

»
»

**MODELOS ESTOCÁSTICOS NO GERENCIAMENTO
DOS RECURSOS HÍDRICOS**

Jerson Kelman

AGRADECIMENTOS

Este capítulo foi escrito de setembro de 1986 a maio de 1987, período em que estive trabalhando no exterior. Sem a cooperação e a revisão de Jorge Machado Damazio, através da intensa troca de correspondência, o projeto não teria sido realizado. O processo de revisão contou com a colaboração de Paulo Holanda Salles.

O texto foi parcialmente redigido num trem suburbano, no trajeto entre minha casa e o local de trabalho. Os garranchos resultantes foram habilmente decodificados e datilografados por Ana Maria Costa Danielli.

Celeste, Diana e Rafael foram tolerantes em vários fins de semana em que a minha presença foi um tanto ausente.

1. PLANEJAMENTO SOB INCERTEZA E PROCESSOS ESTOCÁSTICOS

Gerenciar um sistema de recursos hídricos significa atuar no sentido de assegurar uma distribuição temporal e espacial da água que melhor se coadune com os interesses de uma comunidade. Nesta genérica definição, a prudente expressão "atuar no sentido de assegurar" é utilizada em lugar de simplesmente "assegurar" porque não se pode estar certo de que o objetivo do gerenciamento será atingido, tantas são as incertezas associadas a qualquer ação que se possa tomar. Por exemplo, um sistema de abastecimento de água pode ser insuficiente num determinado ano, seja devido à ocorrência de uma seca mais severa do que a admitida no planejamento, seja devido a um imprevisto aumento na demanda.

O objetivo deste capítulo é descrever algumas técnicas para quantificar e atribuir probabilidades a eventos futuros relacionados com a disponibilidade de recursos hídricos. Também pretende discutir de que maneira essas probabilidades são afetadas por obras de engenharia, tais como reservatórios de regularização. A demanda futura pelos recursos hídricos, outra fonte de incerteza, não será tratada aqui, porque esse assunto envolve áreas de conhecimento (economia, política, sociologia,...) que fogem ao escopo de um texto sobre hidrologia.

Os modelos e algoritmos matemáticos empregados em hidrologia estocástica são erigidos sobre conceitos de probabilidade e estatística. É, portanto, essencial que o leitor tenha conhecimento dessas disciplinas básicas, pelo menos a nível introdutório.

Muitas questões envolvendo incertezas no gerenciamento de recursos hídricos exigem apenas a aplicação das técnicas que constam do arsenal de um engenheiro versado em probabilidade e estatística. Nesses casos, a solução do problema está frequentemente associada à estimativa do valor esperado de alguma variável aleatória ou à estimativa da probabilidade de algum evento. Esquemáticamente, pode-se dizer que o problema é resolvido em quatro etapas:

- a) coletam-se ou organizam-se as observações da variável aleatória x numa amostra aleatória: $x = (x(1), x(2), \dots, x(n))$;
- b) seleciona-se uma expressão paramétrica para a função densidade de probabilidade de X , $f_x(x; \theta)$ (ou para a função acumulada de probabilidades $F_x(x; \theta)$), onde θ é o parâmetro (pode ser um vetor). A seleção é usualmente feita a partir do exame das propriedades da amostra x , embora em alguns casos seja possível escolher a expressão paramétrica com base em conceitos teóricos;
- c) obtém-se uma estimativa de θ , através da combinação de x com $f_x(x; \theta)$. Testa-se a hipótese de que a amostra x tenha sido sorteadada de $f_x(x; \theta)$;
- d) utiliza-se $f_x(x; \theta)$ para a obtenção da resposta desejada.

Nada se altera nessas quatro etapas se os elementos da amostra aleatória forem embaralhados entre si. Isto é, a ordem em que as observações $x(1), x(2), \dots, x(n)$ foram sorteadas é irrelevante. Existem outros casos, no entanto, em que a ordem é relevante. Por exemplo, se X for o deflúvio anual de uma bacia com solos profundos e grande capacidade de retenção de água, um valor x muito acima da média dificilmente será seguido por um x muito abaixo da média, porque uma certa quantidade de água é retida, de um ano para outro, na forma de armazenamento subterrâneo, que por sua vez alimenta o escoamento básico. Em casos como esse, as observações $x(1), x(2), \dots, x(n)$ não formam uma amostra aleatória e sim uma série temporal. Não se pode falar de uma variável aleatória X e sim de um vetor de variáveis $X(1), X(2), \dots, X(n)$, em geral dependentes entre si. Esta seqüência indexada de variáveis aleatórias é também chamada

de processo estocástico. Diz-se que a série temporal é uma realização do processo estocástico, assim como uma observação da variável aleatória é um possível resultado do "sorteio" da variável aleatória. Quando as variáveis aleatórias $X(1), X(2), \dots, X(n)$ são independentes entre si e igualmente distribuídas, o processo estocástico colapsa numa única variável aleatória e a correspondente série temporal transforma-se numa amostra aleatória. Em resumo, processo aleatório (independente) é um caso particular de processo estocástico (dependente).

Um processo estocástico é dito estacionário quando a distribuição conjunta de $X(t), X(t+1), \dots, X(t+m)$ é idêntica à distribuição conjunta de $X(t+j), X(t+j+1), \dots, X(t+j+m)$. Isto é, num processo estocástico estacionário, as propriedades estatísticas não se alteram com o passar do tempo, embora as correspondentes séries temporais possam ser diferentes entre si.

Devido à periodicidade anual dos processos climáticos, as vazões discretizadas em intervalos inferiores ao ano (dia, mês) corresponderão necessariamente a processos não estacionários. Entretanto, mesmo quando o intervalo de discretização é anual, o correspondente processo estocástico não será rigorosamente estacionário, na medida em que:

- as características fisiográficas das bacias hidrográficas estão em constante evolução devido à erosão e à sedimentação;
- a distribuição da energia solar sofre variações periódicas, o que ocasiona mudanças climáticas; e
- as características fisiográficas das bacias hidrográficas sofrem alterações devido à intervenção humana, por exemplo, construção de reservatórios ou desmatamento.

As variações correspondentes aos tópicos a) e b) processam-se numa escala de tempo tão grande que as tornam irrelevantes para o horizonte de planejamento considerado no equacionamento de problemas de engenharia. Muito mais importante são as variações correspondentes a c). Há circunstâncias em que a intervenção na bacia hidrográfica é tão relevante que a hipótese de estacionariedade não pode ser aceita. Nesses casos, o problema é extremamente complexo, não sendo possível o emprego das técnicas apresentadas neste capítulo. O leitor fará bem consultando o capítulo 3 desse volume. Em outras ocasiões, a intervenção é moderada o suficiente para que a hipótese de estacionariedade possa ser aceita dentro de uma ótica em que se pretenda encontrar uma solução aproximada para um problema de engenharia.

Se a intervenção na bacia hidrográfica for moderada, os problemas que envolvem processos estocásticos serão, em geral, resolvidos em quatro etapas, num procedimento similar ao adotado no caso de uma única variável aleatória:

- coleta-se ou organiza-se uma série temporal do processo estocástico $X(t)$;
- seleciona-se um modelo matemático para o processo estocástico;
- estimam-se os parâmetros do modelo e testam-se o ajuste e a adequação do modelo; e
- utiliza-se o modelo ajustado para a obtenção da resposta desejada.

A etapa a) é objeto do item 4.2 deste capítulo, e não será tratada aqui. Os demais itens tratam das etapas b) e c) e esporadicamente da etapa d).

Para dar uma melhor idéia de como existem situações, no gerenciamento de recursos hídricos, em que o uso do modelo aleatório é suficiente e de como, em outras situações, o modelo estocástico é indispensável, considere o seguinte exemplo:

Exemplo 1

Suponha que uma companhia de abastecimento de água tenha a possibilidade de captar água de um rio situado numa bacia hidrográfica com solos pouco profundos e pouco permeáveis. A vazão do rio é quase exclusivamente composta por escoamento superficial e, portanto, quase tão variável quanto a ocorrência de precipitação. Para assegurar um suprimento de água de acordo com as flutuações do consumo, e não de acordo com o ritmo em que ocorrem as precipitações, a companhia decidiu construir um reservatório de regularização. O problema consiste em dimensionar o volume útil. Para simplificar a questão, admita que, no primeiro mês do ano, chove tão abundantemente que o reservatório enche completamente, qualquer que seja o volume útil que se possa escolher. Nos 11 meses subsequentes, a afluência ao local do reservatório não é tão abundante, sendo possível que ocorra um racionamento. Para simplificar ainda mais, ignore flutuações na afluência e na demanda durante esses 11 meses. Nessas condições, haverá racionamento se, e só se, o reservatório chegar vazio ao final do ano. Isto é:

$$P(R > 0) = P(X = 0) \quad (1.1)$$

Pela equação da continuidade, tem-se que

$$X = \max(0, \min(v, v + Y - d)) \quad (1.2)$$

$$R = \max(0, X - v - Y + d) \quad (1.3)$$

(usa-se letra maiúscula para simbolizar variável aleatória e minúscula para representar um valor de uma variável aleatória ou uma constante).

