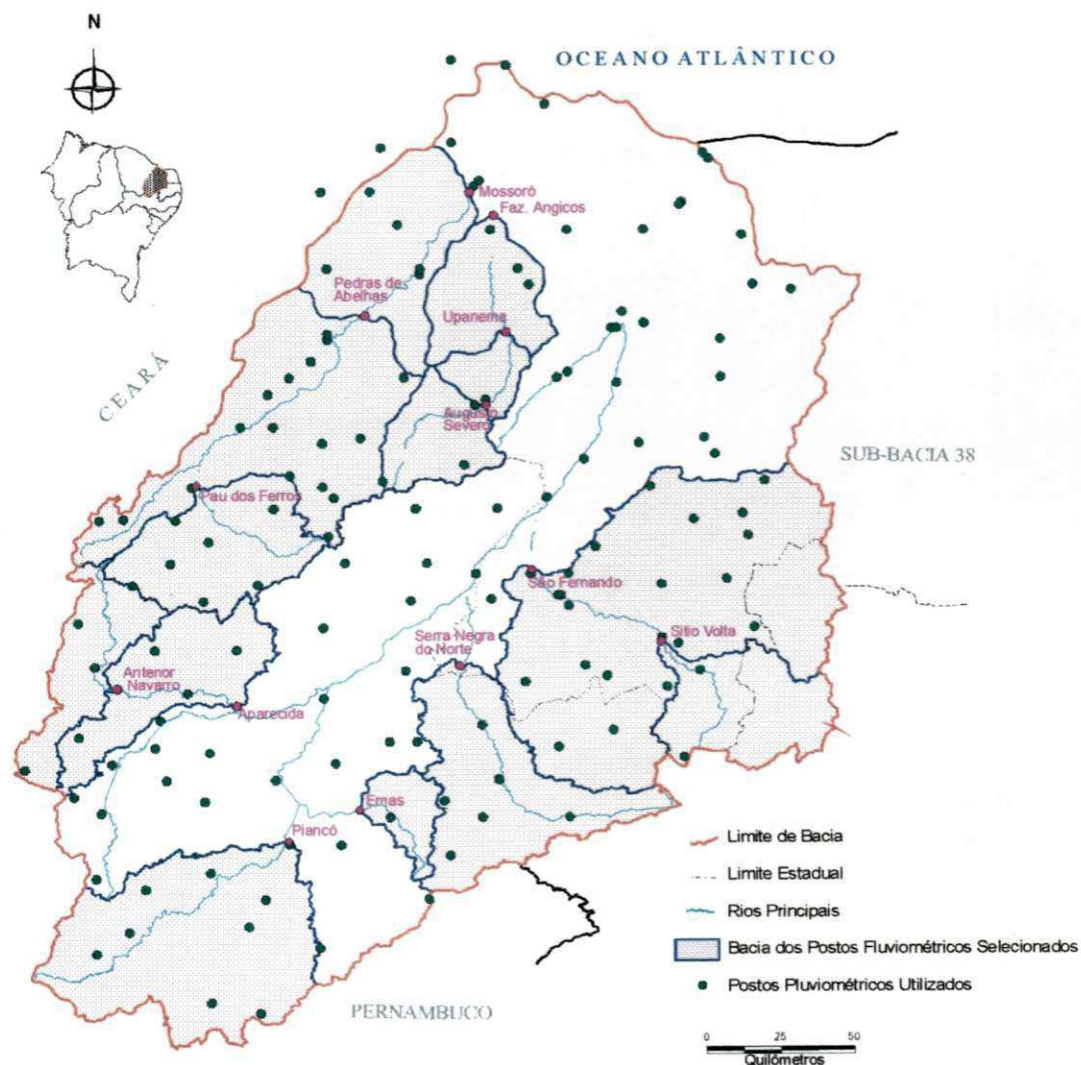


Figura 3.9 – Distribuição dos postos pluviométricos e fluviométricos selecionados (UFPB/ANEEL, 2001).

Tabela 3.5 – Descrição das bacias dos postos fluviométricos.

<i>Código</i>	<i>Posto Fluviométrico</i>	<i>Área de Drenagem (km²)</i>	<i>Período de dados disponíveis</i>	<i>Número total de anos</i>	<i>Número de postos pluviométricos na bacia</i>
37030000	Pau dos Ferros	2.050	1964-83	20	10
37080000	Pedra de Abelhas	6.481	1964-99	36	25
37090000	Mossoró	9.447	1987-88 1990-99	12	32
37150000	Augusto Severo	1.083	1987-99	13	5
37185000	Upanema	1.414	1975-90 1993-99	23	6
37190000	Fazenda Angicos	2.547	1993-99	7	11
37260000	Antenor Navarro	1.514	1965-72 1985-99	22	7
37290000	Aparecida	3.429	1985-99	15	14
37340000	Piancó	4.550	1964-68 1970-89 1991-97 1999	33	11
37360000	Emas	395	1964-65 1967-72	8	5
37440000	Serra Negra	3.027	1922-60	39	9
37515000	Sítio Volta	1.794	1979-87	9	4
37570000	São Fernando	9.700	1963-87	25	23

Tabela 3.6 - Séries sazonais de vazão e precipitação utilizadas.

Ano	Antenor Navarro		Aparecida		Augusto Severo		Emas		Fazenda Angicos		Mossoró		Pau dos Ferros	
	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão
1961														
1962														
1963														
1964							937,80	8,68					923,63	27,50
1965	642,87	6,92					816,62	9,49					705,60	26,95
1966													556,53	1,19
1967	876,80	19,60					1129,98	11,88					1084,1	52,05
1968	600,03	6,90					821,80	7,10					644,84	7,83
1969	679,28	5,98					545,60	3,10					509,14	0,37
1970	373,15	1,50					359,60	0,30					397,44	0,02
1971	577,00	1,33					931,14	4,67					698,51	0,62
1972	519,15	0,87					646,38	0,72					667,64	4,70
1973													718,68	9,44
1974													1197,6	67,94
1975													757,28	15,80
1976													678,44	29,35
1977													847,10	30,67
1978													560,11	7,49
1979													476,74	0,26
1980													689,12	19,03
1981													495,21	9,30
1982													473,16	0,32
1983													374,01	0,02
1984														4,30
1985	1145,7	49,12	1146,7	149,8				17,31						60,59
1986	596,00	3,35	598,00	22,24				18,09						22,52
1987	603,10	11,06	605,10	19,70	411,70	1,34		0,75			394,13	14,75		2,50
1988	684,50	5,54	686,50	21,23	626,10	4,97		9,40			682,55	25,68		0,49
1989	753,10	18,31	755,10	47,79	723,30	17,72		8,83						23,39
1990	212,70	0,02	213,70	2,32	396,30	0,15		0,34			381,55	3,47		0,02
1991	235,30	1,99	237,30	8,62	453,90	0,26		1,40			423,92	10,16		0,68
1992	625,90	3,08	627,90	16,77	393,40	8,43		0,60			491,02	13,52		0,52
1993	375,00	0,02	378,00	1,49	107,60	0,02		0,02	98,20	0,02	127,15	0,09		
1994	348,90	0,09	349,90	1,48	798,60	4,06		3,53	760,63	13,62	571,54	4,58		
1995	700,40	0,58	702,40	17,52	680,90	7,15		1,07	605,80	12,55	620,60	22,37		
1996	696,10	1,60	697,10	24,60	801,00	14,00		3,60	739,33	27,30	618,17	25,63		
1997	588,90	0,02	590,90	6,59	444,90	0,06		0,30	471,97	1,80	565,65	1,93		
1998	317,60	0,02	319,60	0,51	112,60	0,02		0,27	101,43	0,02	189,05	0,02		
1999	445,00	0,02	447,00	3,05	288,60	0,21		2,10	320,60	0,02	394,73	1,08		
Média	572,57	6,27	557,01	22,92	479,92	4,49	773,61	4,94	442,57	7,90	455,01	10,27	672,74	14,68
Desvio padrão	215,55	11,09	241,74	37,34	233,96	5,87	245,68	4,23	278,86	10,44	171,85	9,96	214,95	18,79

Continuação da Tabela 3.6 - Séries sazonais de vazão e precipitação utilizadas.

Ano	Pedra de Abelhas		Piancó		São Fernando		Sítio Volta		Upanema		Ano	Serra Negra do Norte	
	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão	Chuva	Vazão		Chuva	Vazão
1961											1922	346,80	29,10
1962											1923	530,37	40,85
1963					520,32	29,37					1924	532,33	78,34
1964	859,98	62,95	774,90	45,47	748,27	67,49					1925	568,97	42,77
1965	674,27	66,53	572,65	12,08	564,59	71,38					1926	954,13	28,39
1966	517,31	3,92	378,18	0,85	388,02	8,49					1927	716,10	16,48
1967	976,72	107,95	1209,95	75,85	672,31	60,50					1928	452,17	9,84
1968	654,15	28,32	736,13	27,21	537,12	33,62					1929	765,07	30,92
1969	443,08	2,23			457,02	23,72					1930	414,27	10,69
1970	374,02	2,16	360,94	7,16	202,28	3,76					1931	613,77	7,60
1971	643,36	6,53	943,19	52,31	633,17	77,76					1932	165,23	0,25
1972	632,72	17,15	437,99	8,13	457,33	26,12					1933	263,67	16,52
1973	683,15	30,73	689,93	33,02	463,02	36,58					1934	782,40	34,39
1974	1076,1	172,44	1085,55	97,23	893,85	167,5					1935	730,07	45,39
1975	748,60	50,37	750,55	30,23	686,63	73,72			696,94	10,11	1936	351,26	5,45
1976	641,17	21,42	543,69	6,91	440,23	16,14			707,76	11,43	1937	634,52	25,19
1977	740,95	72,36	763,70	34,15	535,94	57,24			773,18	23,44	1938	368,24	15,85
1978	591,87	15,05	499,35	12,77	527,55	29,03			555,30	6,83	1939	502,83	9,86
1979	449,87	4,16	642,81	31,01	354,13	5,67	295,63	0,96	362,35	0,22	1940	636,80	16,75
1980	598,14	38,98	432,82	24,61	342,54	8,95	163,09	2,07	603,17	6,02	1941	387,82	4,39
1981	414,65	31,13	351,49	21,42	364,16	38,37	129,75	9,21	276,65	5,29	1942	202,87	2,64
1982	472,86	5,69	364,47	2,63	363,54	13,37	279,87	1,23	420,00	1,70	1943	283,17	0,93
1983	357,24	0,38	445,53	12,78	189,07	3,54	91,85	0,63	262,50	0,02	1944	475,77	9,93
1984	421,15	32,58	829,30	26,99	330,00	25,37	330,00	4,33	511,05	12,37	1945	729,00	15,58
1985	996,90	151,24	2052,30	166,87	495,10	177,9	495,10	39,36	1389,8	50,80	1946	385,92	7,30
1986	985,25	91,76	668,80	68,78	576,30	77,86	576,30	24,06	748,40	49,67	1947	799,52	12,43
1987	406,50	12,93	414,40	5,69	263,35	13,14	177,50	1,80	415,10	1,60	1948	597,00	15,91
1988	658,40	17,97	575,40	56,30					559,65	5,34	1949	602,37	7,17
1989	665,40	64,02	622,00	33,33					710,15	28,28	1950	625,05	23,21
1990	507,60	2,88							309,00	0,02	1951	364,52	1,52
1991	406,90	8,43	497,00	23,10							1952	515,18	7,22
1992	521,30	8,80	468,00	8,09							1953	227,44	0,93
1993	141,15	0,02	198,20	0,57					98,40	0,02	1954	588,68	8,02
1994	679,45	6,96	447,00	24,49					822,50	12,12	1955	531,78	7,89
1995	688,60	19,49	778,00	38,19					645,40	10,73	1956	615,17	7,15
1996	546,30	21,82	718,00	16,82					788,50	19,86	1957	582,08	10,50
1997	619,25	2,55	556,90	13,10					492,10	1,23	1958	160,70	0,02
1998	254,00	0,02							102,30	0,02	1959	472,75	12,99
1999	413,10	1,13	361,60	8,69					322,25	0,02	1960	598,15	7,97
Média	596,15	32,86	641,54	31,12	480,23	45,86	282,12	9,29	546,63	11,18	Média	514,72	16,11
Desvio padrão	209,07	42,09	335,97	33,28	168,03	45,47	165,42	13,53	282,56	14,63	Desvio padrão	187,14	15,782

3.7.3. CARACTERÍSTICAS FÍSICAS

As extrações das características físicas das bacias neste estudo foram feitas em meio digital através de software de geoprocessamento, como dito anteriormente. O resultado da extração dessas características físicas para as 13 bacias estudadas, encontra-se na Tabela 3.7.

Tabela 3.7 – Características físicas das 13 bacias estudadas.

<i>Características Físicas</i>	<i>Antenor Navarro</i>	<i>Aparecida</i>	<i>Augusto Severo</i>	<i>Emas</i>	<i>Fazenda Angicos</i>	<i>Mossoró</i>	<i>Pau dos Ferros</i>
Código	37260000	37290000	37150000	37360000	37190000	37090000	37030000
Área da Bacia (km ²)	1.514	3.429	1.083	395	2.547	9.447	2.050
Perímetro da Bacia Km	284,09	371,43	154,00	138,26	281,48	556,79	240,23
Cota Inicial (m)	800	800	250	850	250	700	620
Cota Final (m)	200	210	100	200	45	20	200
Desnível (m)	600	590	150	650	205	680	420
Comprimento Total da Drenagem (km)	1.555,81	3.668,13	1.158,13	494,40	2.256,57	7.783,24	2.516,77
Densidade de Drenagem (km ⁻¹)	1,02761	1,06974	1,06937	1,25166	0,88597	0,82388	1,22769
Comprimento do Rio Principal (km)	79,37	142,31	60,36	49,59	141,99	241,15	65,13
Declividade do Rio Principal (km/m)	7,56	4,15	2,48	13,11	1,44	2,82	6,45
Precipitação Média Anual (mm)	824,76	826,31	789,37	778,62	708,55	792,21	863,23

Continuação da Tabela 3.7 – Características físicas das 13 bacias estudadas.

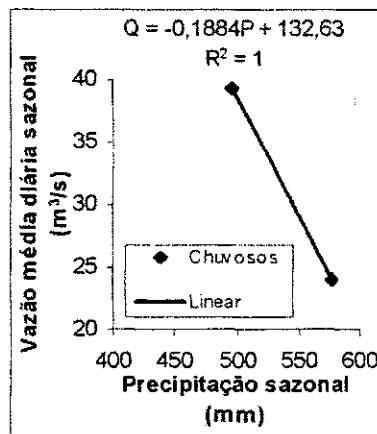
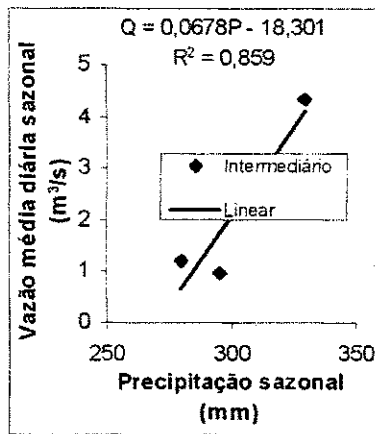
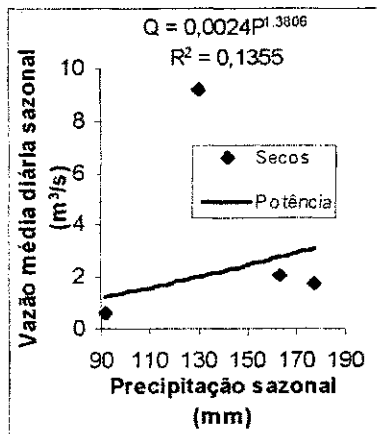
<i>Características Físicas</i>	<i>Pedra de Abelhas</i>	<i>Piancó</i>	<i>São Fernando</i>	<i>Sítio Volta</i>	<i>Upanema</i>	<i>Serra Negra do Norte</i>
Código	37080000	37340000	37570000	37515000	37185000	37440000
Área da Bacia (km ²)	6.481	4.550	9.700	1.794	1.414	3.027
Perímetro da Bacia (km)	455,35	377,77	521,96	244,98	195,96	382,28
Cota Inicial (m)	700	650	600	600	250	720
Cota Final (m)	50	200	100	250	50	250
Desnível (m)	650	450	500	350	200	470
Comprimento Total da Drenagem (km)	7.060,82	6.942,51	9.442,01	1.341,44	1.534,30	3.278,88
Densidade de Drenagem (km ⁻¹)	1,08947	1,52583	0,97340	0,74774	1,08507	1,08321
Comprimento do Rio Principal (km)	168,57	144,74	120,85	59,37	92,66	132,50
Declividade do Rio Principal (km/m)	3,86	3,11	4,14	5,90	2,16	3,55
Precipitação Média Anual (mm)	815,03	907,88	577,52	492,17	767,91	732,15

(mm)

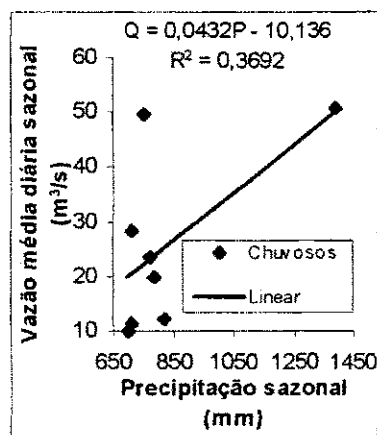
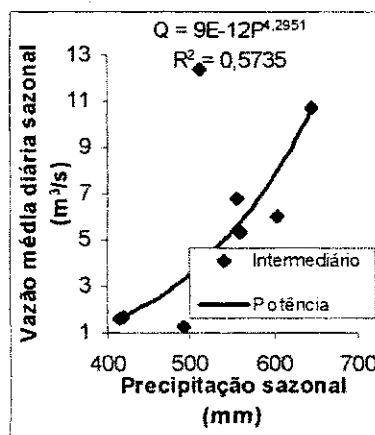
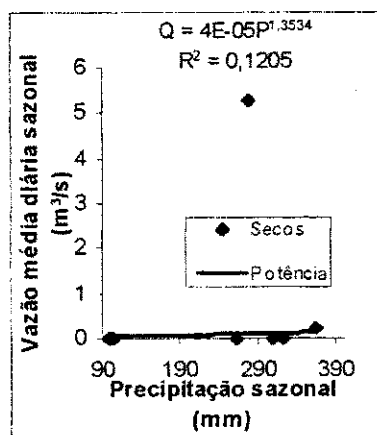
(mm)

(mm)

(j) São Fernando

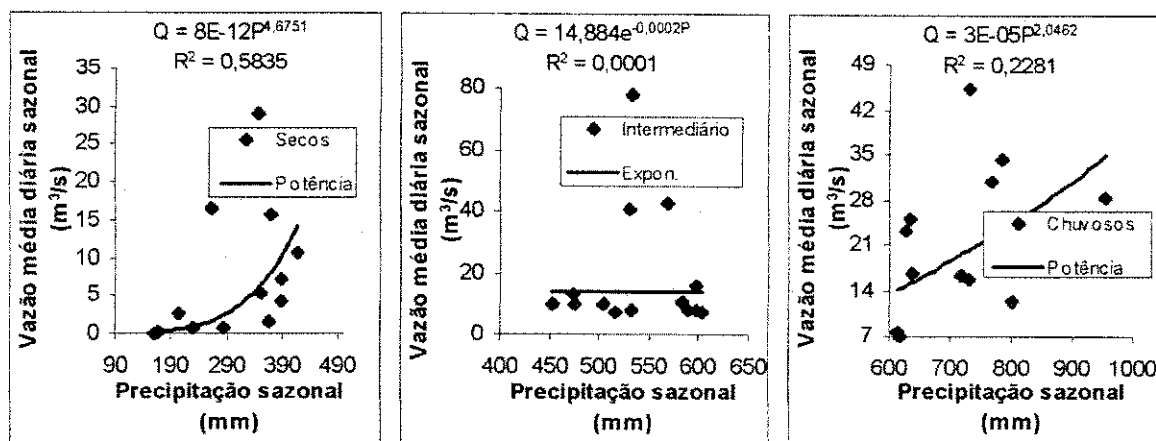


(k) Sítio Volta



(l) Upanena

Figura 4.1 – Relação entre vazão média diária sazonal e precipitação sazonal anual para análise do comportamento das séries com relação aos anos secos, intermediários e chuvosos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