onde

- R – volume racionado ao longo dos 11 meses
- X – volume armazenado no reservatório, ao final do ano
- v – volume útil do reservatório
- Y – volume afluente durante os 11 meses
- d – demanda volumétrica durante os 11 meses

Se a companhia quiser evitar a possibilidade de racionamento, deverá construir um reservatório bem grande, tal que $v > d$. Entretanto, isso pode ser econômica ou fisicamente impossível. Supondo $v < d$, a probabilidade de racionamento diminui quando "v" aumenta. Por outro lado, o custo de construção das obras de barramento e de desapropriação das áreas inundadas cresce com "v". Ou seja, quanto maior for a segurança do sistema de abastecimento, maior será o custo. Problemas desse tipo são freqüentemente resolvidos com base no "pior ano" do registro histórico. Isto é, escolhe-se um volume útil para o reservatório que garanta abastecimento normal caso a mais severa seca observada no passado ocorra novamente, durante a vida útil do empreendimento. No caso específico,

$$v_1 = \max (0, d - \min_i (y(i))) \quad (1.4)$$

onde $y(1), y(2), \dots, y(i), \dots, y(n)$ são observações do volume afluente e "n" é o número de anos do registro histórico.

Pelo menos duas críticas podem ser feitas a esse método de cálculo:

- a) a única certeza que se pode ter sobre o futuro é que o passado não se repetirá, isto é, secas piores do que a mais crítica do passado podem ocorrer, sem que se tenha informação sobre a probabilidade desse evento;
- b) v_1 é uma função não decrescente de n, isto é, quanto mais longo for o registro histórico, maior tenderá a ser o volume de regularização v_1 ; conseqüentemente, se dois locais hidrológicamente equivalentes estiverem competindo para sediar a construção de um reservatório, o local com registro mais curto tenderá a levar vantagem, o que é um contra-senso.

Um critério razoável para a escolha de v consiste em minimizar a soma do valor esperado do prejuízo de um racionamento num ano qualquer com o custo anual referente à construção do reservatório.

$$\int_0^{d-v_2} h(d-v_2-y) f_Y(y) dy + g(v_2) \quad (1.5)$$

onde

- v_2 - volume de armazenamento ótimo
- $g(\cdot)$ - custo anual referente à construção do reservatório
- $h(\cdot)$ - prejuízo associado com racionamento
- $f_Y(\cdot)$ - função densidade de probabilidade de Y

Um outro critério, igualmente razoável, é definir v de tal maneira que a probabilidade de que haja racionamento seja igual a um pequeno valor α , previamente fixado (por exemplo $\alpha = 0,05$). Dentro dessa ótica, a decisão ótima é tal que

$$P(R > 0) = P(X = 0) = \int_0^{d-v_3} f_Y(y) dy = \alpha \quad (1.6)$$

Tanto a solução v_2 como a v_3 apresentam vantagens sobre a solução v_1 . Seleciona-se uma ou outra dependendo de uma opção político-econômica. Em ambos os casos, o primeiro passo a ser dado pela companhia para tomar uma decisão, num contexto de incerteza, consiste em identificar quais os possíveis eventos subseqüentes à

decisão e quantificar suas respectivas probabilidades. Neste exemplo, essa ação implica na determinação das funções "custo de construção", $g(\cdot)$, e "custo de déficit", $h(\cdot)$, isto se a equação (1.5) for adotada. A modelagem da incerteza, nesse caso, resume-se à determinação da densidade de probabilidade $f_Y(y)$, a partir de uma amostra $y(1), y(2), \dots, y(n)$.

A solução do problema descrito no exemplo anterior foi feita de forma relativamente simples, tanto através da equação (1.5) quanto da equação (1.6), porque foram feitas as suposições de que:

- a) o reservatório sempre enche no mês chuvoso;
- b) a afluência Y num ano qualquer é independente de afluências em anos anteriores, devido a um escoamento básico insignificante e à ausência de "memória" nos processos meteorológicos, de um ano para outro.

Se a suposição a) não for verdadeira, o estado de armazenamento do reservatório ao final do primeiro mês de cada ano não é mais constante. Conseqüentemente, definindo-se $X(t)$ como o armazenamento do reservatório num determinado dia do ano t, tem-se que $X(t), X(t+1), X(t+2), \dots$ formam uma sucessão de variáveis dependentes entre si, isto é, formam um processo estocástico. Analogamente, se a suposição b) não for verdadeira, as afluências anuais $Y(t), Y(t+1), Y(t+2)$ formam um outro processo estocástico. Isto é, conhecida a afluência do ano passado, $y(t-1)$, possivelmente a distribuição condicionada de $Y(t)$, dado $y(t-1)$, será diferente da distribuição de $Y(t+1)$, dado $y(t-1)$, que será diferente da distribuição de $Y(t+2)$, dado $y(t-1)$, etc. Conseqüentemente, também $X(t)$ será um processo estocástico.

Nesse caso mais geral, em que $X(t)$ é um processo estocástico, seria ainda possível dimensionar o reservatório com base no período mais crítico observado no passado, através da aplicação da curva de massa ou diagrama de Rippl. Entretanto, as mesmas críticas apresentadas na seqüência da equação (1.4) poderiam ser repetidas com relação a essa solução. As equações (1.5) e (1.6) não poderiam ser utilizadas, mas os métodos a serem desenvolvidos a seguir permitirão a solução do problema de dimensionamento de reservatórios, e de diversos outros, sob a ótica de quantificação da probabilidade de falha e/ou do valor esperado dos benefícios de um empreendimento hídrico.

2. TEORIA ESTOCÁSTICA DE RESERVATÓRIOS

2.1 Análise da operação e dimensionamento

Um processo estocástico é definido como markoviano quando a distribuição de probabilidades dos valores futuros do processo depende apenas do último valor da série temporal. Isto é, num processo markoviano, as observações anteriores ao instante t são irrelevantes para a modelagem do processo após t, desde que se conheça $x(t)$. Mais formalmente,

$$F(X(t+\tau) \mid x(t), x(u), x(v), \dots) = F(X(t+\tau) \mid x(t)) \quad (2.1a)$$

$t > u > v, \quad \tau > 0$

As séries temporais em hidrologia são, em geral, medidas a intervalos discretos de tempo, por exemplo, uma vez por dia. Mesmo quando a medição é contínua, como no caso do linígrafo, o registro costuma ser armazenado sob a forma de uma série temporal em que os valores são registrados a intervalos regulares de tempo. É, portanto, natural que se restrinja o estudo a processos estocásticos discretos no tempo, em cujo caso a equação 2.1 pode ser reescrita:

$$F(X(t+\tau) \mid x(t), x(t-1), x(t-2), \dots, x(1)) = F(X(t+\tau) \mid x(t)) \quad (2.1b)$$

Quando os valores assumidos por $X(t)$ são também discretos ($X(t) = x_0, x_1, x_2, \dots$), o processo markoviano é chamado de "Cadeia de Markov". Frequentemente, no planejamento de recursos hídricos, processos estocásticos que são, na realidade, contínuos, no tempo e nos valores, acabam por serem aproximados por uma Cadeia de Markov, devido à facilidade de associação com algoritmos de otimização, conforme se verá adiante.

Suponha que $X(t)$ possa assumir os valores (chamados "estados") 0, 1, 2, ... i, ... K, K+1. Seja $a(i,j) = P(X(t+1) = i \mid X(t) = j)$ e $a_{t+1}(i) = P(X(t+1) = i)$. Naturalmente,

$$\sum_i a_t(i) = 1, \forall t \quad (2.2)$$

$$\sum_i a(i,j) = 1, \forall j \quad (2.3)$$

$a(i,j)$ é chamada de probabilidade de transição do estado j para o estado i , e a matriz formada pelos elementos $a(i,j)$ é chamada de matriz de transição. Seja A tal matriz, isto é, $\{A\}_{ij} = a(i,j)$. Do teorema da probabilidade total, tem-se que $P(X(t+1) = i) = P(X(t+1) = i \mid X(t) = 0) \cdot P(X(t) = 0) + P(X(t+1) = i \mid X(t) = 1) \cdot P(X(t) = 1) + \dots$. Isto é,

$$a_{t+1}(i) = \sum_j a(i,j) a_t(j) \quad (2.4)$$

$$a_{t+1}(i) = \sum_j a(i,j) a_t(j)$$

Definindo-se o vetor "distribuição de probabilidades" $a_t = (a_t(0), a_t(1), \dots, a_t(k), a_t(k+1))^T$, a equação (2.4) pode ser generalizada para qualquer t :

$$a_{t+1} = A a_t \quad (2.5)$$

onde a_t, a_{t+1} são vetores-coluna com $K+2$ elementos e A é uma matriz com dimensões $(K+2) \times (K+2)$. A distribuição de probabilidade de $X(t+2)$, isto é, a_{t+2} , pode ser facilmente calculada aplicando-se a equação (2.5).

$$a_{t+2} = A a_{t+1}$$

ou

$$a_{t+2} = A(A a_t) = A^2 a_t$$

Em geral,

$$a_{t+m} = A^m a_t \quad (2.6)$$

Pode-se mostrar que $\{A^m\}_{ij}$ é a probabilidade de transição do estado de j para o estado i , em m passos. À medida que m aumenta, menos importante é a informação com relação ao estado de partida j . No limite, o estado de partida é irrelevante:

$$\lim_{m \rightarrow \infty} \{A^m\}_{ij} = \lim_{m \rightarrow \infty} a_{t+m}(i) = a(i) \quad (2.7)$$

O vetor-coluna $a = (a(0), a(1), \dots, a(k), a(k+1))^T$ é a distribuição marginal de probabilidades de $X(t)$. Da equação (2.7) tem-se que

$$\lim_{m \rightarrow \infty} A^m = [a, a, \dots, a] \quad (2.8)$$

Diz-se que a Cadeia de Markov está equilibrada no instante t , quando $a_{t+1} = a_t = a$. Isto é, o sistema está equilibrado se $P(X(t) = i) = P(X(t+\tau) = i)$,

$\forall \tau = 1, 2, \dots, \forall i = 0, 1, \dots, K$. Da equação (2.5) tem-se que

$$\underline{a} = \underline{Aa} \tag{2.9}$$

A equação (2.9) representa um conjunto de $(K+2)$ equações e $(K+2)$ incógnitas, respectivamente a (0) , a (1) , ..., a $(K+1)$, mas não pode ser resolvido diretamente porque qualquer uma das equações é uma combinação linear das demais. Isto é, o posto da matriz A é $K+1$ e não $K+2$. A equação que falta para a resolução do sistema é

$$\sum a(i) = 1 \tag{2.10}$$

A solução é encontrada substituindo-se qualquer uma das equações em (2.9) pela equação (2.10). Alternativamente, pode-se fazer uso da equação (2.8), elevando a matriz de transição A a uma potência m suficientemente grande para que todas as colunas de A^m sejam praticamente idênticas a \underline{a} .

Diversos processos estocásticos em hidrologia podem ser modelados por uma Cadeia de Markov. Por exemplo, Loucks *et al* (1981, cap. 3) mostra uma aplicação dessa abordagem quando $X(t)$ é o deflúvio anual de um rio. No entanto, o uso mais comum é para o caso em que $X(t)$ é função do estoque de água armazenado num reservatório no início do ano t . Essa aplicação da Cadeia de Markov é às vezes chamada de "Análise de Moran", em homenagem ao pioneiro nesse assunto (Moran, 1954).

Ao longo desta sessão serão examinados diversos problemas possíveis de serem resolvidos pela aplicação de conceitos da Cadeia de Markov. Será sempre assumido que v é o volume útil e a unidade de volume é 1% de v . Essa unidade de volume servirá para medir o estoque de água no reservatório, a afluência e a demanda. Isto é, todas as grandezas volumétricas serão padronizadas, dividindo-as pelo volume útil e multiplicando por 100. Dessa maneira, diz-se que o reservatório armazena 60 unidades, se o estoque de água corresponde a 60% do volume útil. Seja $X(t) = j$, $j = 1, 2, \dots, K$, se o armazenamento útil ao início do ano t situa-se na faixa entre $(j - 1)\Delta$ e $j\Delta$, unidades de volume onde $\Delta = 100/K$, $X(t) = 0$ se o reservatório estiver apenas com o volume morto, e $X(t) = K+1$ se o reservatório estiver cheio. Assume-se que:

- a) o volume anual afluente ao reservatório, $Y(t)$, é independente de $Y(\tau)$, $\forall \tau \neq t$;
- b) a distribuição de probabilidades de Y_t é conhecida ou foi previamente estimada: $P(Y_t \leq y) = F_Y(y)$;
- c) qualquer estoque de água situado entre $(j-1)\Delta$ e $j\Delta$ unidades de volume será "arredondado" para $\Delta(j-1+0,5) = \Delta(j-0,5)$. Porém nenhum arredondamento será feito se o reservatório estiver com volume morto ou então cheio;
- d) o consumo anual é d ;
- e) a evaporação é ignorada, ou incluída no consumo;
- f) a afluência $Y(t)$ e o consumo d são uniformemente distribuídos ao longo do ano.

334



Exemplo 2

Suponha $d = 40$, Y normalmente distribuído, $Y \sim N(E(Y) = 50; \text{var}(Y) = 30^2)$, $K = 5$, e o estado inicial de armazenamento igual a 73 unidades. Nesse caso, $\Delta = 100/5 = 20$ e 73 situa-se no intervalo $(60, 80)$. Portanto, $j = 80/20 = 4$ e o volume armazenado são arredondados para $20(4-0,5) = 70$. Haverá uma transição de $x(t) = 4$ para $x(t+1) = 5$ (por exemplo, se $X(t+1) = x(t) + Y(t) - d = 70 + Y(t) - 40$ estiver contido no intervalo $(80, 100)$). A probabilidade dessa transição é igual à probabilidade de que $Y(t)$ pertença ao intervalo $(50, 70)$, isto é, $a(5,4) = \Phi((70 - 50)/30) - \Phi((50 - 50)/30) = 0,25$, onde Φ é a distribuição acumulada de probabilidades normal padrão. Haverá uma transição de $x(t) = 4$ para $x(t+1) = 6$ (cheio), se $Y(t)$ for maior do que 70, isto é, $a(6,4) = 1 - \Phi((70 - 50)/30) = 0,25$. Analogamente, todas as probabilidades $a(i,j)$ poderiam ser calculadas, resultando na seguinte matriz de transição A :

		A						
		j						
i		0	1	2	3	4	5	6
0		0,37	0,25	0,09	0,02	0,00	0,00	0,00
1		0,26	0,25	0,16	0,07	0,02	0,00	0,00
2		0,21	0,25	0,25	0,16	0,07	0,02	0,01
3		0,11	0,16	0,25	0,25	0,16	0,07	0,04
4		0,04	0,07	0,16	0,25	0,25	0,16	0,11
5		0,01	0,02	0,07	0,16	0,25	0,25	0,21
6		0,00	0,00	0,02	0,09	0,25	0,50	0,63

335

Deve-se notar que, de acordo com a equação (2.3), a soma dos elementos de qualquer coluna de A é sempre igual à unidade. A matriz A revela, por exemplo, que a probabilidade de racionamento no ano t é aproximadamente igual 0,37 se o reservatório já começar o ano vazio (transição de vazio para vazio); cai para 0,25, se o estado inicial for $x(t) = 1$ (armazenamento entre 0% e 20% do volume útil); cai para 0,09, se $x(t) = 2$, e para 0,02, se $x(t) = 3$. Se o estoque de água no início do ano for superior a 60% do volume útil, a probabilidade de racionamento é nula. Por outro lado, basta que o estoque de água inicial seja superior a 20% do volume útil para que já haja alguma probabilidade de vertimento (terminar o ano cheio). Se o reservatório já começa o ano cheio, a probabilidade de vertimento é de 0,63. Estes resultados são apenas aproximados, devido às hipóteses simplificadoras adotadas e também ao efeito da discretização. Naturalmente, a precisão dos resultados melhora na medida em que K cresce.

O cálculo das probabilidades de transição, para as hipóteses adotadas no exemplo anterior, e para qualquer valor de K , pode ser generalizado de acordo com as seguintes equações:

3

$$a(i,j) = P(Y \in (\Delta(i-1) + d - \Delta(j-0,5); \Delta i + d - \Delta(j-0,5))) \\ = F_Y(\Delta(i-j+0,5) + d) - F_Y(\Delta(i-j-0,5) + d) \quad (2.11a) \\ i = 1, 2, \dots, K; \quad j = 1, 2, \dots, K$$

$$a(i, K+1) = P(Y \in (\Delta(i-1) + d - 100); (\Delta i + d - 100)) \\ = F_Y(\Delta i + d - 100) - F_Y(\Delta(i-1) + d - 100) \quad (2.11b) \\ i = 1, 2, \dots, K$$

$$a(K+1, j) = P(Y > 100 + d - \Delta(j-0,5)) \\ = 1 - F_Y(100 + d - \Delta(j-0,5)) \quad (2.11c) \\ j = 1, 2, \dots, K$$

$$a(i, 0) = P(Y \in (\Delta(i-1) + d); (\Delta i + d)) \\ = F_Y(\Delta i + d) - F_Y(\Delta(i-1) + d) \quad (2.11d) \\ i = 1, 2, \dots, K$$

$$a(0, j) = P(Y < d - \Delta(j-0,5)) \\ = F_Y(d - \Delta(j-0,5)) \quad (2.11e) \\ j = 1, 2, \dots, K$$

$$a(0, K+1) = P(Y < d - 100) \\ = F_Y(d - 100)$$

$$a(0, 0) = P(Y < d) \\ = F_Y(d) \quad (2.11g)$$

$$a(K+1, 0) = P(Y > d + 100) \\ = 1 - F_Y(d + 100) \quad (2.11h)$$

$$a(K+1, K+1) = P(Y > d) \\ = 1 - F_Y(d) \quad (2.11i)$$

Exemplo 3

Dada a matriz de transição A do exemplo 2, por simples multiplicação sucessiva é possível obter-se as matrizes A², A⁴, A⁸, A¹⁶, A³²...A. Alguns dos resultados são dados a seguir.