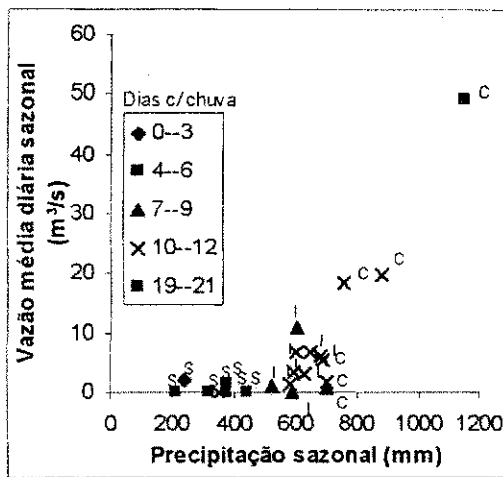


(m) Serra Negra

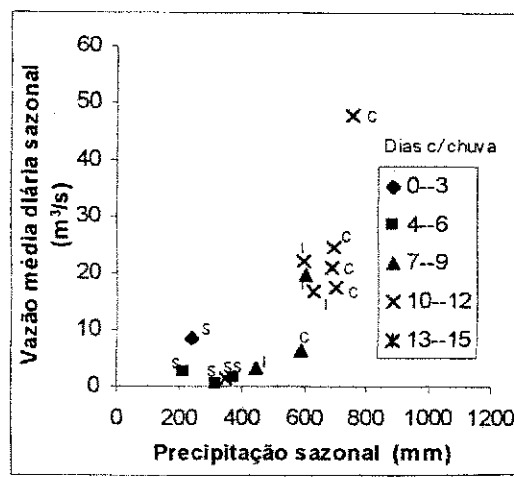
Figura 4.1 – Relação entre vazão média diária sazonal e precipitação sazonal anual para análise do comportamento das séries com relação aos anos secos, intermediários e chuvosos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

4.1.2. ANÁLISE DO PERÍODO CHUVOSO

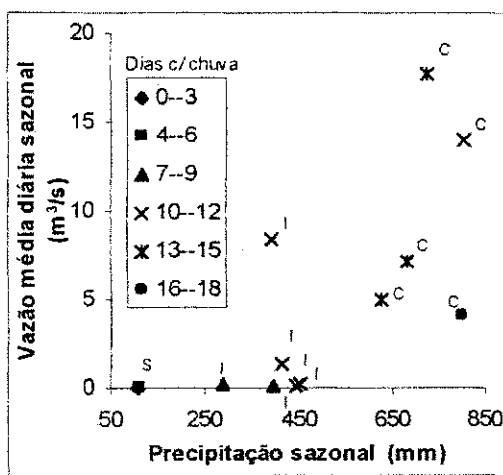
Outros gráficos correspondentes ao período chuvoso da mesma série foram usados para ver se havia uma boa distribuição dos dias de chuva no período estudado e ver a correlação do total pluviométrico com a vazão média diária, tendo em vista que nestes anos com uma boa distribuição pluviométrica e números razoáveis de dias com chuva poderiam apresentar uma contribuição efetiva para a geração do escoamento na bacia. Tal análise é muito usada na agricultura, pra saber especialmente o melhor período de plantio e colheita. Nesta análise pôde-se ver que as chuvas são bem distribuídas durante o período chuvoso nas bacias estudadas, apresentando um bom agrupamento entre os períodos secos, intermediários e chuvosos e o número de dias com chuva. Então nos gráficos a seguir (Figura 4.2) estão os resultados desta etapa, onde as iniciais S, I e C são valores correspondentes respectivamente ao período seco, intermediário e chuvoso; e a legenda ao intervalo do número de dias com chuva.



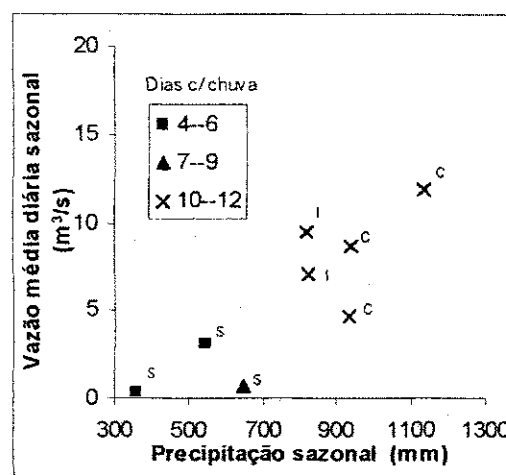
(a) Antenor Navarro



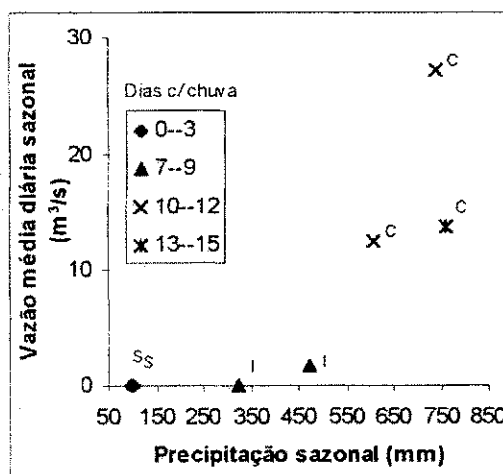
(b) Aparecida



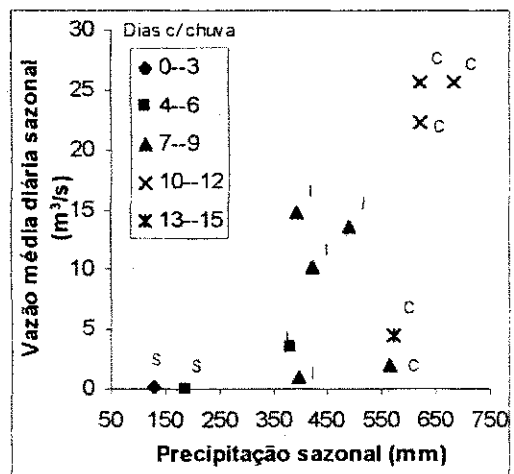
(c) Augusto Severo



(d) Emas

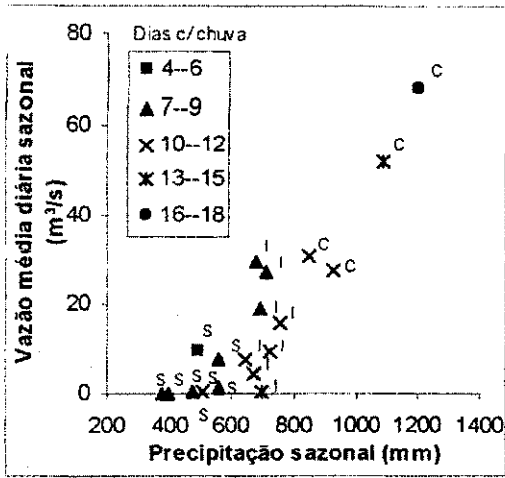


(e) Fazenda Angicos

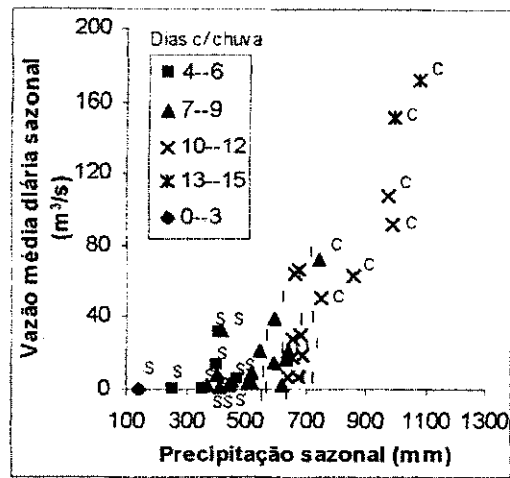


(f) Mossoró

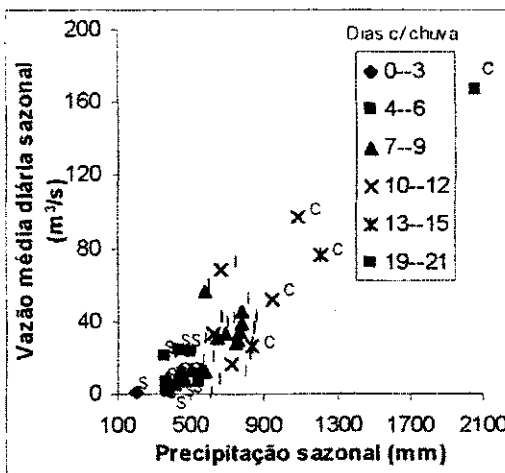
Figura 4.2 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries analisando-se os períodos secos S, intermediários I e chuvosos C e o número de dias com chuva das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sitio Volta; (l) Upanema; e (m) Serra Negra.



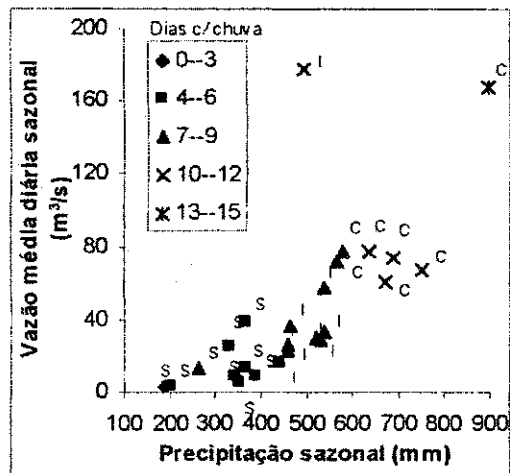
(g) Pau dos Ferros



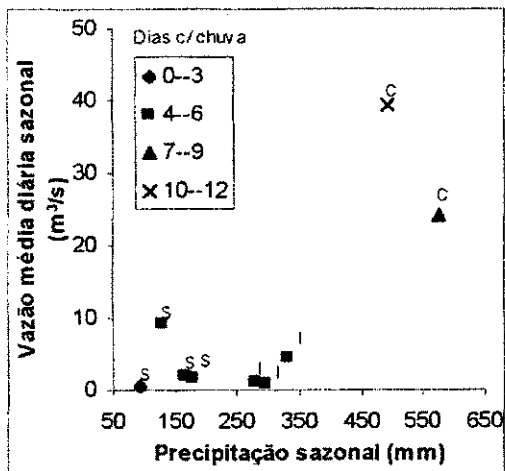
(h) Pedra de Abelhas



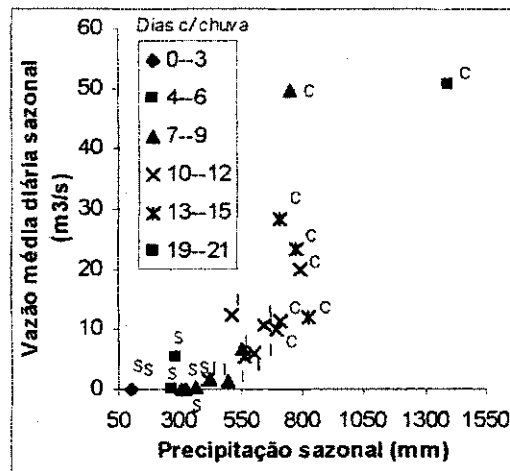
(i) Piancó



(j) São Fernando

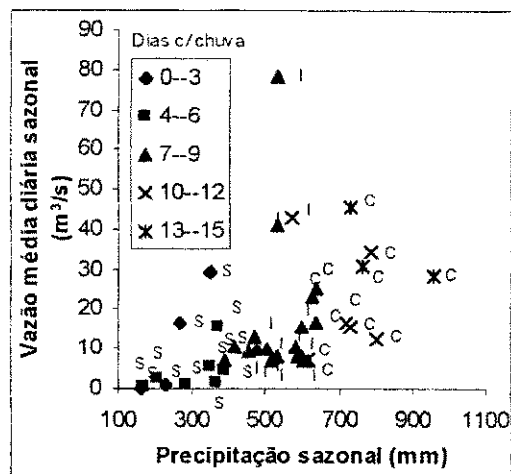


(k) Sítio Volta



(l) Upanema

Figura 4.2 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries analisando-se os períodos secos S, intermediários I e chuvosos C e o número de dias com chuva das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(m) Serra Negra

Figura 4.2 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries analisando-se os períodos secos S, intermediários I e chuvosos C e o número de dias com chuva das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

4.1.3. WINSORIZAÇÃO

A hipótese de se eliminar os extremos de cada série, a Winsorização, não apresentou muita diferença com relação ao R^2 anterior (R^2_{ant}) e o R^2 depois de sanados tais dados (R^2_{fim}); sendo então descartada tal hipótese. Essa análise foi feita eliminando-se apenas os valores de precipitação abaixo de 100mm e acima de 1000mm, isso por haver valores muito altos ou muito baixos nas séries das bacias estudadas, então, esse teste poderia comprovar se tais bacias estavam contribuindo negativamente no resultado encontrado, ou não. Normalmente se retira da amostra, uma faixa de valores geralmente na ordem de 5% do total, mas, como se tratava de uma amostra muito pequena isso não foi possível. Então, foi visto que havia uma certa homogeneidade nos dados utilizados, pois

foram conservadas as características comportamentais ao serem retirados os extremos, não melhorando o resultado. O resultado de tal análise está mostrado na Tabela 4.1.

Tabela 4.1 – Resultado da Winsorização.

Postos	Extremos eliminados		R^2_{ant}	R^2_{fim}
	Ano	Precipitação		
Antenor Navarro	1985	1145,7	0,4636	0,3997
Aparecida	1985	1146,7	0,6486	0,5501
Augusto Severo	-	-	-	-
Emas	1967	1129,9	0,7663	0,7319
Fazenda Angicos	1993	98,2	0,7985	0,7632
Mossoró	-	-	-	-
Pau dos Ferros	1974/67	1197,56/1084,11	0,7242	0,7254
Pedra de Abelhas	1974	1076,1	0,7536	0,7408
	1985	2052,3	0,6865	0,6658
Piancó	1985/1974/1967	2052,3/1085,5/1209,9	0,6865	0,6214
São Fernando	-	-	-	-
Sítio Volta	1983	91,85	0,4032	0,2630
Upanema	1985	1389,85	0,7289	0,7168
	1993	98,4	0,7289	0,6938
Serra Negra	-	-	-	-

4.1.4. DIAGRAMAS DE DISPERSÃO BACIA A BACIA

Foi verificado no estudo de UFPB/ANEEL (2001), já citado anteriormente, que ao analisar-se os valores médios anuais nessa região, o modelo que melhor se ajustaria seria o Potência. Então, foram feitos diagramas de dispersão de vazão e precipitação observadas para se confirmar essa análise para os valores anuais médios sazonais.

Para tal análise foram testados três tipos de modelos: o Linear, o Exponencial e o Potência e observou-se que para a maioria das bacias, assim como no estudo anterior, o melhor modelo ajustado foi o Potência. Daí veio a decisão de se adotar um modelo único para todas as bacias, já que não houve perdas significativas.

Mas, apesar da constatação de que a relação entre as séries de precipitação e vazão das bacias é aproximadamente potência, é nítida a existência de dados bastante dispersos. Ao se analisar, constata-se que isso reflete no resultado do R^2 e essa dispersão pode estar relacionada a alguma falha na coleta e transcrição dos dados, à não homogeneidade na

série ou à falta de correlação simplesmente o que é uma observação óbvia, mas que nenhuma outra observação mais clara pôde ser detectada para tais dispersões.

Os gráficos correspondentes a esta análise estão com os coeficientes de determinação indicados para o período de cada bacia separadamente, mostrados a seguir na Figura 4.3.

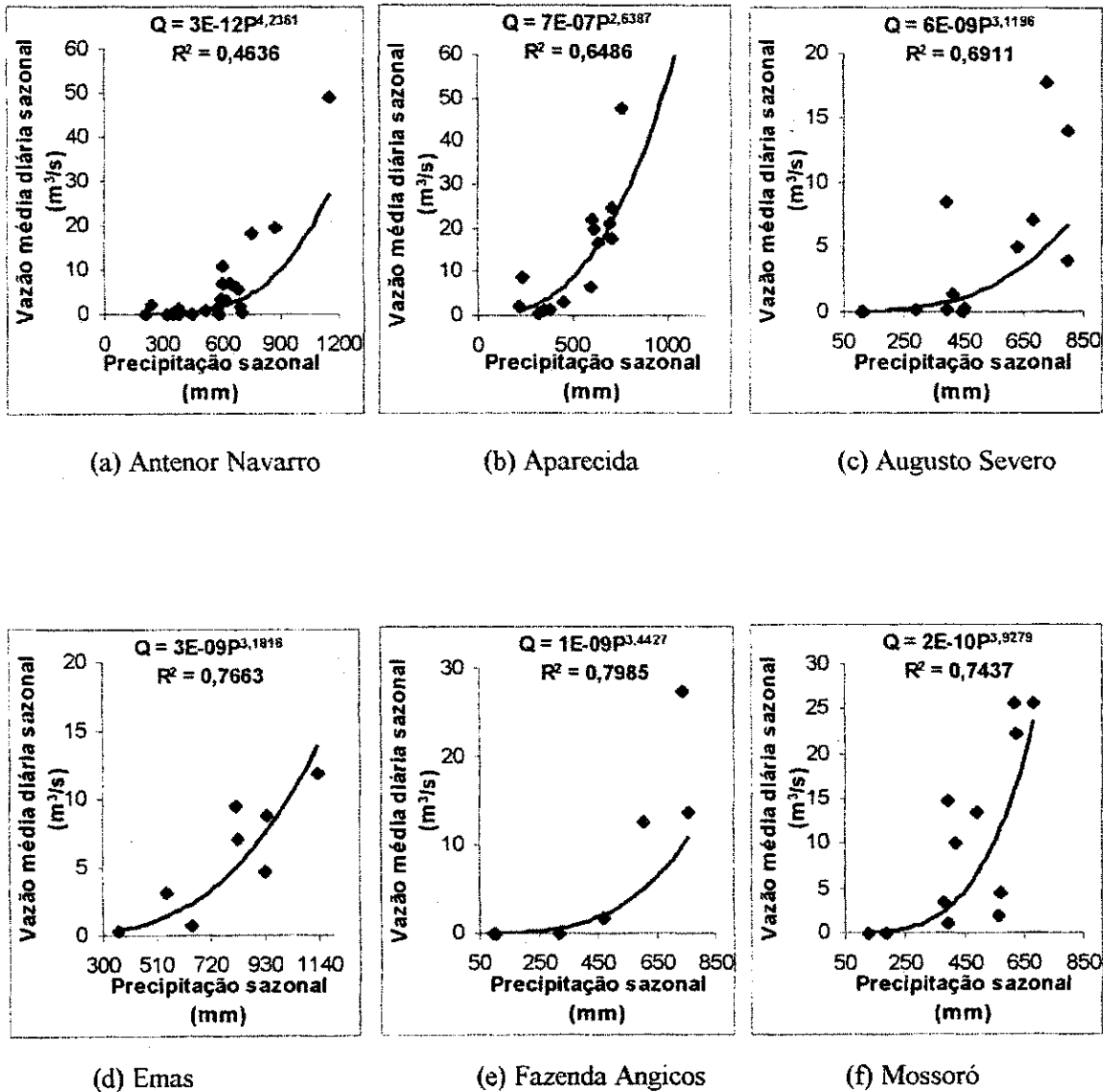
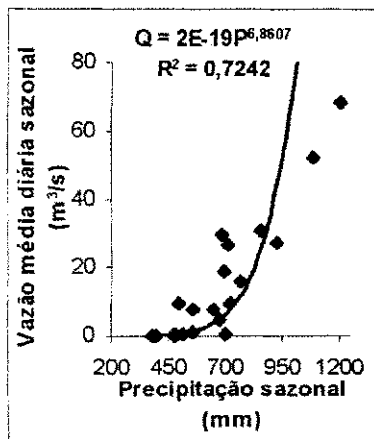
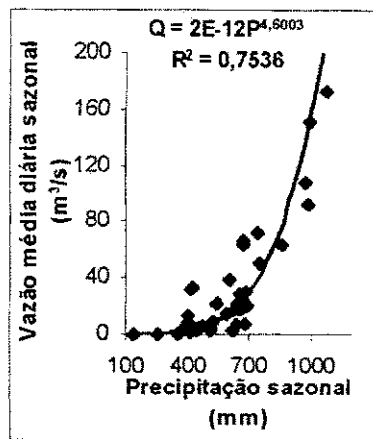


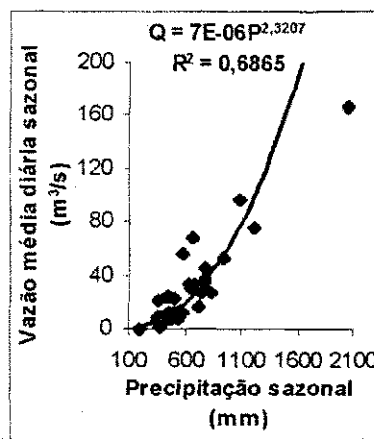
Figura 4.3 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.