Observa-se que as colunas de A^m vão ficando cada vez mais parecidas entre si, à medida que m aumenta. Para m = 32, elas já são idênticas, pelo menos para a precisão adotada, de duas casas decimais. Isso significa que a probabilidade de o reservatório estar vazio, daqui a pelo menos 32 anos, é sempre de 0,03, não importa qual seja o estado de armazenamento atual. Analogamente, a probabilidade de que o reservatório esteja cheio, também daqui a 32 anos, é sempre de 0,40. Na primeira parte deste capítulo formulou-se uma possível abordagem para o dimensionamento de um reservatório, a qual consiste em achar o volume útil v tal que a probabilidade de racionamento num ano qualquer seja igual a um pequeno valor α, previamente fixado. Para o tamanho de reservatório adotado nos exemplos 2 e 3, essa probabilidade é de aproximadamente 0,03. Se o alvo α for 0,05, pode-se considerar a construção de um reservatório menor. O projetista, nesse

A²

i	j						
	0	1	2	3	4	5	6
0	0,22	0,18	0,10	0,05	0,02	0,00	0,00
1	0,20	0,18	0,12	0,07	0,03	0,01	0,01
2	0,22	0,21	0,18	0,12	0,07	0,04	0,03
3	0,17	0,18	0,19	0,17	0,13	0,08	0,07
4	0,11	0,13	0,17	0,19	0,18	0,16	0,14
5	0,05	0,07	0,12	0,17	0,21	0,22	0,22
6	0,03	0,05	0,12	0,23	0,36	0,49	0,53

A⁸

i	j						
	0	1	2	3	4	5	6
0	0,05	0,05	0,04	0,03	0,03	0,03	0,02
1	0,06	0,06	0,05	0,05	0,04	0,03	0,03
2	0,10	0,09	0,09	0,08	0,07	0,06	0,06
3	0,13	0,13	0,12	0,11	0,11	0,11	0,10
4	0,16	0,16	0,16	0,16	0,16	0,16	0,16
5	0,17	0,18	0,18	0,19	0,19	0,20	0,20
6	0,33	0,33	0,36	0,38	0,40	0,41	0,42

A³²

	0	1	2	3	4	5	6
0	0,03	0,03	0,03	0,03	0,03	0,03	0,03
1	0,04	0,04	0,04	0,04	0,04	0,04	0,04
2	0,07	0,07	0,07	0,07	0,07	0,07	0,07
3	0,11	0,11	0,11	0,11	0,11	0,11	0,11
4	0,16	0,16	0,16	0,16	0,16	0,16	0,16
5	0,19	0,19	0,19	0,19	0,19	0,19	0,19
6	0,40	0,40	0,40	0,40	0,40	0,40	0,40

caso, deverá tentar um novo valor para v , que servirá para definir uma nova matriz de transição A . Os valores "padronizados" de d , $E(Y)$ e $var(Y)$ irão aumentar, porque a unidade em que essas grandezas foram expressas corresponde a 1% do volume útil. A repetição do procedimento para se achar nova distribuição marginal de probabilidades para $X(t)$ resulta num novo valor para $a(0)$. Se $a(0)$ for significativamente diferente de α , nova tentativa será feita, diminuindo ou aumentando v se o novo $a(0)$ for menor ou maior do que α . A busca por tentativa e erro, para ser praticável, pressupõe a utilização de um computador.

No item 1 foi apresentado um critério alternativo para dimensionamento de reservatório, que consiste em minimizar a soma do valor esperado dos prejuízos de eventuais racionamentos, num ano qualquer, com a parcela anual da amortização do custo de construção. Como conceber uma matriz de transição que sirva a esse propósito? Uma possibilidade é definir hipotéticos estados de armazenamento negativos.

Seja $X(t) = j$, $j = 1 - d/\delta, 2 - d/\delta, \dots, 0$, se o armazenamento, no início do ano t , é nulo e no ano $t - 1$ o racionamento situa-se na faixa $-(j-1)\delta$ e $-j\delta$. Isto é, o racionamento é discretizado em intervalos δ e substitui-se o estado de armazenamento nulo por d/δ estados negativos. É preciso definir uma nova matriz de transição A , agora de dimensão $(K + 1 + d/\delta) \times (K + 1 + d/\delta)$, em vez de $(K + 2) \times (K + 2)$. Obviamente, na nova matriz, todas as probabilidades de transição $a(i, j)$, $i = 1, 2, \dots, K+1$, $j = 1, 2, \dots, K + 1$ são as mesmas que na antiga matriz (equações 2.11a,b,c). Além disso, $a(i, j)$, $i = 1, 2, 3, \dots, K + 1$ e $j = 1 - d/\delta, 2 - d/\delta, \dots, 0$ da nova matriz é igual a $a(i, 0)$ da antiga matriz - equação (2.11d, e). Resta unicamente definir $a(i, j)$, $i = 1 - d/\delta, 2 - d/\delta, \dots, 0$, $\forall j$, segundo as seguintes equações:

$$a(i, j) = P(Y \in (\delta(i-1) + d - \Delta(j-0,5)); \delta i + d - \Delta(j-0,5)) = F_Y(\delta i + d - \Delta(j-0,5)) - F_Y(\delta(i-1) + d - \Delta(j-0,5))$$

$$i = 2 - d/\delta, 3 - d/\delta, \dots, 0, j = 1, 2, \dots, K \quad (2.12a)$$

$$a(i, K+1) = P(Y \in (\delta(i-1) + d - 100; (\delta i + d - 100))) = F_Y(\delta i + d - 100) - F_Y(\delta(i-1) + d - 100)$$

$$i = 2 - d/\delta, 3 - d/\delta, \dots, 0 \quad (2.12b)$$

$$a(i, j) = P(Y \in (\delta(i-1) + d); \delta i + d)) = F_Y(\delta i + d) - F_Y(\delta(i-1) + d)$$

$$i = 2 - d/\delta, 3 - d/\delta, \dots, 0$$

$$j = 1 - d/\delta, 2 - d/\delta, \dots, 0 \quad (2.12d)$$

$$a(1-d/\delta, j) = P(Y < \delta - \Delta(j - 0,5)) = F_Y(\delta - \Delta(j - 0,5))$$

$$j = 1, 2, \dots, K \quad (2.12e)$$

$$a(1-d/\delta, K+1) = P(Y < \delta - 100) \quad (2.12f)$$

$$a(1-d/\delta, j) = P(Y < \delta - d + d) = F_Y(\delta)$$

$$j = 1 - d/\delta, 2 - d/\delta, \dots, 0 \quad (2.12g)$$

A equação (2.12c) propositalmente não existe, para que se possa manter uma correspondência entre as equações (2.11) e (2.12).

Exemplo 4

Mantendo-se as mesmas condições do exemplo 2, seja $\delta = 10$ unidades de volume. Tem-se $d/\delta = 40/10 = 4$ e a nova matriz de transição é de dimensão 10×10 , em vez de 7×7 . Conforme anteriormente explicado, um grande pedaço da matriz do exemplo 2 é aproveitado e as probabilidades de transição remanescentes são calculadas segundo a equação 2.12. A nova matriz de transição A é:

		A.									
	i	j									
		-3	-2	-1	0	1	2	3	4	5	6
-3	0,09	0,09	0,09	0,09	0,05	0,01	0,00	0,00	0,00	0,00	0,00
-2	0,07	0,07	0,07	0,07	0,04	0,02	0,00	0,00	0,00	0,00	0,00
-1	0,09	0,09	0,09	0,09	0,07	0,02	0,01	0,00	0,00	0,00	0,00
0	0,12	0,12	0,12	0,12	0,09	0,04	0,01	0,00	0,00	0,00	0,00
1	0,26	0,26	0,26	0,26	0,25	0,16	0,07	0,02	0,00	0,00	0,00
2	0,21	0,21	0,21	0,21	0,25	0,25	0,16	0,07	0,02	0,01	0,01
3	0,11	0,11	0,11	0,11	0,16	0,25	0,25	0,16	0,07	0,04	0,04
4	0,04	0,04	0,04	0,04	0,07	0,16	0,25	0,25	0,16	0,11	0,11
5	0,01	0,01	0,01	0,01	0,02	0,07	0,16	0,25	0,25	0,21	0,21
6	0,00	0,00	0,00	0,00	0,00	0,02	0,09	0,25	0,50	0,63	0,63

Como antecipado, as quatro primeiras colunas de A. são iguais entre si e a submatriz correspondente à transição a.(i,j), i = 1, 2, ... K + 1; j = 1, 2, ... K + 1 é idêntica à correspondente submatriz do exemplo 2. A interpretação de cada elemento a. (i, j) precisa ser feita com um pouco de cautela, uma vez que Δ = 20 e δ = 10. Por exemplo, a.(-1, 2) = 0,02 significa que, se o reservatório armazena, no início do ano, de 20% a 40% do volume útil, a probabilidade de que ao longo desse ano ocorra um déficit de 10% a 20% do volume útil (entre 25% e 50% da demanda) é de 0,02.

Como no exemplo 3, é possível calcular as matrizes A²., A⁴., A⁸., ... por simples multiplicação. Também como no exemplo 3, a matriz A.³² já tem todas as colunas praticante iguais a FÓRMULA.. Em particular, interessam a. (0) = 0,012; a. (1) = 0,007; a. (-2) = 0,005; e a. (-3) = 0,005.

Verifica-se que, dado que ocorra um racionamento, a probabilidade de que seja inferior a 25% da demanda é de: P(X.(t) = 0 | X. (t) ≤ 0) = 0,012/ (0,012 + 0,007 + 0,005 + 0,005) = 0,012/0,029 = 0,42.

Suponha-se que os racionamentos sejam classificados em quatro categorias, com diferentes custos, segundo o quadro abaixo:

Categoria r	Racionamento		Custo (\$) c(r)
	Magnitude % volume útil	Magnitude (% demanda)	
1	0 - 10	0 - 25	1
2	10 - 20	25 - 50	5
3	20 - 30	50 - 75	20
4	30 - 40	75 - 100	60

O racionamento é da categoria r se x.(t) = 1-r. Portanto, o valor esperado do custo do racionamento é:

$$E(c(R)) = \sum_{i=-3}^0 c(1-i)a. (i)$$

$$= 60 \times 0,005 + 20 \times 0,005 + 5 \times 0,007 + 1 \times 0,012$$

$$= 0,447$$

A parcela anual para amortização do custo de construção somada à \$ 0,447 fornece o valor esperado do custo anual total associado ao projeto. O projetista deverá examinar algumas alternativas para tamanho de reservatório, repetindo em cada caso o procedimento acima descrito para o cálculo de E(C(R)). Deverá ser escolhida a alternativa que conduzir ao menor custo anual total.

O valor esperado do prejuízo é \$0,447, mas o prejuízo, quando ocorrer, será muito maior. Na realidade, será \$1 ou \$5 ou \$20 ou \$60. Para efeito de planejamento de medidas de emergência, é interessante saber o valor esperado do racionamento, dado que ele ocorre:

custo de

$$E(c(R) | X.(t) \leq 0) = E(c(R)) P(X.(t) \leq 0)$$

$$= 0,447/0,029$$

$$= \$ 15,414$$

Nos exemplos 3 e 4, a atenção foi centrada no cálculo da distribuição marginal de probabilidades de X(t), isto é, na distribuição de probabilidades do armazenamento e do déficit em algum ano futuro, suficientemente distante do presente para que a informação sobre o atual estado de armazenamento seja irrelevante. Entretanto, o projetista pode estar interessado na probabilidade de que o reservatório fique pelo menos uma vez vazio, ao longo de n anos de vida útil. Supondo que o reservatório esteja cheio ao ser inaugurado, esta probabilidade é igual a p = P(X(1) = 0 ou X(2)=0 ou ... ou X(n) = 0 | X(0) = K+1). Note que "p" não é a probabilidade de que o reservatório faça uma transição de cheio para vazio em n "passos". Isto é, p ≠ P(X(n) = 0 | X(0) = K+1). O cálculo de p pode ser simplesmente efetuado a partir de um processo modificado X_Δ (t), tal que o estado X_Δ (t) = 0 seja "absorvente". Isto é, imagine-se que, na cadeia modificada, o "reservatório", ao cair no estado vazio, nunca mais saia de lá. Nestas condições, a probabilidade de que X(t) tenha alguma vez sido nula, para t = 1, 2, ..., n, é igual a P(X_Δ(n) = 0 | X_Δ(0) = k+1). A nova matriz de transição A_Δ é idêntica à matriz de transição original "A" exceto sua primeira coluna, que é substituída por a_Δ (0,0) = P(X_Δ(t+1) = 0 | X_Δ(t) = 0) = 1 e a_Δ (i,0) = P(X_Δ(t+1) = i | X_Δ(t) = 0, i = 1, 2, ... K+1).

Exemplo 5

Mantendo-se as mesmas condições do exemplo 2, a matriz de transição A_Δ é:

		A _Δ						
		j						
i		0	1	2	3	4	5	6
0	1	0,25	0,09	0,02	0,00	0,00	0,00	0,00
1	0	0,25	0,16	0,07	0,02	0,00	0,00	0,00
2	0	0,25	0,25	0,16	0,07	0,02	0,01	0,01
3	0	0,16	0,25	0,25	0,16	0,07	0,04	0,04
4	0	0,07	0,16	0,25	0,25	0,16	0,11	0,11
5	0	0,02	0,07	0,16	0,25	0,25	0,21	0,21
6	0	0,00	0,02	0,09	0,25	0,50	0,63	0,63

Calculando-se as matrizes $A_{\Delta}^2, A_{\Delta}^4, A_{\Delta}^8, A_{\Delta}^{10} = A_{\Delta}^2 \cdot A_{\Delta}^8, A_{\Delta}^{20}, A_{\Delta}^{30}$, obtém-se

$$\{A_{\Delta}^n\}_{0,6} = P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = 6) = P(X(1) = 0 \text{ ou}$$

$$X(2) = 0 \text{ ou } \dots X(n) = 0 | X(0) = 6)$$

A tabela seguinte resume os resultados alcançados:

n										
10	20	30	40	50	60	70	80	90	100	
0,09	0,21	0,32	0,41	0,49	0,56	0,61	0,67	0,71	0,75	

$$P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = 6)$$

Percebe-se que, para o caso examinado, a probabilidade de que pelo menos uma vez ocorra algum racionamento ao longo da vida útil típica de 50 anos é de aproximadamente 1/2. É interessante investigar como as probabilidades se alterariam caso o reservatório fosse inaugurado com apenas metade do volume útil, em vez de inteiramente cheio. Trata-se de verificar os valores

$$\{A_{\Delta}^n\}_{0,3} = P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = 3) = P(X(1) = 0 \text{ ou}$$

$$X(2) = 0 \text{ ou } \dots X(n) = 0 | X(0) = 3)$$

A tabela seguinte resume os resultados alcançados:

n										
10	20	30	40	50	60	70	80	90	100	
0,20	0,31	0,40	0,48	0,55	0,61	0,66	0,71	0,75	0,78	

$$P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = 3)$$

Comparando-se as duas tabelas, nota-se que a diferença entre valores correspondentes diminui à medida que n aumenta. Isso ocorre porque, em ambos os casos, há convergência para a unidade:

$$\lim_{n \rightarrow \infty} P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = 6) = \lim_{n \rightarrow \infty} (X_{\Delta}(n) = 0 | X(0) = 3) = 1$$

O problema de dimensionamento do volume útil pode ser visto sob outra ótica. Seja uma seqüência de volumes afluentes ao local do reservatório, $Y(1), Y(2) \dots Y(n)$. A aplicação da curva de massa (Rippl, 1883) permite que se calcule o mínimo volume útil Z_n capaz de garantir uma demanda anual igual a d. Qualquer outra seqüência $y(1), y(2), \dots, y(n)$ resultaria num outro valor para Z_n . Isto é, como $Y(1), Y(2), \dots Y(n)$ é um processo estocástico, Z_n é uma variável aleatória. Gomide (1975) chamou Z_n de máximo déficit acumulado e determinou sua distribuição a partir da seguinte igualdade:

$$P(Z_n > v) = P(X_{\Delta}(n) = 0 | X_{\Delta}(0) = K+1) \tag{2.13}$$

onde v é o volume útil, "alvo" a partir do qual a matriz A_{Δ} é calculada.

Para o caso particular em que $d = E(Y)$, Gomide determinou a distribuição assintótica de Z_n , bem como os momentos

$$E(Z_n) \cong 1,2533 \sqrt{n \text{ var}(Y)} \tag{2.14a}$$

$$\sqrt{\text{var}(Z_n)} \cong \sqrt{0,2661 n \text{ var}(Y)} \tag{2.14b}$$

Esses resultados assintóticos não dependem da distribuição de probabilidades de Y e são importantes porque a distribuição exata da variável padronizada $(Z_n - E(Z_n))/\sqrt{\text{var}(Z_n)}$ já é bem aproximada pela distribuição assintótica, mesmo para pequenos valores de n. O Quadro 47 fornece a distribuição assintótica da variável padronizada.