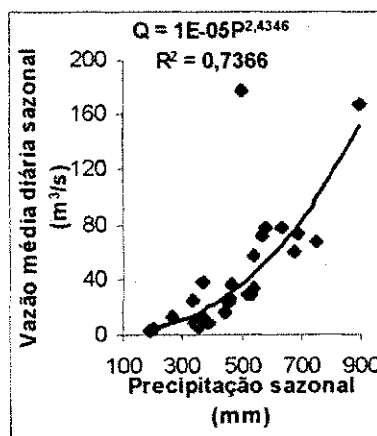
(g) Pau dos Ferros



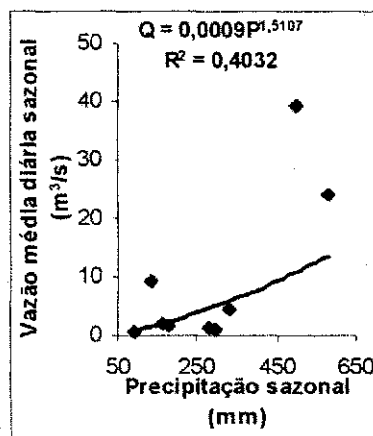
(h) Pedra de Abelhas



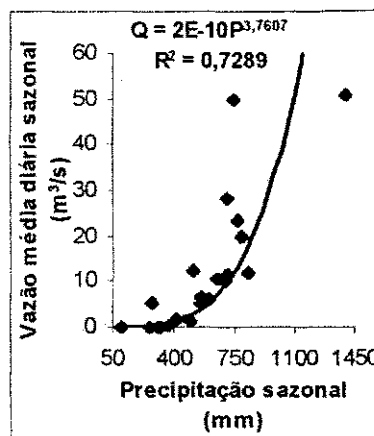
(i) Piancó



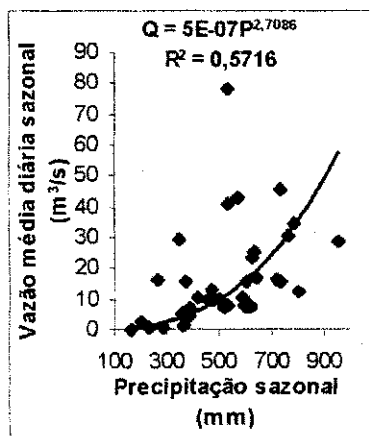
(j) São Fernando



(k) Sítio Volta



(l) Upanema



(m) Serra Negra

Figura 4.3 – Relação entre vazão média diária sazonal anual e precipitação sazonal anual para análise do comportamento das séries das bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

Sabendo-se que o comportamento de cada bacia foi identificado; que o melhor ajuste aos dados seria mesmo o tipo Potência; e visto que esse modelo pode ser aplicado a todas as bacias sem grandes perdas, partiu-se para a construção da tabela para o ajuste do modelo, através da regressão, com a junção das séries sazonais de todas as bacias ano a ano e a estas foram aplicados o logaritmo (Tabela A.1).

4.2. ANÁLISE DE REGRESSÃO

À frente, na Tabela 4.2, tem-se o resultado da correlação entre as variáveis independentes (P – precipitação (mm); A – área da bacia (km²); DD – densidade total de drenagem (km⁻¹); L – comprimento do rio principal (km); DEC – declividade do rio principal (km/m)) e a variável dependente (Q – vazão média diária sazonal (m³/s)) na análise estatística para o modelo, ou seja a matriz de correlação de todas as variáveis uma com as outras. Na primeira linha tem-se o valor da correlação de Pearson (r) entre as variáveis, na segunda linha tem-se o nível descritivo associado ao teste de hipótese nula - Ho.

Ao analisar-se os dados desta tabela, pode-se comprovar que de todas as variáveis independentes escolhidas para descrever o modelo, a declividade do rio principal é a única variável inversamente proporcional à variável dependente e que esta não apresenta nenhuma correlação com a variável dependente observada. Logo, nesta primeira análise, esta variável já é uma forte candidata à exclusão do modelo. Pode-se observar ainda que existe uma correlação muito boa entre a área e o comprimento do rio principal ($r = 0,8179$) o que não se tornaria uma boa opção mantê-las no modelo, ou seja, apenas uma destas deveria permanecer, mas, as duas variáveis foram mantidas quando na análise de regressão do programa SOC, e o que explica isso é que de certa forma houve alguma melhoria do modelo quando se manteve as duas variáveis. Para a precipitação, em função de se pretender fazer previsões de vazão a partir de previsões de precipitação, esta foi sempre mantida no modelo em todas as interações geradas, como já foi explicado neste, e os resultados comprovaram ainda a sua boa aceitação no modelo.

Tabela 4.2 – Matriz de correlação de Pearson r e nível descritivo do teste $H_0: \text{Correl} = 0$, entre as variáveis do modelo.

	$\ln Q$	$\ln P$	$\ln A$	$\ln DD$	$\ln DEC$	$\ln L$
$\ln Q$	1 0,000					
$\ln P$	0,6736 0,000	1 0,000				
$\ln A$	0,3536 0,000	-0,0184 0,767	1 0,000			
$\ln DD$	0,1592 0,010	0,3071 0,000	-0,0791 0,202	1 0,000		
$\ln DEC$	-0,0048 0,938	0,1528 0,013	-0,2763 0,000	0,0164 0,791	1 0,000	
$\ln L$	0,2621 0,000	0,0079 0,898	0,8179 0,000	0,008 0,898	-0,4802 0,000	1 0,000

Para a análise da matriz de correlação pode-se verificar que uma correlação é boa, se esta estiver dentro do intervalo $0,6 \leq r \leq 1$ (valores correspondentes à primeira linha de cada uma das relações na Tabela 4.2) e $\alpha = 0,05$ ou $\alpha = 0,01$ (valores correspondentes à linha logo abaixo dos valores de correlação na Tabela 4.2), do teste de hipótese $H_0: \text{Correl} = 0$. Sendo pois, o valor negativo devido ao resultado ser inversamente proporcional.

Na análise de regressão que o SOC registrou, a melhor iteração foi a que usou todas as variáveis envolvidas. Esta sempre é escolhida observando-se principalmente o melhor coeficiente de determinação dentro da análise de regressão, seguido do teste-F, da análise do coeficiente de variação que representa a dispersão, ou seja, se os dados são mais ou menos precisos e do resultado da estatística descritiva (média, variância e desvio padrão). Visto isso, parte-se para a montagem da equação através dos resultados fornecidos por essa iteração. O resultado da análise descritiva da melhor iteração deste estudo, encontra-se descrito na Tabela 4.3,

Tabela 4.3 – Estatística descritiva das variáveis para a construção do modelo.

<i>Variável</i>	<i>Média</i>	<i>Variância</i>	<i>Desvio Padrão</i>
ln Q	1,73889313	5,06963440	2,25158486
ln P	6,20755725	0,25688826	0,50684146
ln A	8,05156489	0,58348451	0,76386158
ln DD	0,09221374	0,02690159	0,16401705
ln DEC	1,35480916	0,18674077	0,43213513
ln L	4,73015267	0,15761990	0,39701373

As tabelas abaixo mostram a análise de variância também fornecida pelo ajuste do modelo na regressão. Sendo que, na Tabela 4.4 se encontra a análise estatística da escolha do modelo. A Tabela 4.5, mostra os resultados da análise das estimativas do ajuste dos parâmetros para montagem da equação, ou seja, o valor do coeficiente de cada uma das variáveis da equação proposta. E na Tabela 5.6 encontra-se as demais estatísticas, incluindo esta o coeficiente de determinação da regressão.

Tabela 4.4 - Análise de variância.

<i>Fonte de variação</i>	<i>gl</i>	<i>Soma de quadrados</i>	<i>Quadrado médio</i>	<i>Valor F</i>	<i>PR > F</i>
Modelo	5	788,42033703	157,68406741	75,4872	0,0000
Resíduo	256	534,75424197	2,08888376		
Total	261	1323,17457901			

Tabela 4.5 - Análise das estimativas para o modelo.

<i>Parâmetro</i>	<i>Estimativa</i>	<i>Desvio Padrão</i>	<i>Valor T </i>	<i>PR > T </i>
Coef. linear	-24,06889576	1,71099312	14,0672	0,000
P	3,09700970	0,18872253	16,4104	0,000
A	1,44407725	0,21206675	6,8095	0,000
DD	-0,18972312	0,57905179	0,3276	0,744
DEC	-0,30414633	0,24725704	1,2301	0,221
L	-0,97557158	0,44711353	2,1819	0,031

Tabela 4.6 – Estatísticas para a análise do modelo.

<i>Parâmetro</i>	<i>Valor</i>
Média	1,7388931
Raiz Quad. QMres.	1,4452971
Coef. Variação	83,1159255
Coef. Determinação	0,5958551
Coef. Determinação Ajustado	0,5879617

O modelo desenvolvido neste estudo apresenta um conjunto de variáveis independentes que melhor representa o comportamento hidrológico da região homogênea, ou seja, representa a variável dependente, que neste caso é a vazão Q . Este é escrito da seguinte forma, já demonstrado anteriormente:

$$Q = a P^b A^c L^d \quad (4.1)$$

onde a , b e c são os parâmetros encontrados na análise das estimativas (Tabela 4.5), respectivamente: a exponencial do valor da estimativa correspondente ao coeficiente linear; a estimativa correspondente ao valor do parâmetro P ; a estimativa correspondente ao valor do parâmetro A ; e a estimativa correspondente ao valor do parâmetro L . E ainda, P , A , e L que são as variáveis independentes medidas, que já foram citadas antes, mas para não deixar dúvidas, estas são respectivamente a precipitação na bacia, a área da bacia e o comprimento do rio principal desta, e Q é a variável dependente e neste caso, a vazão desejada. Ou seja, é obtida com essa relação, a vazão da bacia em função das características físicas desta e da variável meteorológica prevista, a previsão de precipitação.

Pode-se ver também na Tabela 4.6, que o coeficiente de determinação para esta interação foi aproximadamente $R^2 = 0,60$, o que significa que, como em uma estimativa pontual, cerca de 60% da variância de Q é explicada pelas variáveis P , A e L . Além disso, dado um Q , 40% da variância permanece não explicada, logo, mais da metade da variância de Q está associada com mudanças nessas variáveis P , A e L .

Assim, para o total de 262 anos, das 162 estações pluviométricas e 13 estações fluviométricas, consideradas neste estudo, e considerando as variáveis independentes: P , A , DD , DEC e L , a equação obtida é:

$$Q = 3,5238 \cdot 10^{-11} P^{3,0970} A^{1,4441} L^{-0,9756} \quad (4.2)$$

As demais variáveis como a densidade de drenagem e a declividade do rio principal, segundo o teste-F aplicado, ao nível de significância de 0,05, como é visto na Tabela 4.5, não contribuem para a melhoria das equações. A equação 4.2 é, portanto, desenvolvida com as 13 bacias estudadas, e que pode ser considerada apta para uso nas bacias sem dados para a região estudada neste. Devendo esta ser usada com cautela, em virtude de ter sido derivada de um pequeno número de postos, com séries curtas, em bacias com forte impacto antrópico e não sendo possível um período único de dados devido a séries não serem completas.

Uma vez de posse da equação, o usuário pode utilizá-la nas bacias onde não se têm dados, se o mesmo desejar estimar o valor da vazão média diária sazonal em um certo local, dentro da região estudada. Para isso então, o mesmo deve apenas calcular as características físicas da bacia (área e comprimento do rio principal) e junto com a precipitação média sazonal observada ou a previsão de precipitação sazonal, substituí-las na equação 4.2, e obter a vazão média diária prevista para o período chuvoso em questão.

E ainda, objetivando-se dar uma margem de confiança na aplicação da equação encontrada (4.2), fez-se uso da técnica de cálculo do intervalo de confiança, já que este dá uma idéia da qualidade da regressão. O intervalo de confiança informa o intervalo com o qual a probabilidade de $1 - \alpha$ das estimativas para os valores fixados das vazões observadas, incluir os verdadeiros valores, desconhecidos, da vazão calculada. Logo, não se tem a certeza de que o intervalo obtido inclua a vazão calculada, mas, sabe-se que para $\alpha=0,05$ há uma chance de 95% de o valor estimado está dentro do intervalo inferior (Int. Inf) e superior (Int. Sup).

Encontra-se no anexo A.2 a tabela contendo os limites encontrados nesta etapa.

4.3. ANÁLISE DO AJUSTE E VALIDAÇÃO DO MODELO

Graficamente, na fase de ajuste, o modelo apresentou um coeficiente de determinação de $R^2 = 0,67$ para todas as bacias juntas (Figura 4.4). Esta análise gráfica entre os valores da vazão observada versus a vazão calculada, fazendo-se uso da equação resultante do ajuste para com as 13 bacias estudadas, mostrou que não houve nenhuma singularidade que comprometesse o modelo, apesar de os dados não mostrarem uma boa aproximação com a reta de igual valor (reta 45°), e apresentar algumas dispersões. A

Figura 4.5. mostra o com mais detalhe a Figura 4.4, onde esta apresenta as vazões apenas até 60 m³/s.

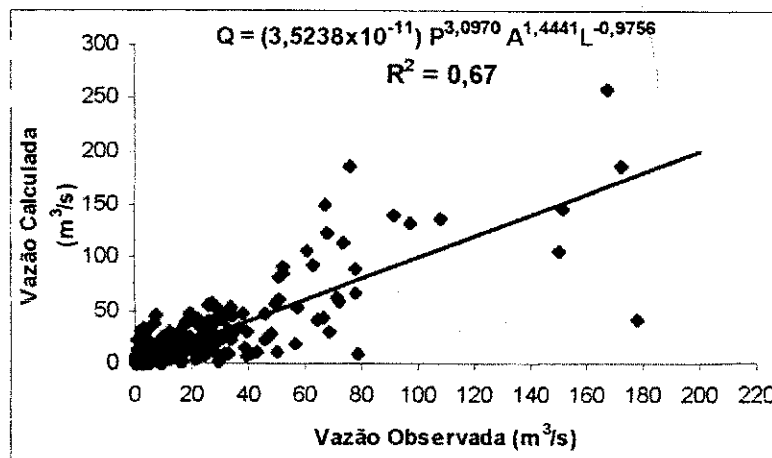


Figura 4.4 – Vazão calculada versus vazão observada para todos os dados de todas as bacias na fase de ajuste do modelo.

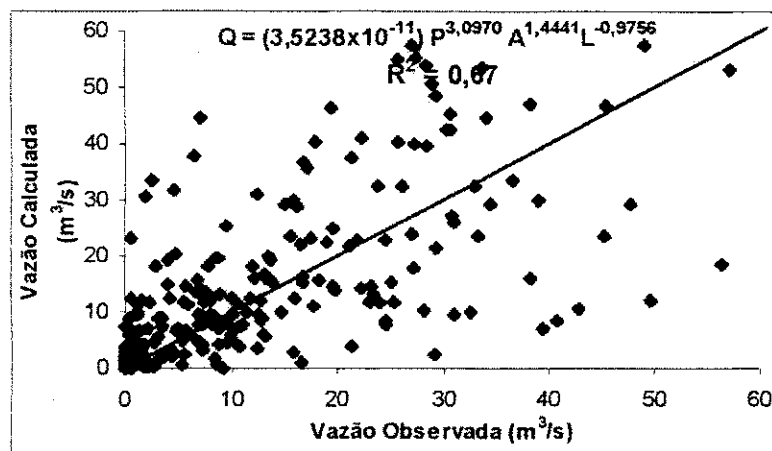


Figura 4.5 – Detalhe da Figura 4.4: vazão calculada versus vazão observada para todos os dados de todas as bacias na fase de ajuste do modelo.

Quando se retira qualquer bacia e ajusta-se o modelo às demais, não há grandes mudanças no resultado encontrado, estas apresentam R² com uma semelhança bastante animadora. Isso mostra que as bacias tiveram parcelas iguais de contribuição para a geração do modelo (ver Tabela 4.7 a seguir).

Tabela 4.7 - Resultados da validação cruzada.

<i>Bacia usada na validação</i>	<i>Equação obtida sem os dados da bacia</i>	<i>R² do ajuste</i>	<i>R² da validação: apenas com a bacia</i>
Antenor Navarro	$Q = 4,9113 \times 10^{-11} P^{3,0214} A^{1,2104} L^{-0,6563}$	$R^2 = 0,66$	$R^2 = 0,89$
Aparecida	$Q = 5,8447 \times 10^{-11} P^{3,1042} A^{1,4906} L^{-1,1615}$	$R^2 = 0,66$	$R^2 = 0,97$
Augusto Severo	$Q = 2,3768 \times 10^{-10} P^{3,0840} A^{1,4591} L^{-1,3134}$	$R^2 = 0,64$	$R^2 = 0,51$
Emas	$Q = 3,9018 \times 10^{-11} P^{3,0784} A^{1,7042} L^{-1,3216}$	$R^2 = 0,65$	$R^2 = 0,70$
Faz. Angicos	$Q = 8,6213 \times 10^{-11} P^{3,0447} A^{1,3610} L^{-0,9065}$	$R^2 = 0,67$	$R^2 = 0,80$
Mossoró	$Q = 9,4246 \times 10^{-12} P^{3,0503} A^{1,3309} L^{-0,4396}$	$R^2 = 0,69$	$R^2 = 0,58$
Pau dos Ferros	$Q = 2,7714 \times 10^{-10} P^{3,0335} A^{1,7355} L^{-1,8397}$	$R^2 = 0,62$	$R^2 = 0,86$
P. de Abelhas	$Q = 2,5139 \times 10^{-11} P^{2,9480} A^{1,4423} L^{-0,7316}$	$R^2 = 0,60$	$R^2 = 0,85$
Piancó	$Q = 3,0454 \times 10^{-11} P^{3,2657} A^{1,2939} L^{-0,9409}$	$R^2 = 0,68$	$R^2 = 0,69$
São Fernando	$Q = 4,1743 \times 10^{-11} P^{3,1083} A^{0,9372} L^{-0,2039}$	$R^2 = 0,71$	$R^2 = 0,51$
Sito Volta	$Q = 1,2442 \times 10^{-12} P^{3,3463} A^{1,2600} L^{-0,3105}$	$R^2 = 0,66$	$R^2 = 0,65$
Upanema	$Q = 1,0170 \times 10^{-10} P^{2,9653} A^{1,4108} L^{-0,9673}$	$R^2 = 0,67$	$R^2 = 0,56$
Serra Negra	$Q = 5,5580 \times 10^{-11} P^{3,1586} A^{1,6580} L^{-1,5287}$	$R^2 = 0,66$	$R^2 = 0,13$

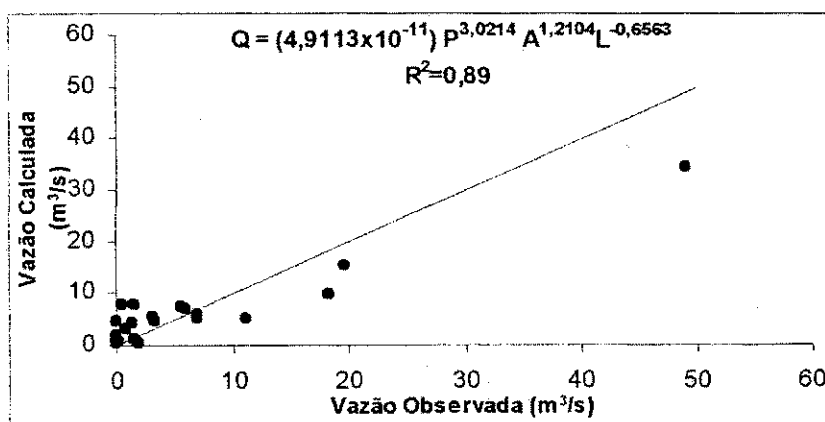
Deve-se informar que, o ano de 1985 do posto Piancó foi retirado da série apenas na fase de validação do modelo por este apresentar um valor muito alto comparado com os demais de toda a série. Neste ano a precipitação sazonal registrada foi de 2052,30 mm, o que não é um fato normal para a região. Vale salientar que o mesmo foi usado na fase de ajuste, haja vista que a série neste caso foi logaritimizada tornando este valor normalizado, porém não mais tão diferente dos demais.

A validação do modelo usando a técnica da validação cruzada, onde se deixa sempre uma bacia de fora na calibração para usá-la para validar o modelo, mostrou que boa parte dos coeficientes de determinação equiparou-se aos do ajuste, e que não houve muita variação com relação aos parâmetros encontrados nas equações com a retirada de uma bacia do modelo. Como foi mostrado na Tabela 4.7. Os resultados nesta tabela encontram-se dispostos da seguinte maneira: as bacias usadas neste estudo encontram-se na primeira coluna; na segunda coluna, tem-se a equação que foi gerada com os dados de 12 bacias, tendo como exceção os dados da bacia da linha correspondente; na terceira coluna encontra-se o coeficiente de determinação R^2 , calculado no ajuste da equação; na última

coluna tem-se o coeficiente de determinação R^2 , resultante da aplicação da equação na bacia da linha correspondente.

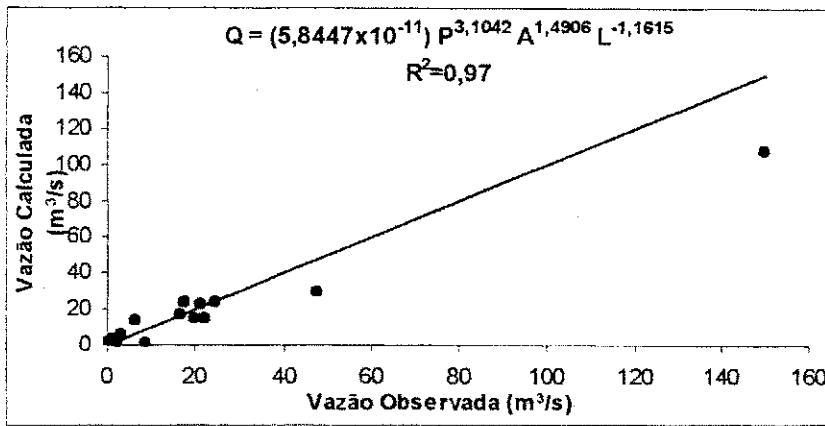
Nesta etapa, os resultados dos coeficientes de correlação melhoram bastante para 6 bacias, mantêm-se iguais em 2 e decrescem também em 5. Para o posto Serra Negra, o R^2 é tão baixo que desaconselha o uso dos seus dados. O que aconteceu neste caso foi que a série que gerou esta equação é bem de um período bem distinto (1961 a 1999) da série de Serra Negra (1922 a 1960) que ficou fora do ajuste, podendo assim explicar as diferenças.

Os gráficos da validação (Figura 4.6) também apresentaram uma boa aproximação com a reta de igual valor (reta 45°), com apenas algumas dispersões que justificam os 33% da variância que permaneceu não explicada pelas variáveis independentes. Neles pode-se observar ainda que em alguns casos há uma certa tendência positiva, ou seja, há uma superestimação no modelo, como é o caso de Augusto Severo, Mossoró e Pau dos Ferros, enquanto que, nas bacias de São Fernando e Sítio Volta, houve o inverso; isto é, o modelo é subestimado. Os resultados das vazões calculadas e observadas também podem ser visualizados em gráficos de barras (Figuras 4.7), onde nestes encontram-se ainda, os resíduos, para que se tenha uma melhor visualização dos anos que apresentaram valores discrepantes.

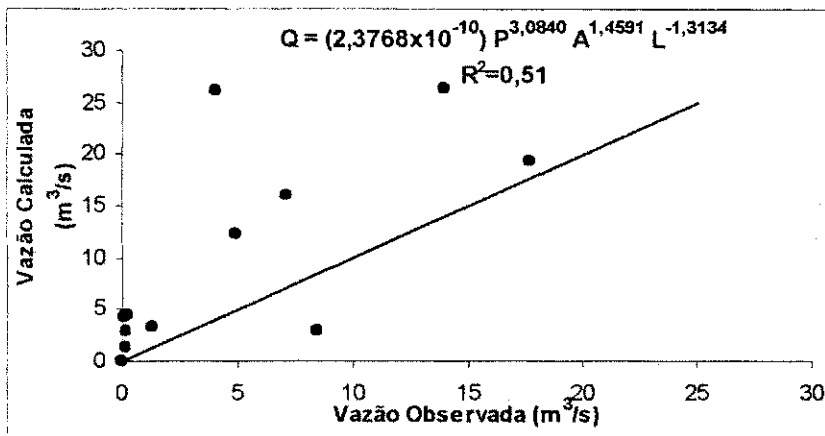


(a) Antenor Navarro

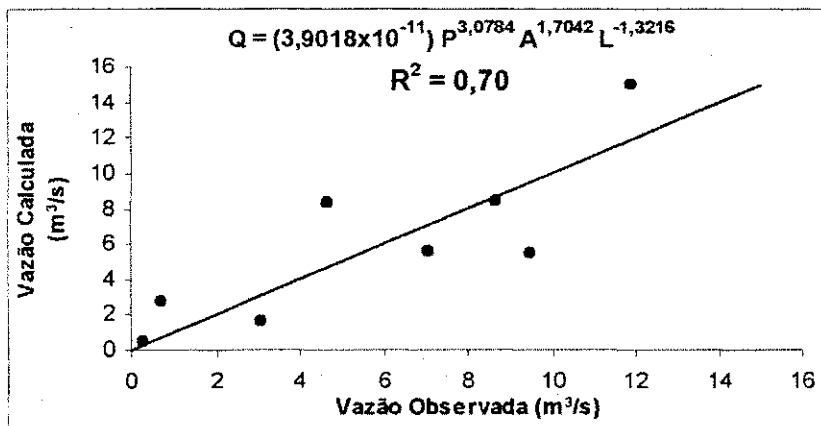
Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.



(b) Aparecida

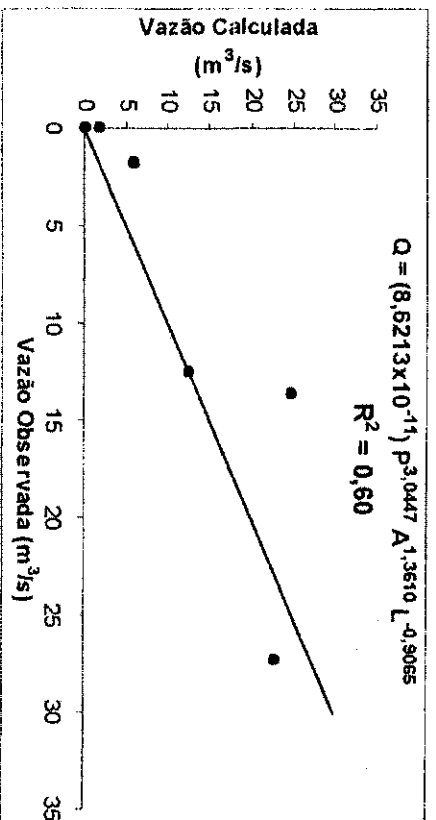


(c) Augusto Severo

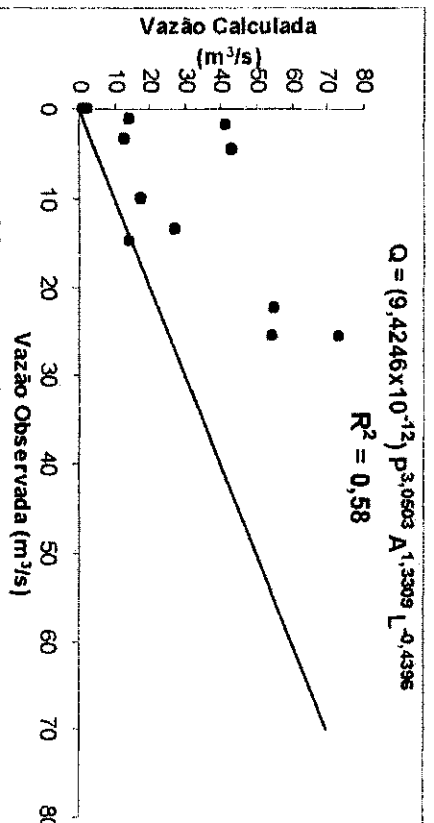


(d) Emas

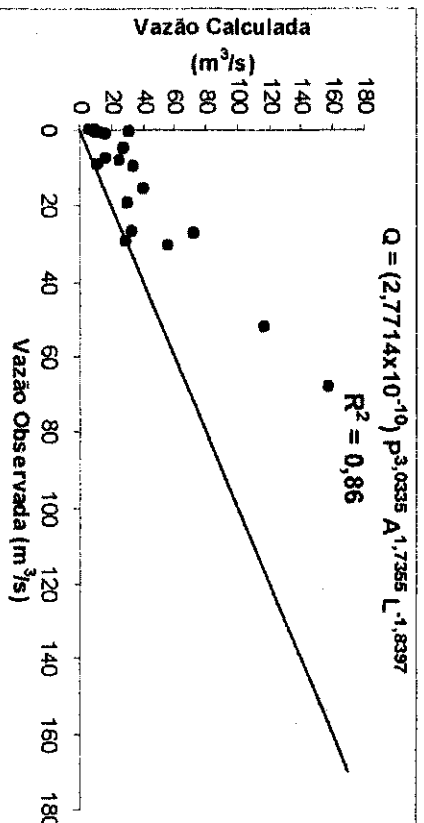
Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(c) Fazenda Angicos

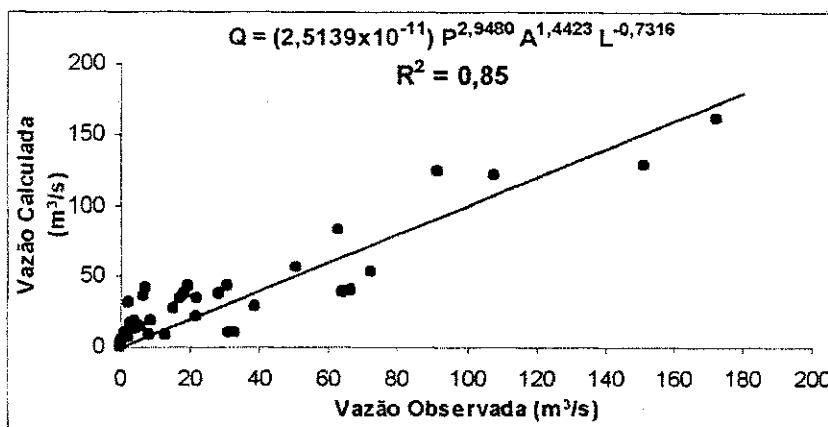


(f) Mossoró

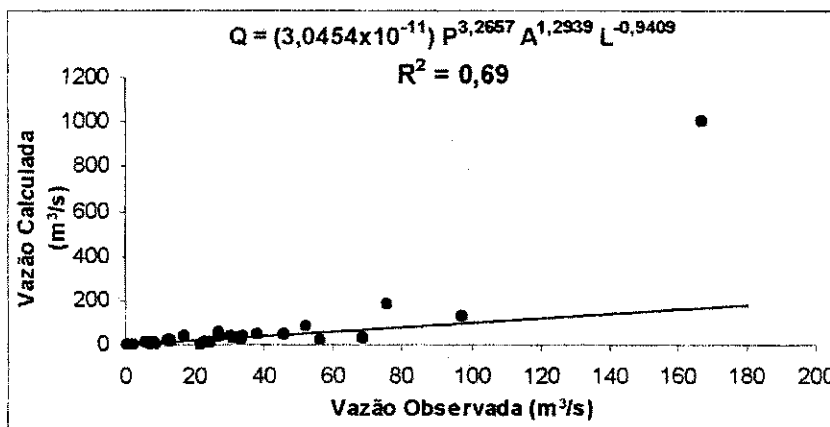


(g) Pau dos Ferros

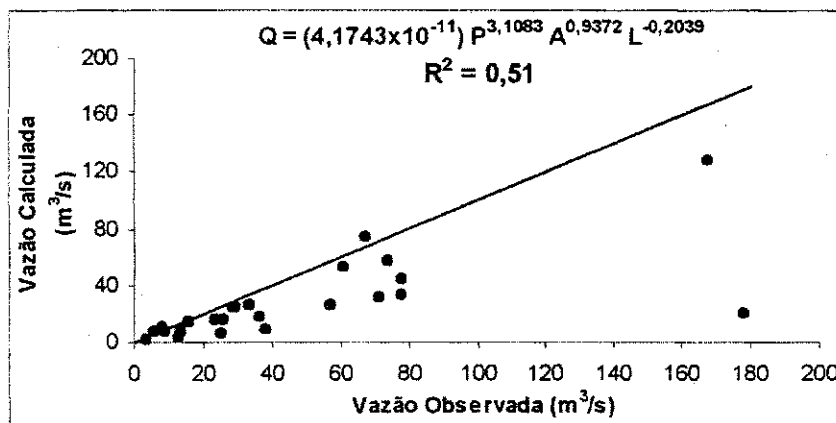
Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos pluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(h) Pedra de Abelhas

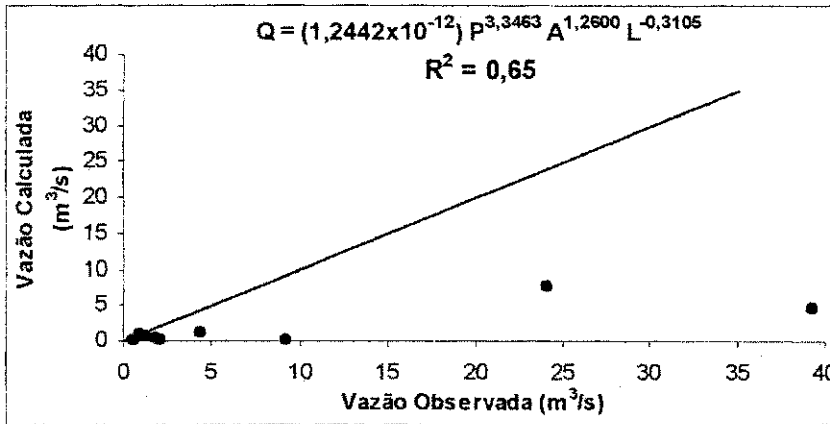


(i) Piancó

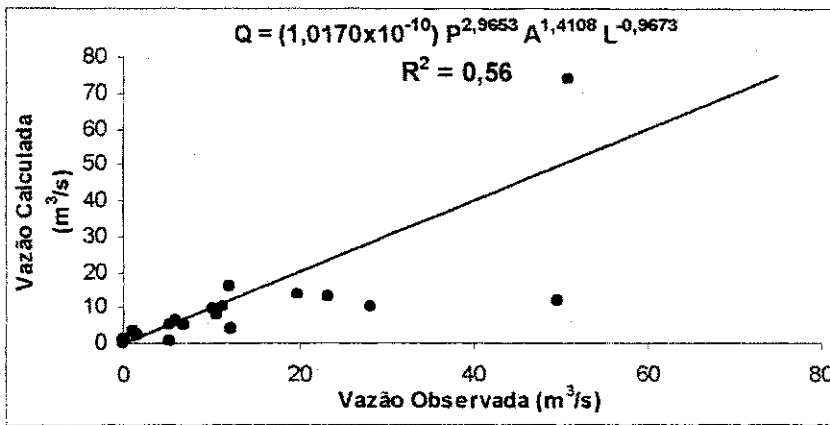


(j) São Fernando

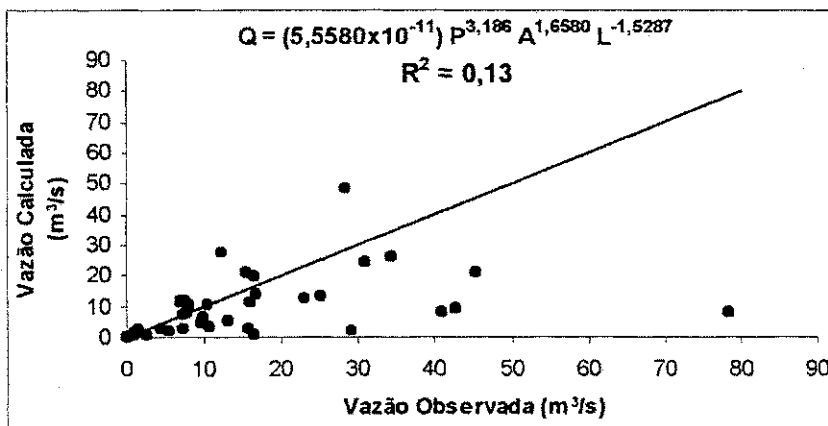
Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(k) Sitio Volta