Quadro 47. Distribuição assintótica do máximo déficit.

u	P(U ≤ u)	u	P(U ≤ u)	u	P(U ≤ u)
-1,6	0,002	-0,2	0,501	1,2	0,878
-1,5	0,006	-0,1	0,542	1,3	0,891
-1,4	0,016	0,0	0,580	1,4	0,904
-1,3	0,034	0,1	0,616	1,5	0,915
-1,2	0,059	0,2	0,650	1,6	0,925
-1,1	0,092	0,3	0,682	1,7	0,934
-1,0	0,132	0,4	0,711	1,8	0,942
-0,9	0,176	0,5	0,739	1,9	0,949
-0,8	0,222	0,6	0,764	2,0	0,955
-0,7	0,270	0,7	0,787	2,1	0,961
-0,6	0,319	0,8	0,809	2,2	0,966
-0,5	0,367	0,9	0,828	2,3	0,971
-0,4	0,413	1,0	0,846	2,4	0,975
-0,3	0,458	1,1	0,863	2,5	0,978

$$U = (z_n - E(Z_n))/(\text{var}(Z_n))^{1/2}$$

Exemplo 6

Mantendo-se as mesmas condições do exemplo 2, o projetista pode estar interessado em qual deve ser o volume útil se a demanda d for aumentada. O limite superior para d é o valor esperado da afluência, $E(Y)$, já que não é concebível "retirar do rio" mais água do que em média lá está disponível. Se d for aumentado do valor $d = 40$ para o limite superior $d = E(Y) = 50$, a probabilidade de algum racionamento ao longo da vida útil de 50 anos certamente será maior do que 0,50 (exemplo 5). Se o projetista quiser manter o mesmo nível de confiabilidade, ele deverá aumentar o volume útil.

Uma maneira de se determinar de quanto exatamente o volume útil deve ser aumentado consiste em realizar um processo de tentativa e erro, com cada tentativa para v definindo uma nova matriz A e, portanto, uma nova probabilidade de transição $A_{0,6}^n$. O processo converge quando $A_{0,6}^n$ for suficientemente próximo da "meta", no caso 0,50 (exemplo 5).

Uma maneira alternativa, bem mais simples, consiste em tomar partido da equação (2.13): fixando-se o lado direito em 0,5, o Quadro 47 fornece o valor de

$$\frac{v' - E(Z_n)}{\sqrt{\text{var}(Z_n)}} = -0,20$$

onde v' é o novo volume útil do reservatório, neste caso coincidente com a mediana de Z_n .

As equações (2.14) fornecem

$$E(Z_n) = 266, \quad \sqrt{\text{var}(Z_n)} = 109$$

Portanto,

$$v' = 244$$

Nota-se que o aumento na demanda de 80% da afluência para 100% (regularização total) implica num aumento de volume útil de 144%, mantendo-se o mesmo nível de confiabilidade.

Diante de tão significativo aumento, o projetista possivelmente desistirá da regularização total ou então diminuirá o padrão de confiabilidade, ou ambos. O Quadro 47 permite que se faça uma rápida avaliação do que acontece, caso se mantenha $v = 100$ e se aumente a demanda para $d = 50$. A probabilidade de que o reservatório fique alguma vez vazio ao longo de $n = 50$ anos é igual a

$$P(Z_n > 100) = P\left(\frac{Z_n - E(Z_n)}{\sqrt{\text{var}(Z_n)}} > \frac{100 - 266}{109}\right) =$$

$$= 1 - P\left(\frac{Z_n - E(Z_n)}{\sqrt{\text{var}(Z_n)}}\right) = 0,995$$

Da mesma forma que o projetista se interessa pela probabilidade de algum racionamento ao longo da vida útil e não apenas num ano qualquer, ele deverá se interessar pelo valor esperado do prejuízo associado com eventuais racionamentos ao longo do intervalo (t, n) , dado o estado de armazenamento $x(t): b(t, x(t))$. Em particular haverá interesse em determinar $b(0, K+1)$, isto é, o valor esperado do prejuízo ao longo da vida útil sendo o reservatório inaugurado completamente cheio.

Seja $c(r)$ o custo associado com um racionamento na faixa $((r-1)\delta; r\delta), r > 0$ (exemplo 4) e $c(r) = 0, r \leq 0$. A probabilidade de que o prejuízo, no último ano de operação do reservatório, venha a ser igual a $c(r)$, dado que $x(n-1) = j$, é igual a $a(1-r, j)$, sendo $a(i, j)$ calculado pela equação (2.12). Portanto, para o último ano da vida útil, tem-se:

$$b(n-1, j) = E(c(R)|t=n-1, x(t)=j) = \sum_i a(i, j) c(1-i) \forall j \tag{2.15}$$

Quando $t = n - 2$, pode ocorrer um racionamento no penúltimo ano da vida útil ou no último ou em ambos. Um pouco de reflexão leva a:

$$b(n-2, j) = \sum_i a(i, j) (c(1-i) + b(n-1, i)) \tag{2.16}$$

Essa equação revela que o valor esperado do custo nos dois últimos anos é igual ao "valor esperado do custo imediato" (durante o penúltimo ano), $\sum a(i, j) c(1-i)$, mais o "valor esperado do custo remanescente" (durante o último ano), $\sum a(i, j) b(n-1, i)$. Como as duas parcelas referem-se a desembolsos feitos em épocas diferentes, é conveniente multiplicar a segunda parcela pela taxa de atualização γ (menor ou igual à unidade), para tornar as duas parcelas financeiramente mensuráveis.

$$b(n-2, j) = \sum_i a(i, j) (c(1-i) + \gamma b(n-1, i)) \tag{2.17}$$

Generalizando o conceito de "custo imediato" e "custo remanescente", tem-se, em geral:

$$b(t-1, j) = \sum_i a(i, j) (c(1-i) + \gamma b(t, i)) \forall j \tag{2.18}$$

Exemplo 7

Mantendo-se as mesmas condições do exemplo 4, a utilização das equações (2.15) – uma única vez – e (2.18) – diversas vezes, recursivamente –, para $\gamma = 1$, resulta nos seguintes valores para $b(t, j)$:

Estado X(t)=j	Armazenamento (% volume útil)	b(t, j)						
		t=0	t=10	t=20	t=30	t=40	t=49	t=50
6	100	22,18	17,22	12,26	7,30	2,48	0,00	0
5	80-100	22,52	17,56	12,60	7,65	2,80	0,00	0
4	60-80	23,77	18,82	13,86	8,90	3,95	0,02	0
3	40-60	26,20	21,24	16,28	11,31	6,20	0,17	0
2	20-40	30,50	25,54	20,58	15,60	10,30	1,02	0
1	0-20	37,87	32,91	27,95	22,96	17,50	4,16	0
0	0	43,25	38,29	33,33	28,34	22,76	7,43	0

Esse quadro informa que o valor esperado do custo de eventuais racionamentos, ao longo dos 50 anos de vida útil, supondo o estado inicial cheio, é de \$22,18. Se houver interesse em antecipar a inauguração do reservatório, para quando o estoque de água for, por exemplo, de 50% do volume útil, o acréscimo de custo ao longo da vida útil será de 26,20-22,18 = \$4,02. A antecipação só deve ser realizada se a vantagem imediata suplantará esse valor.

2. 2 Otimização da Operação

O custo de um racionamento, em geral, varia não-linearmente com a magnitude do racionamento, porque o primeiro 1% da demanda não suprida resulta em inconvenientes sem importância (suspensão de lavagem de carros, por exemplo), ao passo que o último 1% resulta em gravíssimos problemas (as mais elementares necessidades humanas não conseguem ser atendidas). Foi por essa razão que, nos exemplos 4 e 6, os racionamentos nas faixas (0,25), (25,50), (50,75), (75,100), em percentuais da demanda, foram arbitrariamente associados com custos respectivamente iguais a \$1, \$5, \$20 e \$60. Raciocínio similar pode ser feito para outros usos da água, como por exemplo produção de energia elétrica, onde um racionamento de hidreletricidade pode significar que parte da demanda energética passa a ser atendida com o acionamento das usinas termelétricas, até um certo limite, a partir do qual se processa um corte no fornecimento.

Suponha que, num certo ano t, tudo indica que no ano t+1 cerca de 40% da demanda deverá ser cortada. Isto é, no ano t+1, o prejuízo do racionamento deverá ser de \$5 (mantidos os valores do exemplo). O administrador cogita em acionar um esquema de racionamento preventivo. Seu raciocínio é de que se, durante o ano t, 20% da demanda for cortada, o custo será de \$1, deixando 20% da demanda para ser cortada no ano t+1, ao custo de \$1. Dessa maneira, o prejuízo total é de \$1 + \$1 = \$2, em vez de \$5.