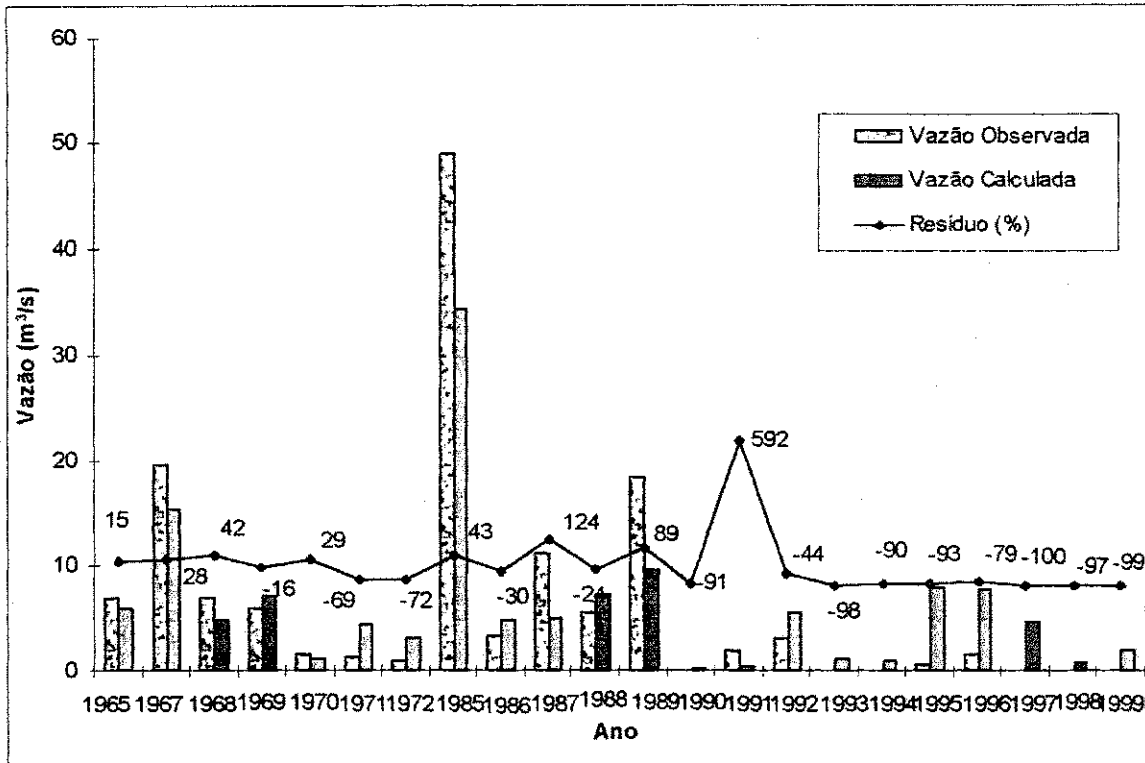


(l) Upanema

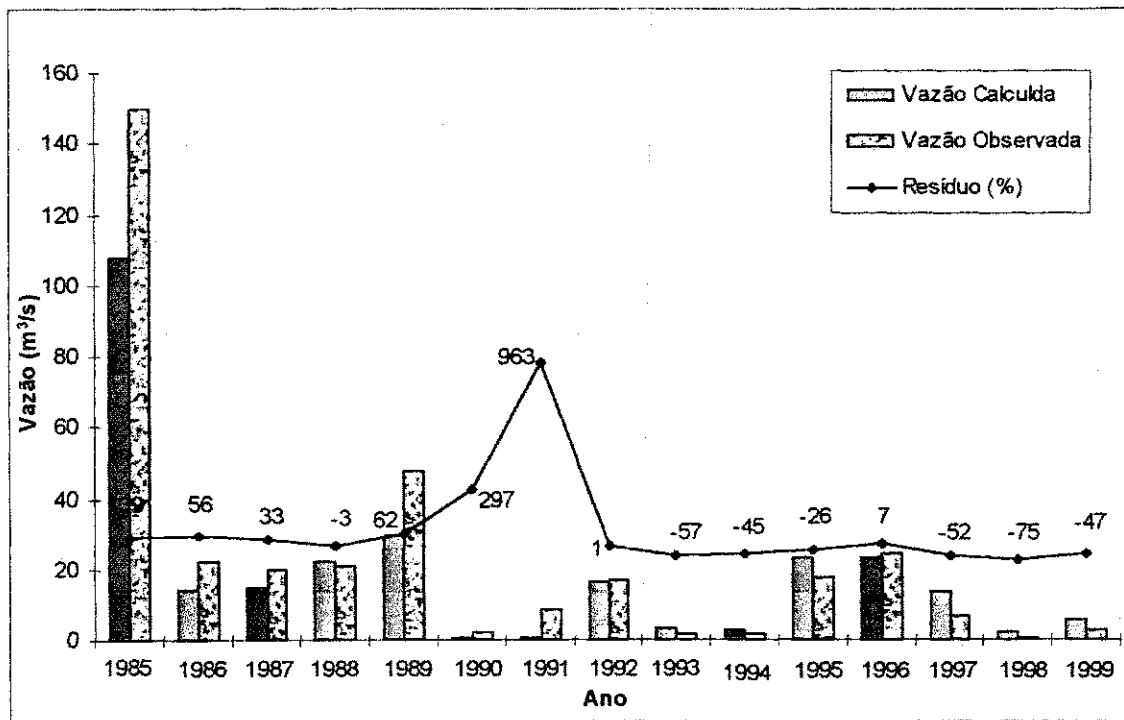


(m) Serra Negra.

Figura 4.6 – Resultados da Validação Cruzada, apenas com uso das bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

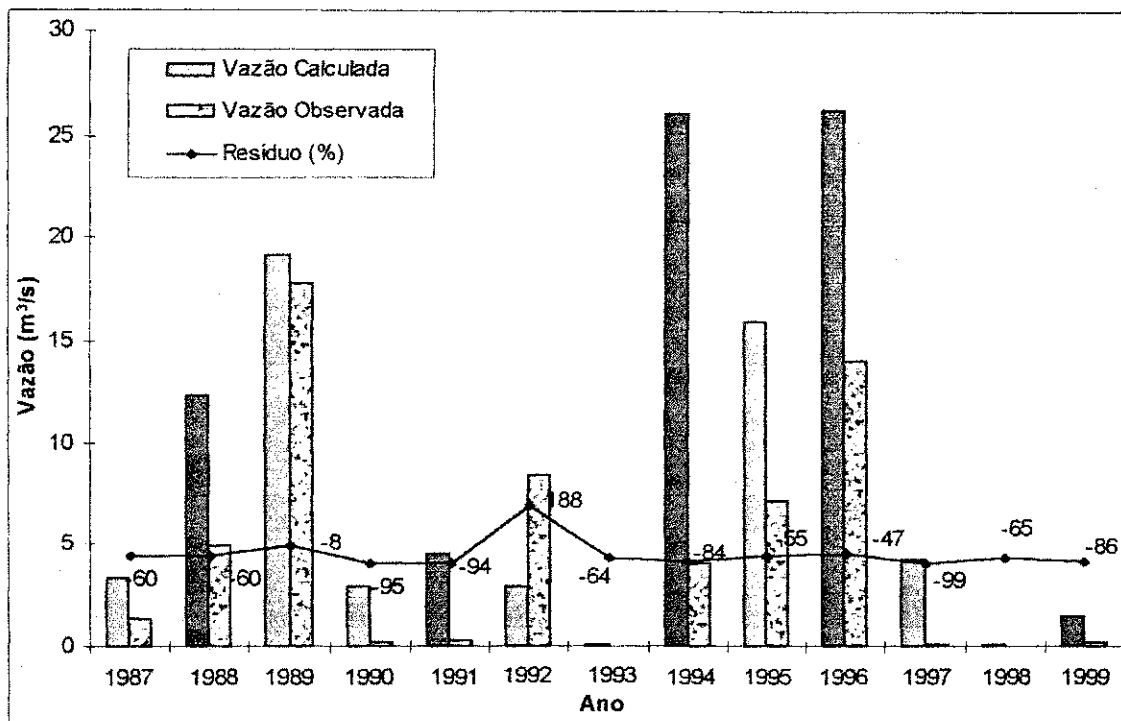


(a) Antenor Navarro

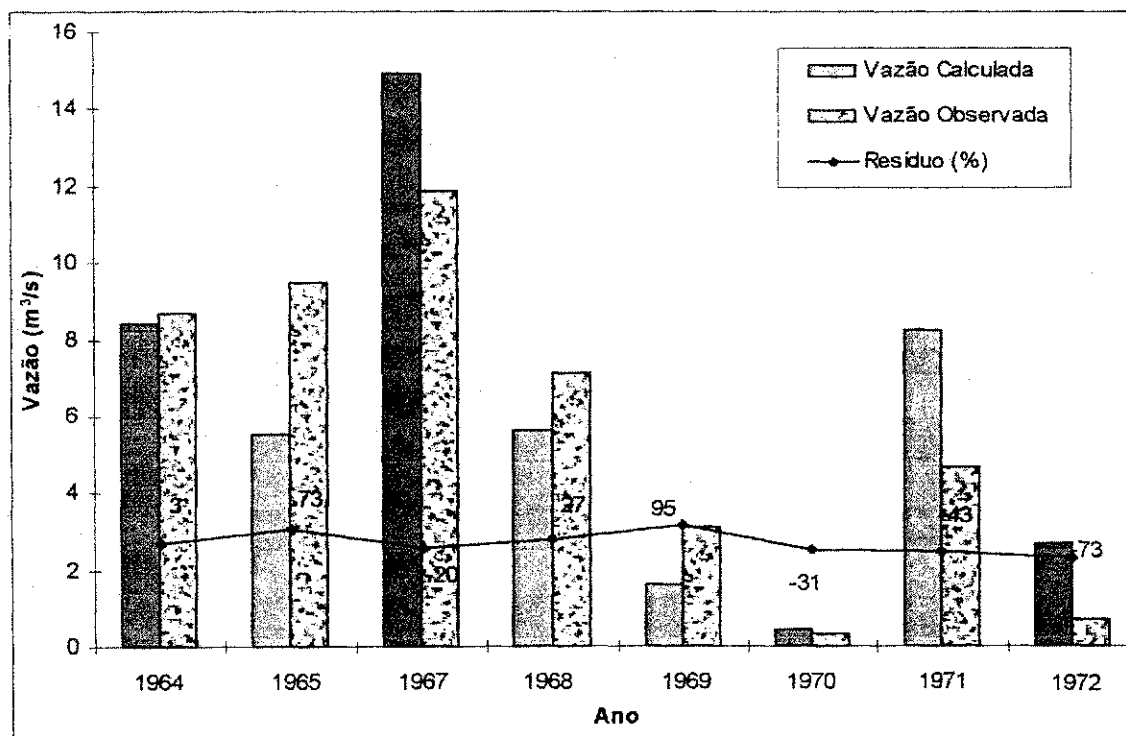


(b) Aparecida

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.

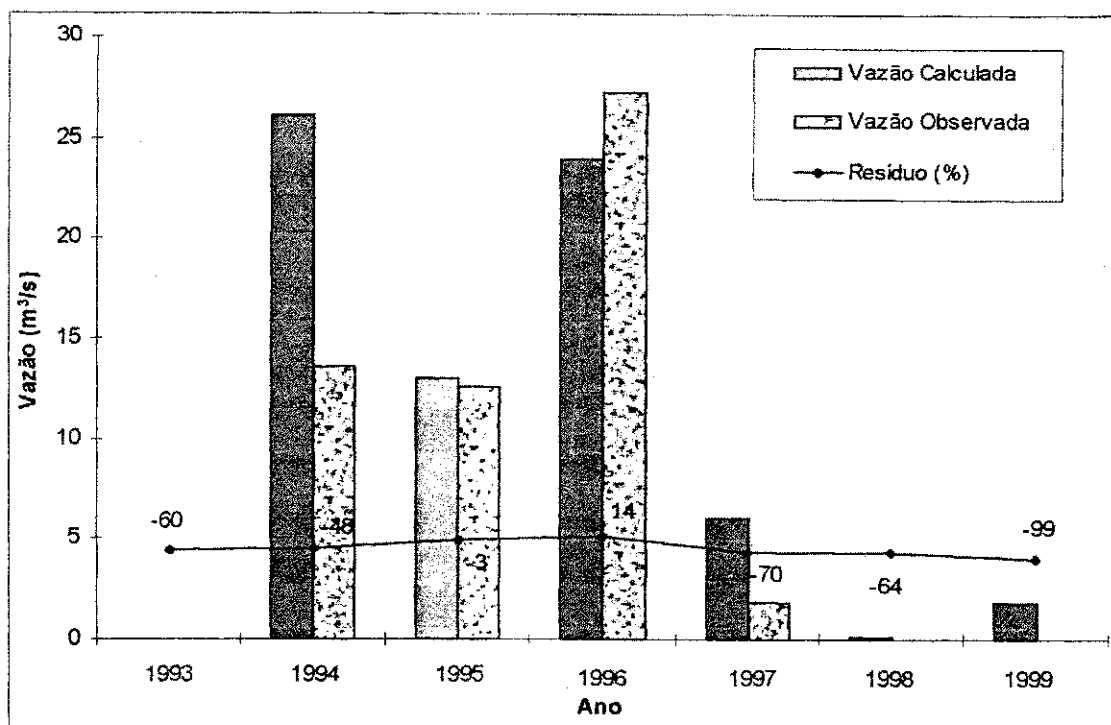


(c) Augusto Severo

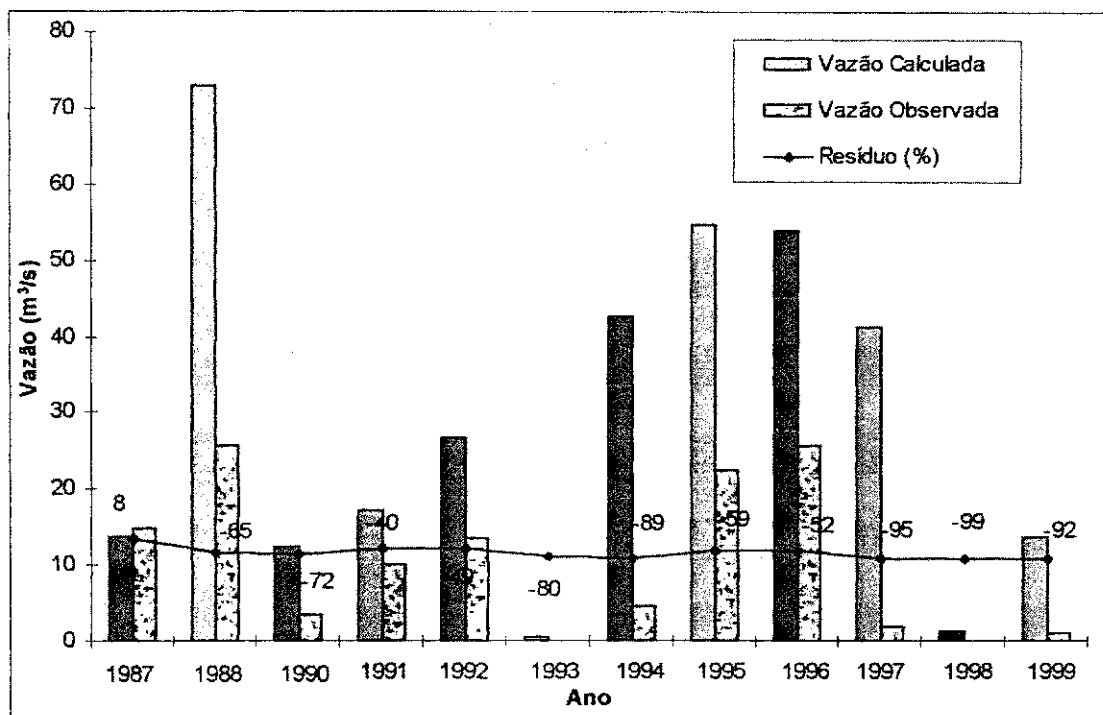


(d) Emas

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

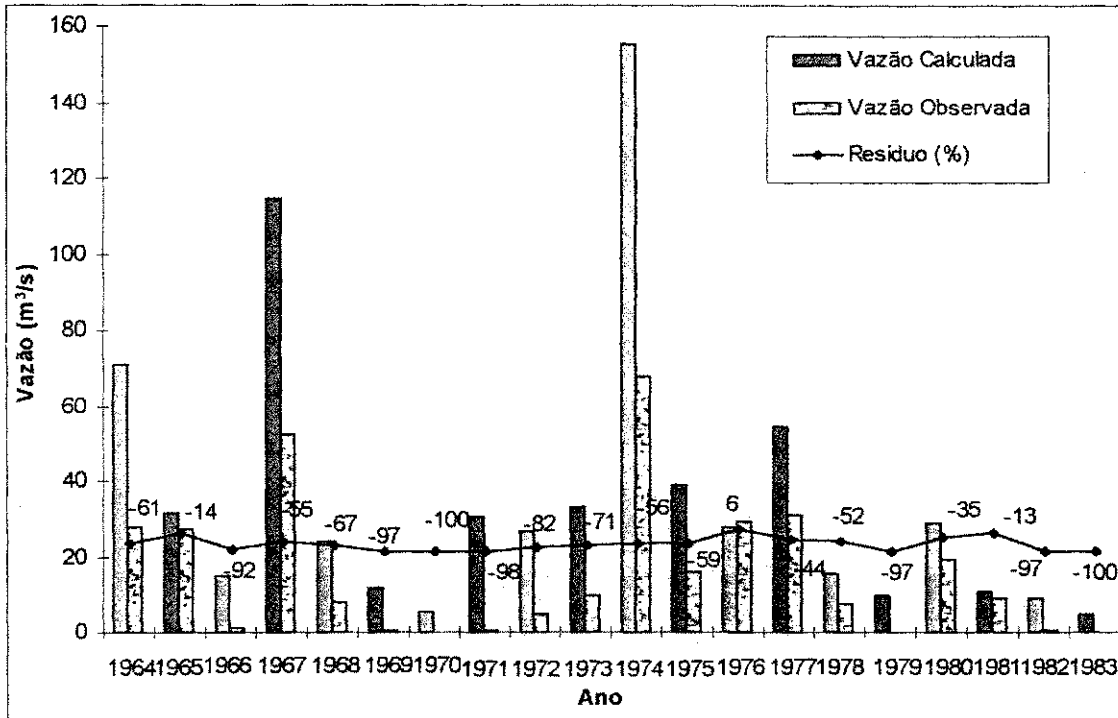


(e) Fazenda Angicos

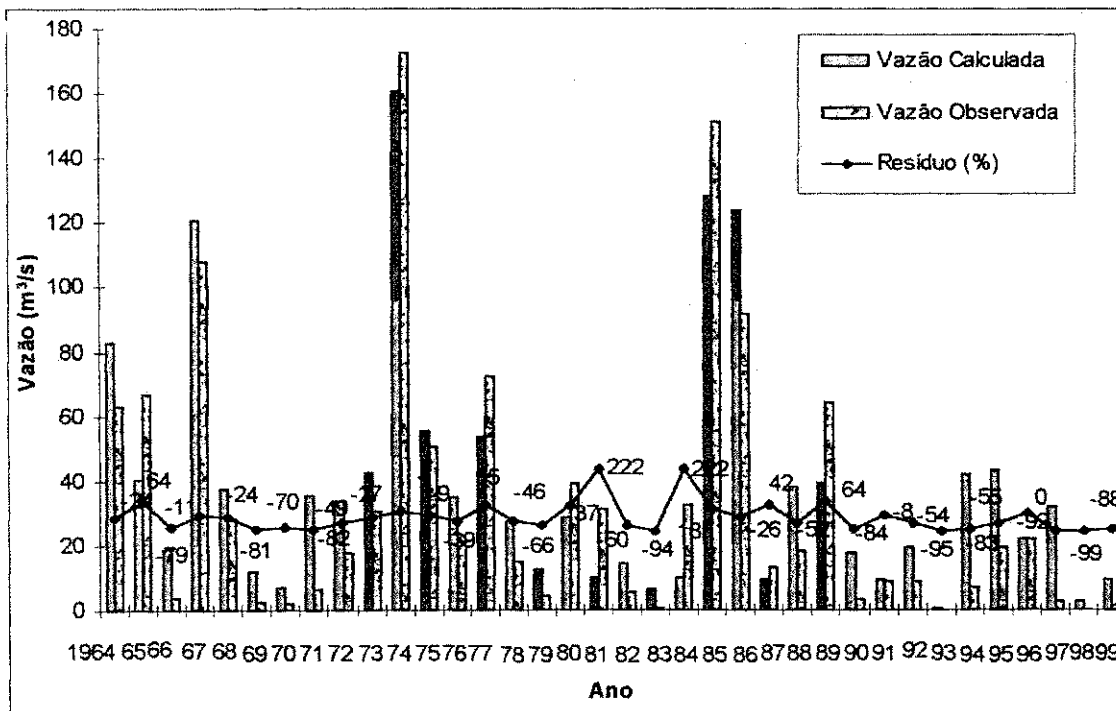


(f) Mossoró

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

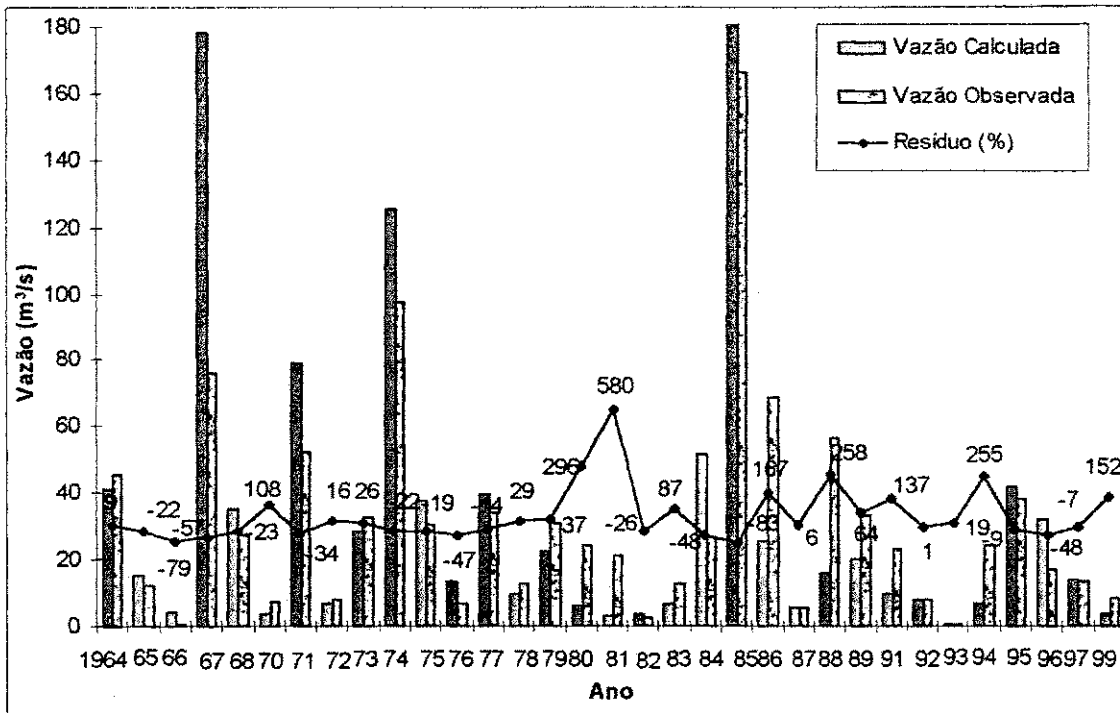


(g) Pau dos Ferros

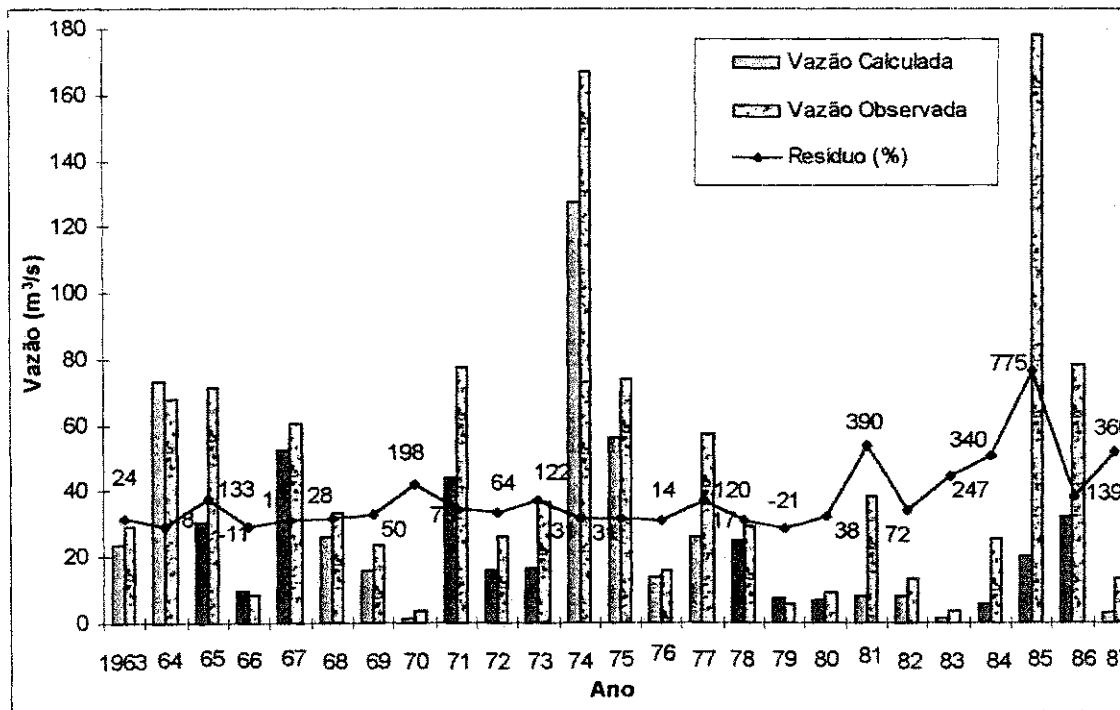


(h) Pedra de Abelhas

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sitio Volta; (l) Upanema; e (m) Serra Negra (continuação).

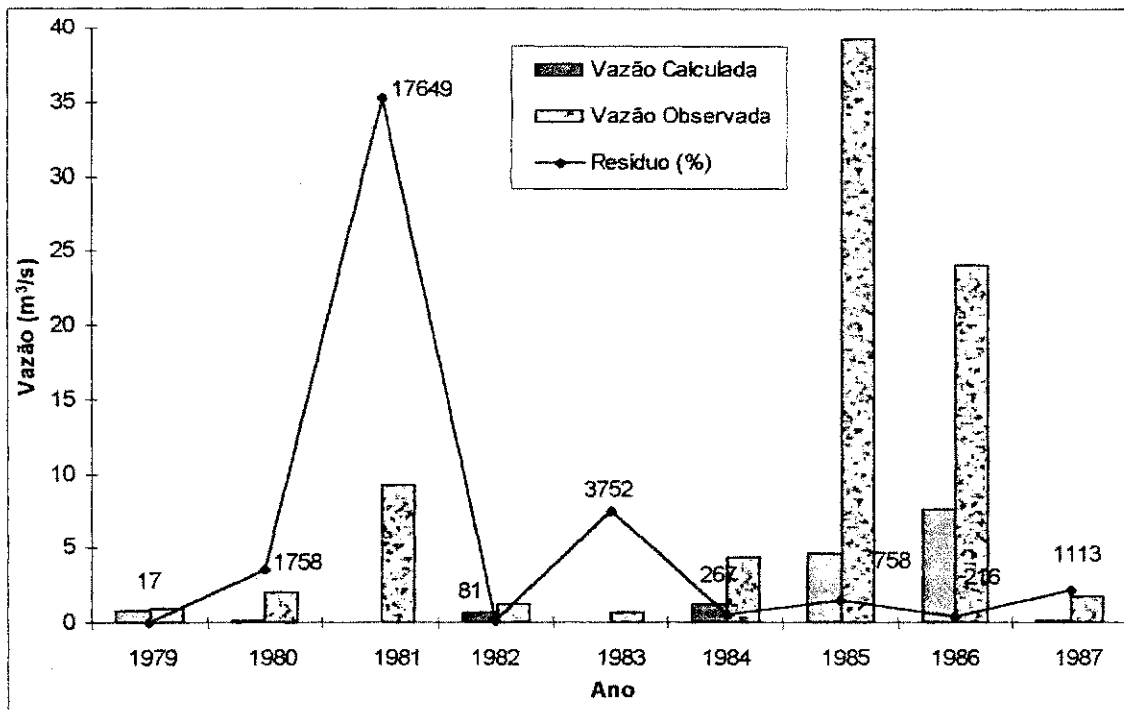


(i) Piancó

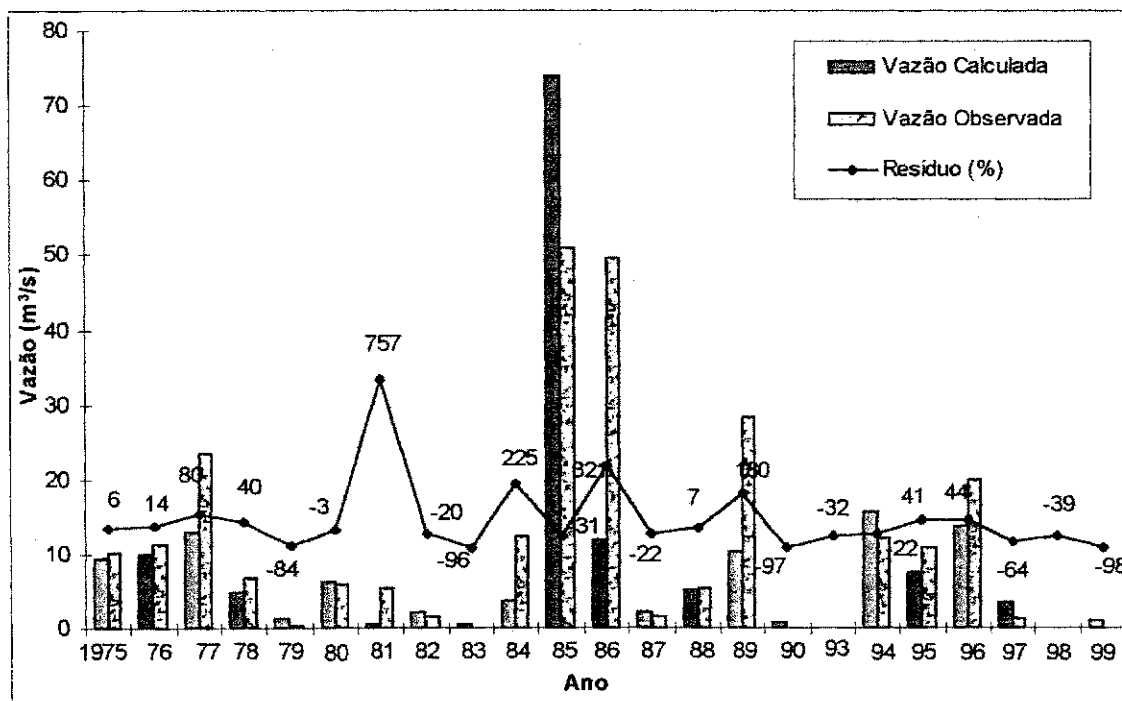


(j) São Fernando

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

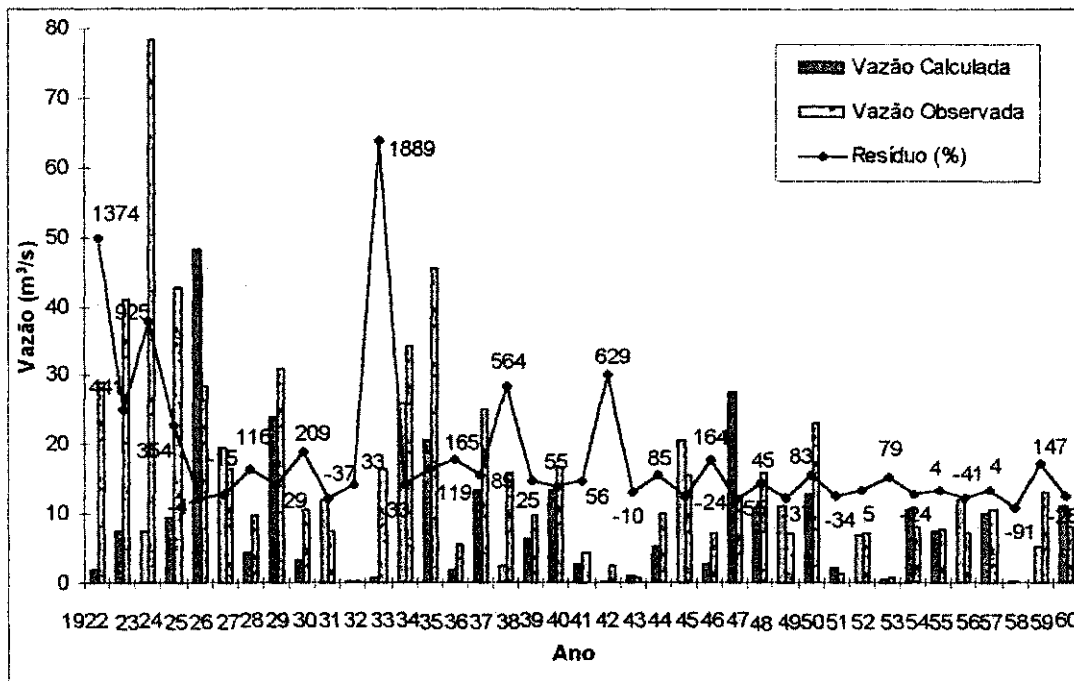


(k) Sítio Volta



(l) Upanema

Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(m) Serra Negra

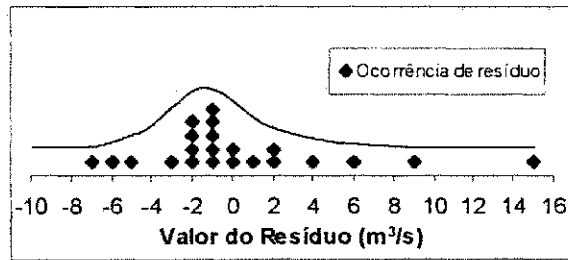
Figura 4.7 – Resultados da validação para as bacias hidrográficas dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

4.4. ANÁLISE DOS RESÍDUOS

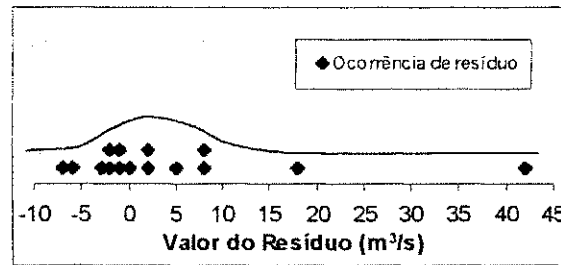
Nesta etapa do trabalho, os resíduos, a parte do valor observado que o ajuste não foi capaz de explicar, foram analisados graficamente seguindo os três critérios definidos anteriormente:

(1) Distribuição de frequência: Nessa análise, foi observado que a distribuição de frequência dos resíduos se é bem parecida com a distribuição normal de probabilidade com média zero, sendo que, para algumas bacias isto é visto bem claramente, enquanto outras mostram certas distorções que não permitem ver a forma da distribuição centrada e definida, como é visto nos gráficos da Figura 4.8 para as bacias de Antenor Navarro, Augusto Severo, Mossoró, Pau dos Ferros, Pedra de Abelhas, São Fernando e Sítio Volta. Onde as mesmas mostram um

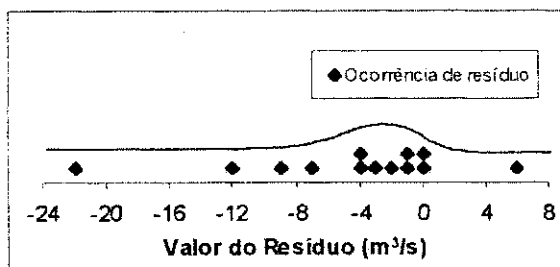
deslocamento na posição central na maioria para o lado negativo, o que para uma distribuição normal não pode acontecer, mostrando assim, que deveria ser feita alguma mudança no modelo de forma a melhorar esta subestimação para estas bacias.



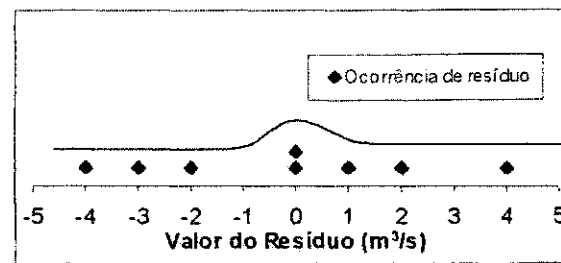
(a) Antenor Navarro



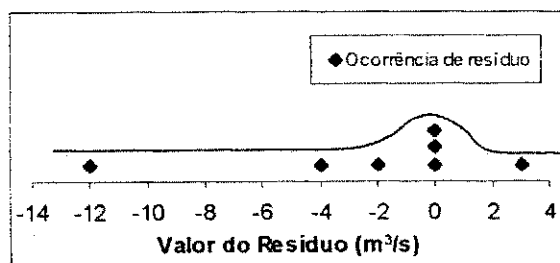
(b) Aparecida



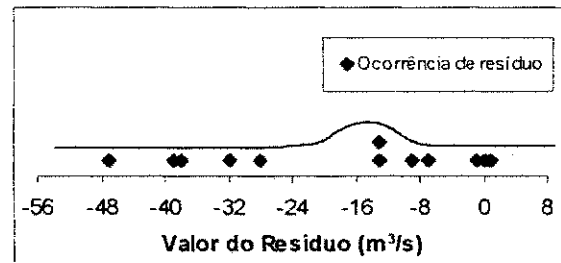
(c) Augusto Severo



(d) Emas

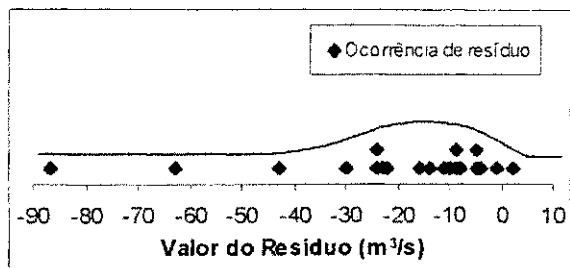


(e) Fazenda Angicos

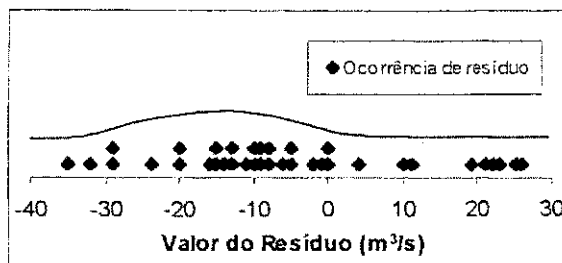


(f) Mossoró

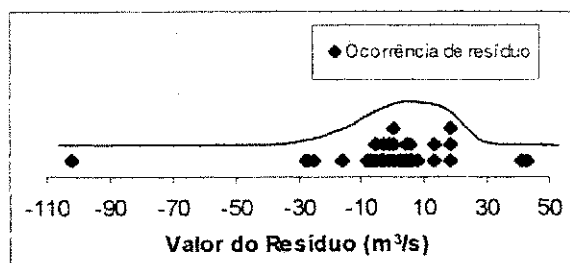
Figura 4.8 – Resultados da distribuição de frequência dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.