Como nunca se sabe qual vai ser a evolução futura do reservatório, o raciocínio acima pode ser generalizado, dizendo-se que um racionamento preventivo deve ser efetuado sempre que resulte numa diminuição do valor esperado do custo de racionamento. O que se procura é uma política de operação que defina, para cada estágio

(ano) t e para cada estado de armazenamento x(t), qual deve ser o racionamento preventivo e qual será o valor esperado do custo de operação (custo de racionamento, nesse caso), seguindo-se a política ótima de operação b*(t, x(t)).

Seja a "variável de decisão" $\Omega = 0, 1, 2, \dots, d/\delta - 1$, que corresponde a racionamentos preventivos respectivamente iguais a 0, $\delta, 2\delta, \dots, d - \delta$. O racionamento real pode ser maior do que o preventivo, num ano t qualquer, se o reservatório esvaziar completamente, isto é, se $x(t+1) = i \leq 0$. Seja o custo do racionamento igual a c(r), quando o racionamento real estiver situado na faixa entre (r-1) δ e r δ , $r > 0$ e seja $c(r) = 0, r \leq 0$. Quando uma decisão Ω é tomada, o custo do racionamento será c(r), onde

$$r = \Omega + \max(0, 1-i) \tag{2.19}$$

Para um certo estágio "t", suponha que a função b*(t+1, i) seja conhecida, qualquer que seja x(t+1) = i. Para cada estado x(t) = j e decisão Ω , é possível calcular a função b(t, j, Ω), definida por:

$$b(t, j, \Omega) = \sum_i a_+(i, j, \Omega) (c(r) + \gamma b^*(t+1, i)) \tag{2.20}$$

onde

b(t, j, Ω) - valor esperado do curso de operação, quando se toma a decisão Ω .

γ - taxa de atualização (equação(2.17))

$a_+(i, j, \Omega)$ - probabilidade de transição do estado j para o estado "i", quando se toma a decisão Ω .

O cálculo de $a_+(i, j, \Omega)$ é feito através da equação (2.12), substituindo-se o valor de "d" por d - $\Omega\delta$.

A melhor decisão, $\Omega^*(t, j)$, será aquela que minimize b(t, j, Ω). Portanto,

$$b^*(t, j) = \min_{\Omega} b(t, j, \Omega) \tag{2.21}$$

As equações (2.19) a (2.21) definem um algoritmo recursivo para cálculo da política ótima de operação. Resta definir a condição de fronteira, para t = n-1.

Assume-se a hipótese de que, ao final da vida útil, o reservatório é desativado, o que significa que é irrelevante o estado de armazenamento

$$b^*(n, i) = 0 \tag{2.22}$$

Os leitores familiarizados com técnicas de otimização reconhecerão nas equações (2.19) a (2.22) o algoritmo conhecido como programação dinâmica estocástica.

Exemplo 8

Mantendo-se as mesmas condições do exemplo 6, a utilização das equações (2.19) a (2.22), para $\gamma=1$, resulta nos seguintes valores para b*(t, j):

Estado X.(t)=j	Armazenamento (% volume útil)	b*(t, j)						
		t=0	t=10	t=20	t=30	t=40	t=49	t=50
6	100	13,37	10,46	7,54	4,63	1,74	0,00	0
5	80-100	13,57	10,66	7,75	4,83	1,94	0,00	0
4	60-80	14,31	11,40	8,48	5,57	2,66	0,02	0
3	40-60	15,76	12,85	9,94	7,02	4,08	0,17	0
2	20-40	18,19	15,28	12,37	9,45	6,50	1,02	0
1	0-20	22,22	19,32	16,40	13,49	10,49	3,54	0
0	0	25,08	22,17	19,26	16,34	13,34	5,21	0

Comparando-se os valores dessa tabela com os correspondentes valores da tabela do exercício 6, verifica-se a importância, nesse caso, do racionamento preventivo. Por exemplo, o valor esperado do custo de operação, racionamentos preventivos e/ou involuntários, ao longo de 50 anos de vida útil, e supondo o estado inicial cheio, cai de \$22,18 para \$ 13,37.

A política ótima de operação, nesse caso é, para $t \geq 47$:

$$X.(t) \leq 1 \rightarrow \Omega = 1; X.(t) > 1 \rightarrow \Omega = 0$$

para $t < 47$:

$$X.(t) \leq 2 \rightarrow \Omega = 1; X.(t) > 2 \rightarrow \Omega = 0$$

Para os últimos três anos de operação ($t = 47, 48, 49$), só se aplica um racionamento preventivo se o reservatório já começar o ano com um armazenamento abaixo de 20% do volume útil. Faz-se, nesse caso, um racionamento preventivo de 10% do volume útil ($\Omega = 1$). Para todos os demais anos ($t = 0, 1, 2, \dots, 49$), faz-se um racionamento preventivo de 10% do volume útil ($\Omega = 1$), se o estoque no reservatório, no início do ano, for inferior a 40%. Em nenhuma circunstância, a política ótima recomenda uma decisão Ω maior do que 1. Isto é, racionamentos maiores do que 10% do volume útil não são aconselháveis.

A consideração de racionamentos preventivos altera a matriz de transição entre os estados. Por exemplo, para $t < 47$, apenas as colunas $j > 2$ da nova matriz de transição "A*." serão idênticas às correspondentes colunas da matriz "A". (exemplo 4), quando nenhum racionamento preventivo foi considerado. As demais colunas corresponderão a probabilidades de transição $a.(i., j.)$, calculadas pela equação 2.12, substituindo 40 (valor de "d") por 30 (valor de d-δ).

A nova matriz "A*." é:

		A*.								
i	j									
		-2	-1	0	1	2	3	4	5	6
-2		0,09	0,09	0,09	0,05	0,02	0,00	0,00	0,00	0,00
-1		0,09	0,09	0,09	0,07	0,02	0,01	0,00	0,00	0,00
0		0,12	0,12	0,12	0,09	0,04	0,01	0,00	0,00	0,00
1		0,26	0,26	0,26	0,25	0,16	0,07	0,02	0,00	0,00
2		0,21	0,21	0,21	0,25	0,25	0,16	0,07	0,02	0,01
3		0,11	0,11	0,11	0,16	0,25	0,25	0,16	0,07	0,04
4		0,04	0,04	0,04	0,07	0,16	0,25	0,25	0,16	0,11
5		0,01	0,01	0,01	0,02	0,07	0,16	0,25	0,25	0,21
6		0,00	0,00	0,00	0,00	0,02	0,09	0,25	0,50	0,63

Nota-se, como esperado, que $a*(i, j) \leq a.(i, j)$, para valores de $i = -3, -2, -1, 0$ e $\forall j$.

Existe uma outra técnica para o cálculo da função $b*(t, j)$, e conseqüentemente da decisão ótima $\Omega*(t, j)$, que dispensa o emprego do esquema recursivo representado pelas equações 2.19 a 2.22. O método alternativo toma partido do fato de que para valores de "t" suficientemente pequenos, isto é, para anos suficientemente distantes do fim do horizonte de operação, a decisão ótima $\Omega*(t, j)$ é estável de ano para ano e $b*(t, j)$ varia linearmente com t, como se pode observar na tabela do exercício 8. Howard (1960, capítulo 4) descreve o método alternativo, chamado de "iteração de política".

2.3. Miscelânea

O cálculo numérico da distribuição de probabilidades de $X(t)$ e dos processos assemelhados $X.(t)$ e $X_{\Delta}(t)$ foi o objeto principal deste item. A rigor, todos os cálculos dependeram apenas das relações $d/v, \delta/v, E(Y)/v$ e $\sqrt{\text{var}(Y)}/v$, bem como da distribuição de probabilidades de $Y(t)$. Existem diversos resultados disponíveis na literatura que fornecem expressões matemáticas para a distribuição de $X(t)$, para diferentes hipóteses quanto à distribuição de $Y(t)$. Lloyd (1967) fornece um revisão sobre o assunto.

A distribuição de probabilidades do máximo déficit acumulado Z_n , dado no Quadro 47, é apenas um dos resultados de uma linha de pesquisa conhecida por "análise de propriedades de somas parciais" (Gomide, 1985, 1983 e 1986). Existem diversos resultados disponíveis na literatura relacionados não apenas a Z_n , mas também aos momentos e à distribuição de probabilidades da variável aleatória "máxima amplitude". Durante anos, essa variável aleatória desempenhou um papel central na escolha de modelos estocásticos para o processo $Y(t)$, devido ao chamado "fenômeno de Hurst" (item 4.4).