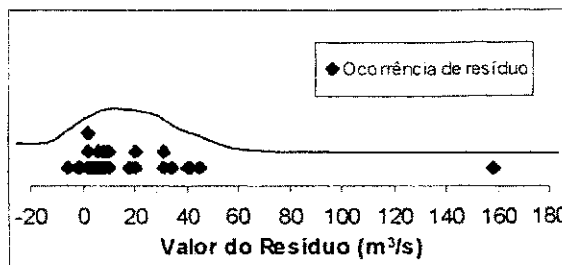
(g) Pau dos Ferros



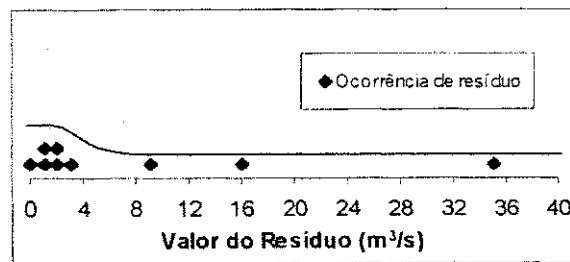
(h) Pedra de Abelhas



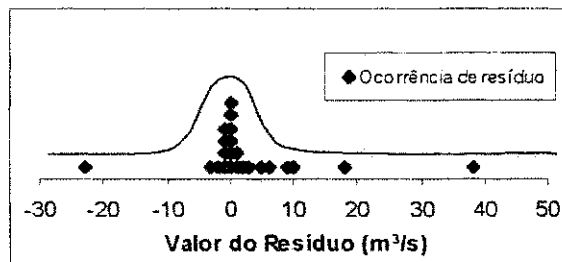
(i) Piancó



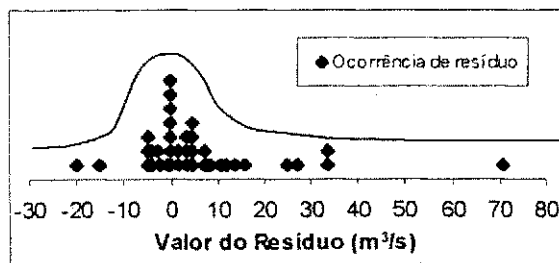
(j) São Fernando



(k) Sítio Volta



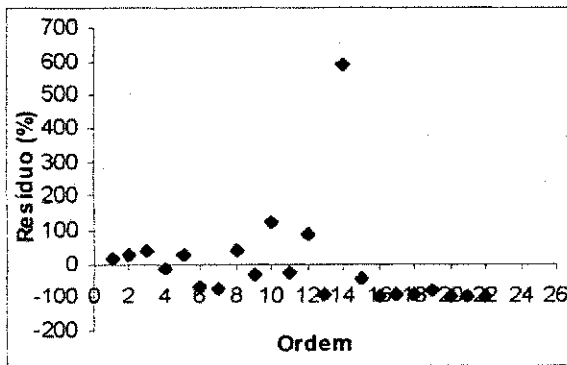
(l) Upanema



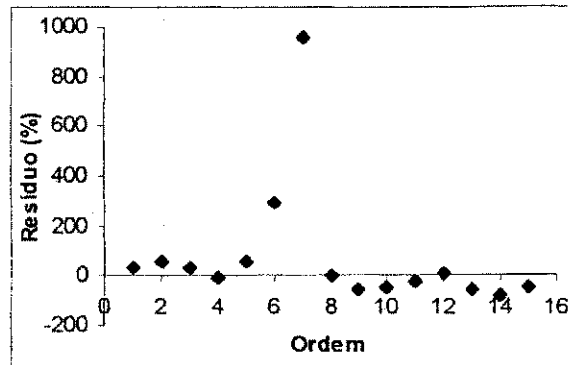
(m) Serra Negra

Figura 4.8 – Resultados da distribuição de frequência dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

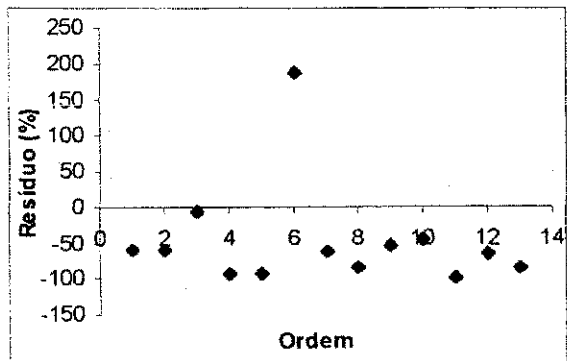
(2) Distribuição cronológica: nessa análise, foi observado que a distribuição dos resíduos seguindo uma determinada ordem cronológica, não sofreu mudanças bruscas com o tempo, ou seja, os resíduos mostram ser aleatórios, pois há muita variação dos valores observados nesta fase. Apenas pequenas mudanças, não muito significativas, foram detectadas nas bacias hidrográficas dos postos fluviométricos de Fazenda Angicos (Figura 4.9.e), São Fernando (Figura 4.9.j) e Serra Negra (Figura 4.9.m), como mostra a linha pontilhada nestes.



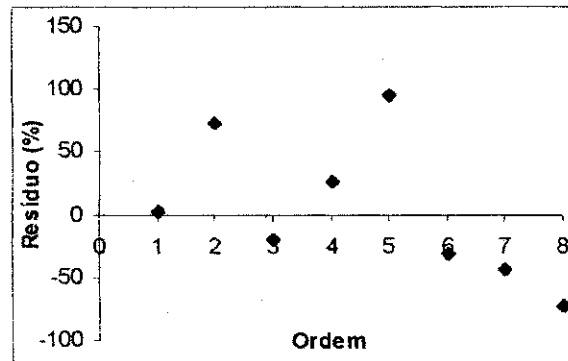
(a) Antenor Navarro



(b) Aparecida

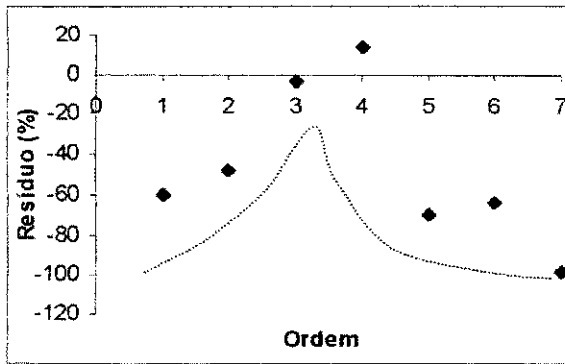


(c) Augusto Severo

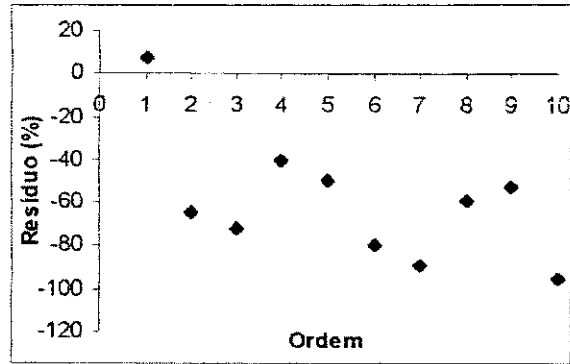


(d) Emas

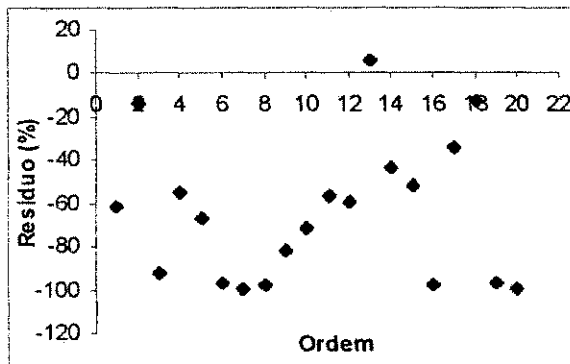
Figura 4.9 – Resultados da distribuição cronológica dos resíduos na fase de validação cruzada para as bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.



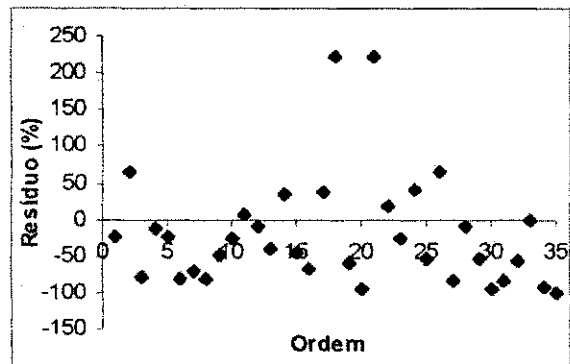
(e) Fazenda Angicos



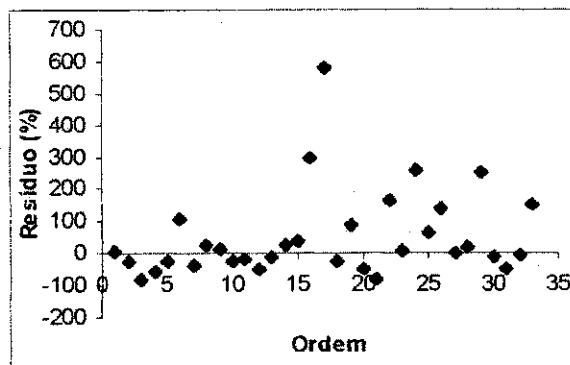
(f) Mossoró



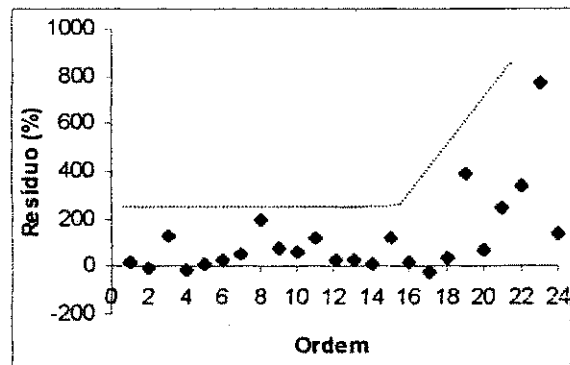
(g) Pau dos Ferros



(h) Pedra de Abelhas

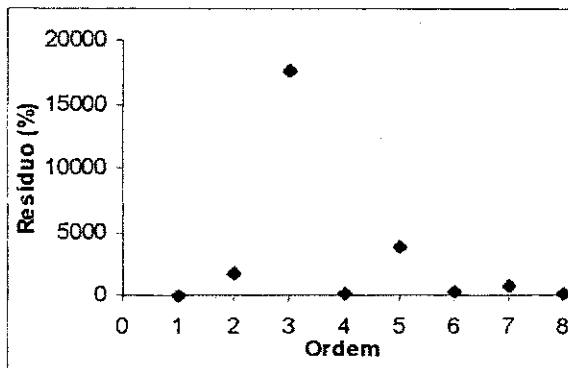


(i) Piancó

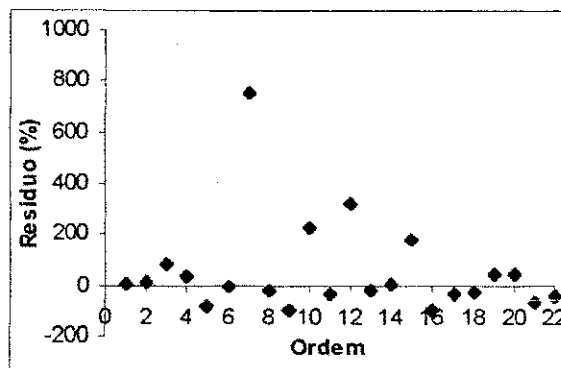


(j) São Fernando

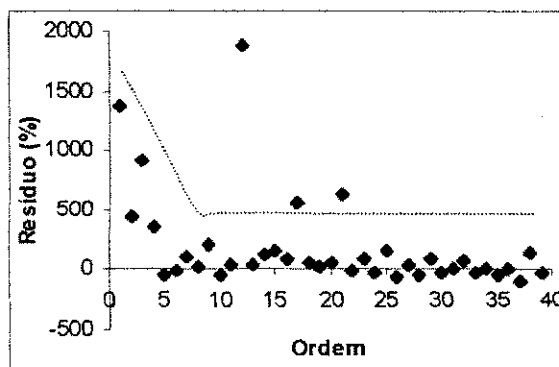
Figura 4.9 – Resultados da distribuição cronológica dos resíduos na fase de validação cruzada para as bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(k) Sítio Volta



(l) Upanema

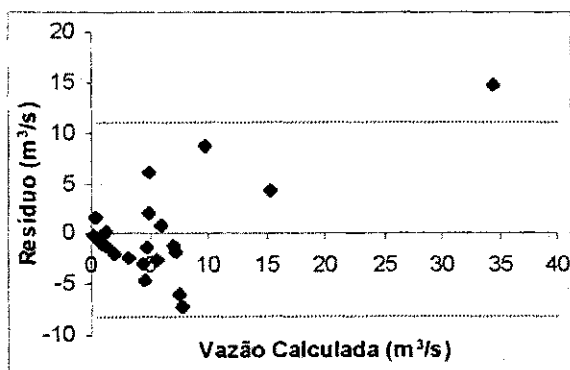


(m) Serra Negra

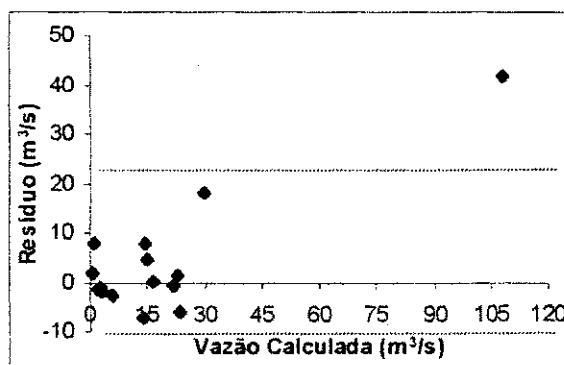
Figura 4.9 – Resultados da distribuição cronológica dos resíduos na fase de validação cruzada para as bacias dos postos fluviométricos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

(3) Distribuição com o valor calculado: ao analisar os gráficos dos resíduos em relação ao valor da vazão calculada, pode-se ver que há uma distribuição seguindo uma faixa linear, representada pelas retas pontilhadas neste, possível de ser analisado por não haver nenhuma correlação entre eles, em quase todas as bacias, com exceção das bacias hidrográficas dos postos fluviométricos de Augusto Severo (Figura 4.10.e), Mossoró (Figura 4.10.f) e Pau dos Ferros

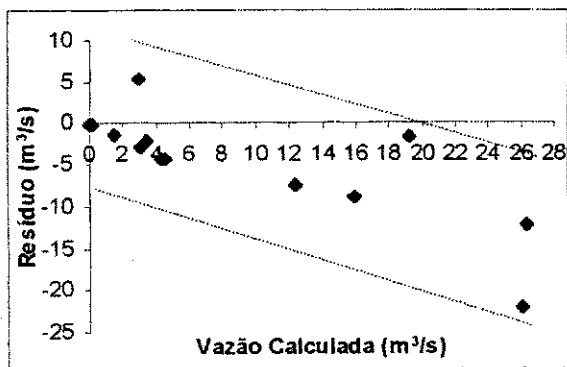
(Figura 4.10.g) que apresentaram uma tendência que deveria ser investigada. A presença de tendência sugere, em geral, a necessidade de incluir outras variáveis exploratórias na regressão acarretando mudanças no modelo. Mas, ao observar os demais resultados na validação para estes casos, não foi visto nenhuma distorção que corrobore esta necessidade. Todos os gráficos desta análise encontram-se na Figura 4.10 a seguir.



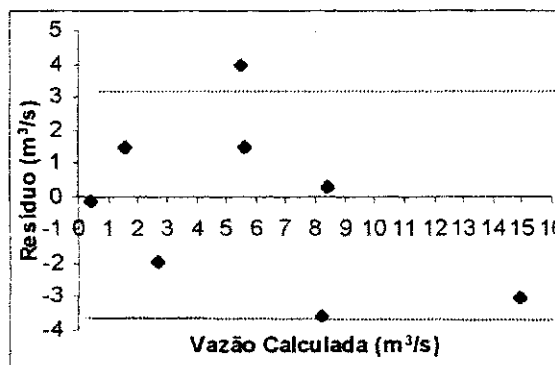
(a) Antenor Navarro



(b) Aparecida

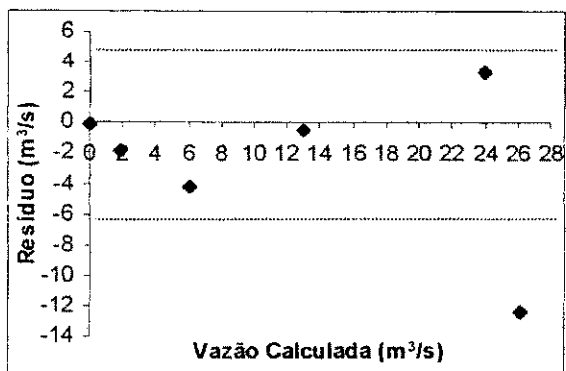


(c) Augusto Severo

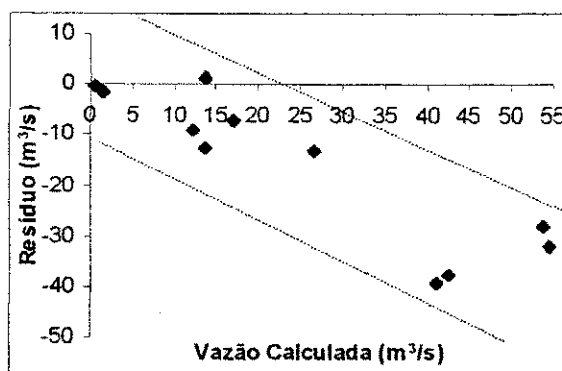


(d) Emas

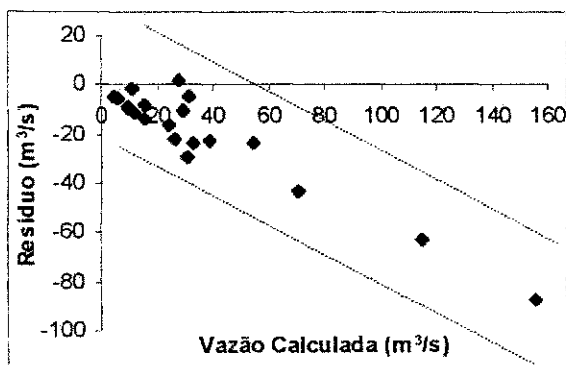
Figura 4.10 – Resultados da distribuição com valor calculado dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra.



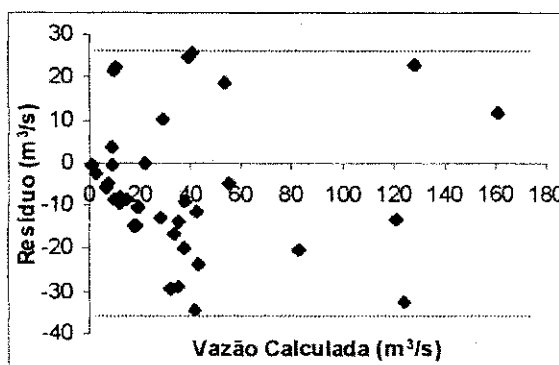
(e) Fazenda Angicos



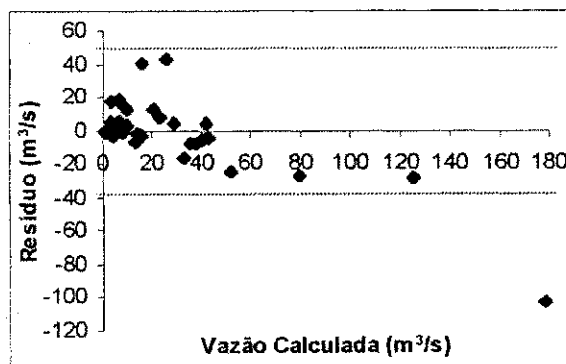
(f) Mossoró



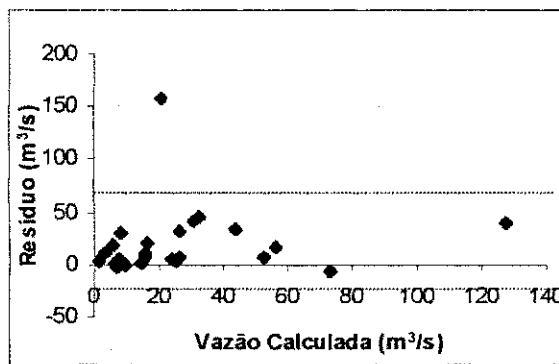
(g) Pau dos Ferros



(h) Pedra de Abelhas

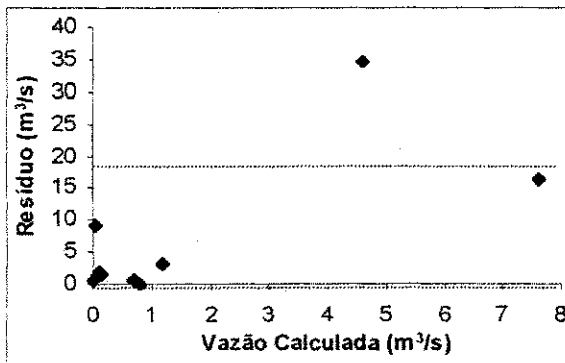


(i) Piancó

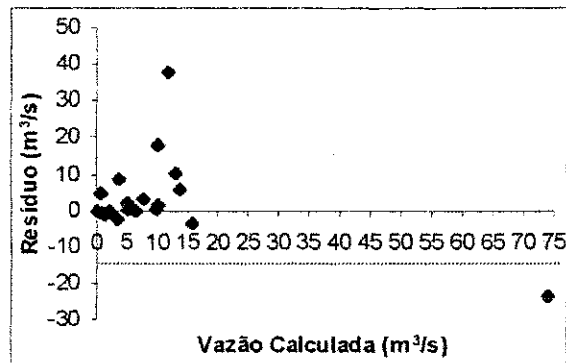


(j) São Fernando

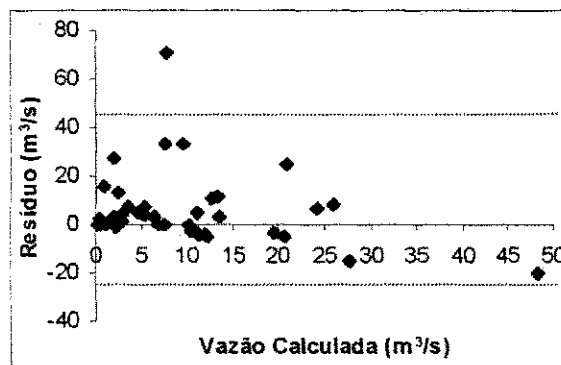
Figura 4.10 – Resultados da distribuição com valor calculado dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).



(k) Sítio Volta



(l) Upanema



(m) Serra Negra

Figura 4.10 – Resultados da distribuição com valor calculado dos resíduos na fase de validação cruzada para as bacias dos postos de: (a) Antenor Navarro; (b) Aparecida; (c) Augusto Severo; (d) Emas; (e) Fazenda Angicos; (f) Mossoró; (g) Pau dos Ferros; (h) Pedra de Abelhas; (i) Piancó; (j) São Fernando; (k) Sítio Volta; (l) Upanema; e (m) Serra Negra (continuação).

Os testes (1), (2) e (3), aplicados neste estudo, servem tanto para a validação do modelo, como mostra a condição de aplicação do modelo proposto, pois permitem uma detecção visual qualitativa de situações a serem investigadas. Estes testes mostram, portanto, os problemas que se pode esperar ao aplicar este modelo a outras bacias.

Após todas estas análises e embora tendo visto que nada impede a utilização da equação (4.2), apresentada anteriormente, em bacias sem dados e com características hidrológicas semelhante à região estudada, uma observação foi feita de forma a oferecer

uma contribuição mais adequada ao modelo proposto: a remoção dos dados da bacia hidrográfica do posto fluviométrico de Serra Negra do modelo, pois apesar de tais dados não terem mostrado grande influência no ajuste do modelo, (ver Figura 4.11) os mesmos não podem ser usados neste estudo, como foi visto na validação cruzada e pôde ser confirmado na Figura 4.12, que mostra o resultado gráfico do uso dos dados da bacia de Serra Negra na equação que foi gerada com as séries de todas as outras bacias usadas neste estudo. Isso pode ser confirmado nas Tabelas 4.8, 4.9, 4.10, 4.11, 4.12.

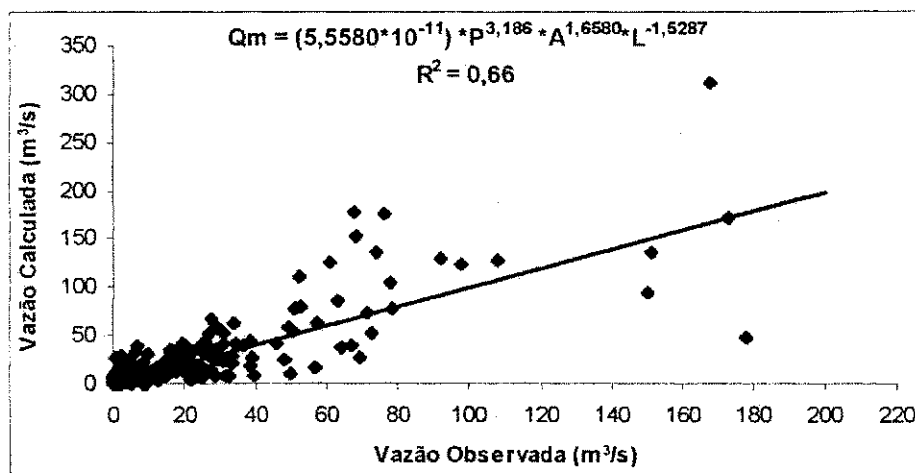


Figura 4.11 – Ajuste para todas as bacias estudadas, com exceção da bacia de Serra Negra.

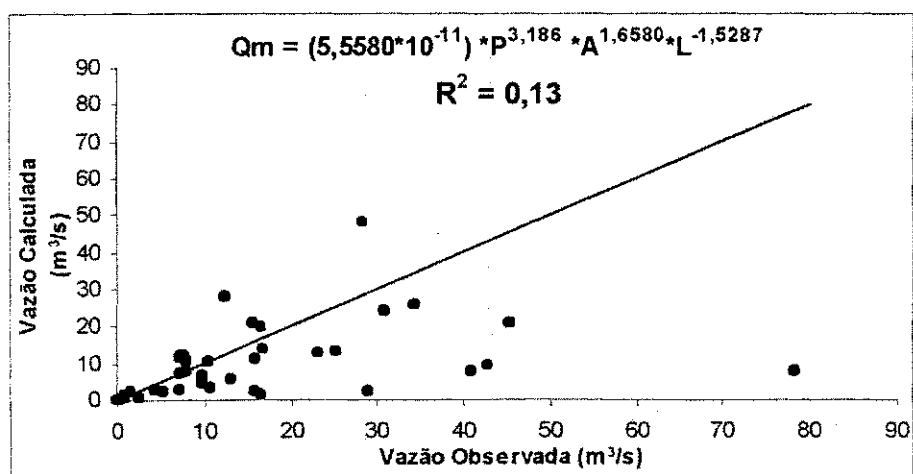


Figura 4.12 – Validação cruzada com a bacia de Serra Negra.

4.5 ANÁLISE DE REGRESSÃO SEM A BACIA DE SERRA NEGRA

Para a geração de uma nova equação, sem a bacia de Serra Negra, foram encontradas as seguintes tabelas de análises estatísticas:

Tabela 4.8 – Matriz de correlação e nível descritivo do teste $H_0: \text{Correl} = 0$, entre as variáveis do modelo.

	<i>lnQ</i>	<i>lnP</i>	<i>lnA</i>	<i>lnDD</i>	<i>lnDEC</i>	<i>LnL</i>
<i>lnQ</i>	1 0,000					
<i>lnP</i>	0,6733 0,000	1 0,000				
<i>lnA</i>	0,3728 0,000	-0,0142 0,833	1 0,000			
<i>lnDD</i>	0,1627 0,015	0,3195 0,000	-0,0875 0,193	1 0,000		
<i>lnDEC</i>	-0,0202 0,764	0,1404 0,036	-0,3063 0,000	0,0211 0,754	1 0,000	
<i>lnL</i>	0,2646 0,000	0,0172 0,798	0,8344 0,000	0,065 0,923	-0,4949 0,000	1 0,000

Tabela 4.9 – Estatística descritiva das variáveis para a construção do modelo.

<i>Variável</i>	<i>Média</i>	<i>Variância</i>	<i>Desvio Padrão</i>
<i>ln Q</i>	1,66542601	5,51464745	2,34832865
<i>ln P</i>	6,21520179	0,26985210	0,51947290
<i>ln A</i>	8,05255605	0,70006686	0,83669998
<i>ln DD</i>	0,09502242	0,03166475	0,17794592
<i>ln DEC</i>	1,37708520	0,22354507	0,47280553
<i>ln L</i>	4,70130045	0,18130055	0,42579403

Tabela 4.10 - Análise de variância.

<i>Fonte de variação</i>	<i>gl</i>	<i>Soma de quadrados</i>	<i>Quadrado médio</i>	<i>Valor F</i>	<i>PR > F</i>
Modelo	5	756,54767300	151,30953460	70,2029	0,0000
Resíduo	217	467,70406153	2,15531826		
Total	222	1224,25173453			

Tabela 4.11 - Análise das estimativas para o modelo.

<i>Parâmetro</i>	<i>Estimativa</i>	<i>Desvio Padrão</i>	<i>Valor T </i>	<i>PR > T </i>
Coef. linear	-23,61316986	1,79826303	13,1311	0,000
P	3,15855803	0,20336141	15,5317	0,000
A	1,65796975	0,22294478	7,4367	0,000
DD	-0,07286280	0,59205420	0,1231	0,902
DEC	-0,36993552	0,25048815	1,4769	0,142
L	-1,52872971	0,47979026	3,1862	0,002

Tabela 4.12 – Estatísticas para a análise do modelo.

<i>Parâmetro</i>	<i>Valor</i>
Média	1,665426
Raiz Quad. QMres.	1,4681002
Coef. Variação	88,1516328
Coef. Determinação	0,6179674
Coef. Determinação Ajust	0,6091648

Portanto o modelo de previsão da vazão média diária sazonal, obtida com as 12 bacias da região semi-árida, para ser usado em bacias onde não se têm dados hidrométricos observados, a ser sugerida por este estudo seria:

$$Q = 5,5580 \times 10^{-11} P^{3,1586} A^{1,6580} L^{-1,5287} \quad (4.3)$$

5. CONCLUSÕES E RECOMENDAÇÕES

5.1 CONCLUSÕES

O desenvolvimento da metodologia para a obtenção do modelo proposto neste estudo, bem como sua aplicação na região das bacias dos rios Piranhas-Açu-Apodi permite tirar as seguintes conclusões:

- ✓ Há possibilidade de se prever a vazão média diária sazonal em bacias hidrográficas sem dados hidrométricos, sendo esta mais uma contribuição para a tentativa de se amenizar, em especial, problemas causados por eventos extremos, como secas ou cheias, consequência da variação temporal e espacial da precipitação.
- ✓ Os resultados obtidos com a aplicação do modelo desenvolvido neste estudo se mostraram satisfatórios, apresentando boas correlações com os dados observados na fase de validação do modelo.
- ✓ Verificou-se, também, a importância de se utilizar a análise gráfica como critério de análise, e não apenas a verificação dos coeficientes de correlação. A análise gráfica dos desvios entre os valores observados e calculados pela equação mostra uma capacidade de representação da relação entre as variáveis compatíveis com os valores dos índices estatísticos de ajuste e, nesta etapa, a equação recomendada não apresentou nenhum ponto singular que comprometa sua utilização. Entretanto, foi verificado que a bacia de Serra Negra contribuiu negativamente para a geração da equação proposta. Logo uma outra equação foi sugerida com a retirada desta bacia. Isso prova a importância da boa definição das séries em estudos hidrológicos.
- ✓ Uma forma adequada para se obter um melhor resultado seria a obtenção de dados com uma série homogênea, onde pudesse ser adquirido um período único. Mas, é notória a necessidade de medições dos processos espaciais (chuva, vazão) em pontos da bacia (postos pluviométricos) ou seções de um rio (postos fluviométricos) de forma a tornar mais

confiável tais estudos. Pois como citou Tucci (2002): *“Se os dados não possuem qualidade ou não foram identificados e sanados os seus erros, a própria regionalização será tendenciosa e com resultados inadequados”*.

- ✓ As limitações do modelo são atribuídas devido o trabalho ter sido feito com um pequeno número de postos (13), com séries curtas e épocas distintas.
- ✓ Conclui-se, portanto, que o método de regionalização para prever vazões a partir de previsões sazonais de precipitação, da maneira que foi testada neste estudo, fazendo uso das características físicas da bacia, pode e deve ser utilizado como técnica de prever vazões em locais onde não existam dados hidrométricos observados, mas que disponham de informações pluviométricas e cartográficas.
- ✓ Esta metodologia foi desenvolvida e testada através de uma aplicação em uma determinada região do semi-árido nordestino, logo poderá ser empregada em outras regiões com o mesmo objetivo e assim, a equação correspondente poderá ser determinada e usada para previsões de vazões médias diárias sazonais.

5.2 RECOMENDAÇÕES

Baseado nos estudos e nos conhecimentos aqui adquiridos faz-se então algumas recomendações a trabalhos posteriores:

- Verificar se a utilização de outros dados homogeneizados e preenchidos as falhas acarretariam em melhorias na obtenção dos resultados;
- Fazer uso da equação sugerida para outra aplicação prática do modelo na realidade, verificando o erro da previsibilidade dos modelos de previsão de precipitação, o erro da regressão quando da obtenção da equação, e então definir o erro total a que se submeteria se fosse usado tal modelo nesta aplicação.
- Fazer uma união do modelo com a ferramenta de geoprocessamento num ambiente interativo de SIG, de forma a tornar mais fácil a sua utilização.

REFERÊNCIAS BIBLIOGRÁFICAS

- BARROS, M. T. L. (2003). Sistema de alerta a inundaç o da regi o leste do Estado de S o Paulo. (Professor da Escola Polit cnica da Universidade de S o Paulo (EPUSP) e diretor do Centro Tecnol gico de Hidr ulica do DAEE). Dispon vel on-line em www.geocities.com/capitolHill/Senate.htm, 01 Abril 2003.
- BARROSO, L. C., BARROSO, M. M. A., FILHO, F. F. C., CARVALHO, M. L. B., MAIA, M. L. (1987). *C culo num rico com aplica es*. UFMG. 2^a edi o. Harbra Ltda.
- BERTONI, J. C.; TUCCI, C. E. M. (2000). *Precipita o*. Hidrologia: Ci ncia e aplica o. UFRGS: ABRH. v.4, c.5. 2^a edi o. Rio Grande do Sul.
- BERTONI, J.C.; TUCCI, C.E.M.; CLARKE, R.T. (1992). *Rainfall-based real-time flood forecasting*. Journal of Hydrology n. 131 p 313-339. Apud BRUN, G. W.; TUCCI, C. E. M. – Previs o em tempo real do volume afluente ao reservat rio de Ernestina. RBRH – Revista Brasileira de Recursos H dricos. Vol. 6, n.2, p 73-79, Abr/Jun 2001.
- BRAGA, C. C. (1992). *Classifica o de regi es pluviometricamente homog neas atrav s da an lise multivariada*. In: VII Congresso Brasileiro de Meteorologia. Vol. 1. S o Paulo, 1992.
- BRUN, G. W.; TUCCI, C. E. M. (2001). *Previs o em tempo real do volume afluente ao reservat rio de Ernestina*. RBRH – Revista Brasileira de Recursos H dricos. Vol. 6, n.2, p 73-79, Abr/Jun 2001.
- BUSSAB, W. O.; MORETTIN, P. A. (1987). *Estatistica b sica – m todos quantitativos*. 4^a edi o. Atual Editora Ltda.
- CAUVIN, C.; RIMBERT, S. (1976). *La lecture num rique d s cartes th matiques*. Ed. Universitaires. Fribourg.

- COSTA, M. T. (2003). Atenção para a previsão do tempo - *Novo sistema vai trazer salto de qualidade nas informações sobre chuvas, temperaturas, enchentes, geadas...* Disponível on-line em http://www.unicamp.br/unicamp/unicamp_hoje/ju/jul2000/pagina4-Ju153.html, 01 Abril 2003.
- CRESPO, A. A. (1994). *Estatística fácil* – 11^a edição. Editora Saraiva.
- CRUZ, J. (1987). *Estudo de previsão de cheias em tempo real com modelos empíricos e semi-conceituais*. Dissertação de mestrado COPPE UFRJ.
- DINIZ, L. S.; CLARKE, R. T. (2001). *Regionalização de parâmetros de modelo chuva-vazão usando redes neurais*. In: XIV Simpósio Brasileiro de Recursos Hídricos – V Simpósio de Hidráulica e Recursos Hídricos dos países de língua oficial portuguesa: gestão de Recursos Hídricos – O desafio da prática. Aracaju / SE, 2001.
- DRAPER, N. R.; SMITH, H. (1981). *Applied Regression Analysis*, 2a edição, John Wiley and Sons, New York, USA.
- DUAN, Q.; SOROOSHIAN, S.; GUPTA, V.K. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *WATER RESOURCES RESEARCH*, v.28, n.4, p.1015-1031. Apud DINIZ, L. S.; CLARKE, R. T. (2001). *Regionalização de parâmetros de modelo chuva-vazão usando redes neurais*. In: XIV Simpósio Brasileiro de Recursos Hídricos – V Simpósio de Hidráulica e Recursos Hídricos dos países de língua oficial portuguesa: gestão de Recursos Hídricos – O desafio da prática. Aracaju / SE, 2001.
- DUNN, O. J.; CLARK, V. A. (1974). *Applied Statistics: Analysis of variance and regression*. By John Wiley & Sons, Inc. United States of America. (Wiley Series in probability and mathematical statistics).
- ELESBON, A. A. A.; FIOROT, P.; CESCNETO, E.; COSER, M.; MENDONÇA, A. S. F. (2002). *Regionalização de vazões máximas, médias e mínimas para o estado do Espírito Santo visando futuras aplicações de SIG em sua automatização* – ABRH. In: VI Simpósio de Recursos Hídricos do Nordeste. Maceió - AL, 2002.
- ELETROBRÁS (1985). Centrais Elétricas Brasileiras S.A. *Metodologia para Regionalização de Vazões*. Rio de Janeiro.
- FAVORETO, R. S.; ROHN, M. C.; MINE, M. R. M. (2001). *A técnica de rede neural artificial aplicada na previsão de vazão*. In: XIV Simpósio Brasileiro de Recursos Hídricos – V Simpósio de Hidráulica e Recursos Hídricos dos países de língua oficial portuguesa: gestão de Recursos Hídricos – O desafio da prática. Aracaju / SE, 2001.