A extensão da abordagem mostrada neste capítulo, para o caso em que as aflúncias em anos consecutivos ($Y(t), Y(t+1)$) forem variáveis dependentes, poderia, em princípio, ser feita considerando-se a variável de estado bivariada ($X(t), Y(t)$). A transição de ($X(t), Y(t)$) para ($X(t+1), Y(t+1)$), quando viável, seria associada à probabilidade de $Y(t+1)$, dado $Y(t) = y(t)$. Se a variável $Y(t)$ for discretizada em K' diferentes valores, a matriz de transição teria dimensão $K'(K+2)$ por $K'(K+2)$. Dependendo dos valores de K

e K' , a matriz de transição pode ser excessivamente grande, tornando o problema difícil de ser tratado. Essa intratabilidade computacional torna-se ainda mais acentuada quando o dimensionamento e/ou a otimização da operação envolve diversos reservatórios. Nesse caso, o método experimental, também chamado de Monte Carlo, associado a modelos de simulação, é a abordagem mais indicada. Assunto para o próximo item.

3.0 MÉTODO DE MONTE CARLO

3.1 Descrição do método

Suponha que se pretenda planejar ou operar algum aproveitamento hídrico num local com uma série histórica de n anos. Essa série é usualmente utilizada como dado de entrada para um modelo de simulação e/ou otimização que produza como saída a grandeza relevante para o projeto. Por exemplo, o volume de reservatório necessário para regularizar uma descarga prefixada, o armazenamento "meta" ao final de cada mês ou a descarga máxima que o vertedor deve ser capaz de dar passagem. A Figura 112 apresenta a série histórica de uma seção do rio na qual se pretende construir uma barragem que crie um reservatório de regularização. O modelo de simulação/otimização, nesse caso, é simplesmente o diagrama de Rippl, que produz como saída v , o volume de reservatório.

Como a série histórica é apenas uma das possíveis realizações de um processo estocástico, pode-se imaginar que a Natureza "sorteou" a série histórica segundo algum conjunto de leis probabilísticas. Um novo sorteio redundaria numa outra série, diferente da histórica, mas igualmente provável. Mais adiante serão apresentados modelos que procuram aproximar esse comportamento estocástico. Cada um desses modelos permite que artificialmente se façam tantos sorteios quanto forem necessários para o estudo em foco.

Cada sorteio estará associado a uma série de vazões, chamada de série sintética. Como essas séries serão todas distintas entre si, podem-se obter diversos resultados provenientes de simulações, ao invés de um único resultado, que seria obtido caso apenas a série histórica estivesse disponível. Dessa forma, a informação contida na série histórica pode ser mais completamente extraída. A Figura 115 mostra que a série histórica é dado de entrada para o modelo estocástico, que produz tantas séries sintéticas quanto se queira. Como a cada série estocástica estará associado um tamanho de reservatório, é possível inferir a distribuição de probabilidades de V a partir da amostra aleatória (v_1, v_2, v_3, \dots).

Caso se pretenda definir o volume do reservatório de forma tal que a probabilidade de sucesso seja α , deve-se adotar o valor v_α . Caso se pretenda definir o volume pelo método tradicional, isto é, v_h , pode-se avaliar a probabilidade de sucesso β . Nesse contexto, sucesso significa o atendimento da descarga que se escolheu para regularizar, num horizonte de planejamento igual ao da série histórica.

É importante ressaltar, no entanto, que, quando se projeta para v_α , o horizonte de planejamento pode ser diferente, inclusive mais longo, do que o número de anos de série histórica. Basta que as séries sintéticas tenham o comprimento do desejado horizonte de planejamento.

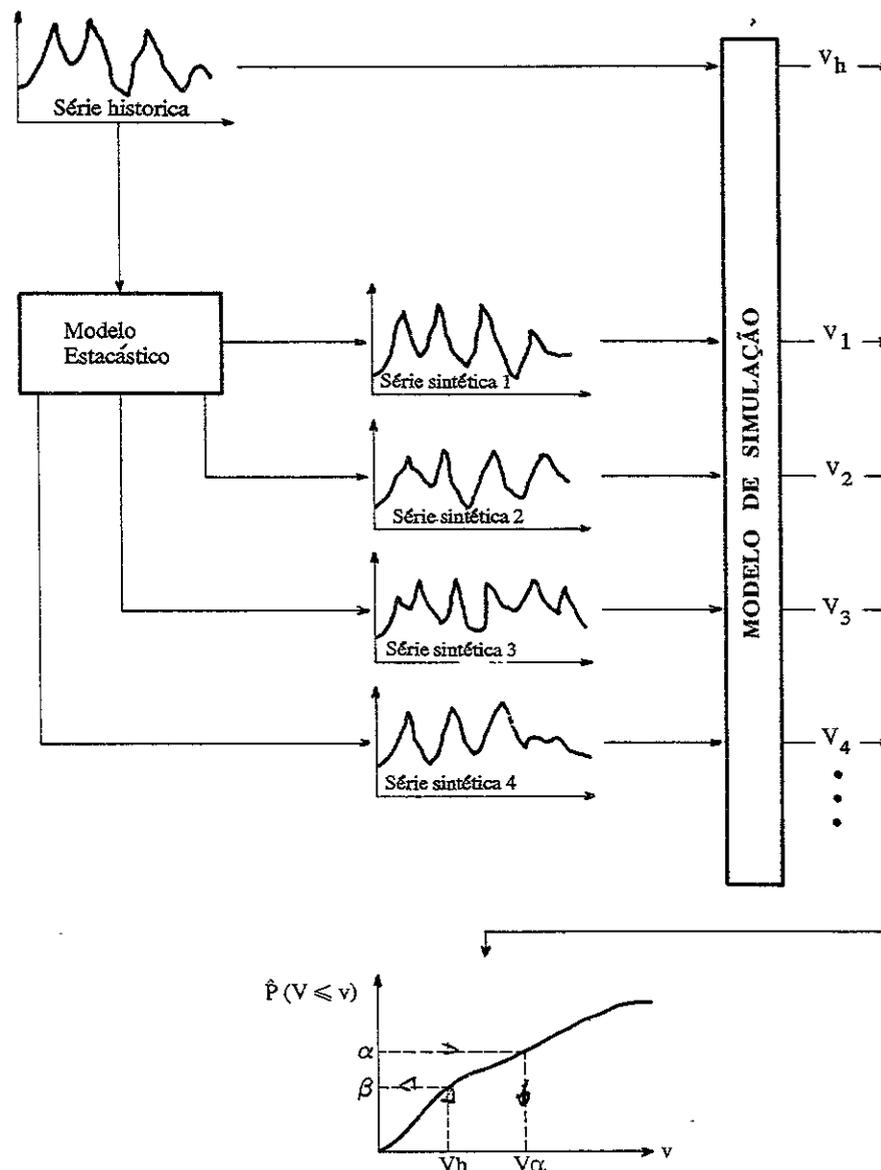


Figura 115. Representação esquemática do método de Monte Carlo

Exemplo 9

Como no exemplo 2, suponha que a demanda anual a ser atendida por um reservatório seja $d = 40$ unidades de volume e as afluições anuais ao reservatório sejam variáveis independentes normalmente distribuídas - $Y \sim N(E(Y) = 50, \text{var}(Y) = 30^2)$. Suponha ainda que uma série histórica com $n = 50$ anos esteja disponível e que a aplicação do diagrama de Rippl a essa série resulte num tamanho de reservatório $v_h = 100$. Essa é a menor capacidade que pode ter o reservatório para garantir o suprimento, caso nos próximos 50 anos se repitam as afluições observadas nos últimos 50 anos. O problema é definir qual a probabilidade de sucesso β associado a v_h e qual o tamanho de reservatório associado, por exemplo, a $\alpha = 0,95$.

A solução pode ser encontrada simulando-se o comportamento de um reservatório "sem fundo", um número grande de vezes, por exemplo mil vezes. A idéia é assumir que o reservatório começa cheio e nunca se observa um déficit para cada uma das simulações ao longo do horizonte de planejamento m (neste caso igual a $n = 50$ anos). Portanto, cada simulação produz uma capacidade de armazenamento igual ao armazenamento inicial (cheio) menos o mínimo armazenamento ao longo de 50 anos (vazio). A diferença entre os armazenamentos não é alterada, qualquer que seja o valor que se arbitre para o armazenamento inicial. Em particular, quando se admite o estado inicial nulo, a capacidade de armazenamento de cada simulação será o valor absoluto do mínimo armazenamento, ao longo dos 50 anos.

O modelo estocástico neste caso é extremamente simples. Como a afluição, em cada ano, é assumida independentemente das afluições em outros anos, cada série sintética é produzida sorteando-se 50 observações de uma variável aleatória normalmente distribuída com valor esperado igual a 50 e variância 30^2 .

As mil simulações podem ser facilmente realizadas através de um programa de computador que implemente o seguinte algoritmo:

$$K=1.000, m=50, \mu=50, \sigma=30, d = 40$$

Repita $i= 1, K$

$$v(i)= 0$$

$$x= 0$$

Repita $j= 1, m$

Sorteie uma variável normal padrão, z

$$y=\mu +\sigma*z$$

$$x= \min (0, x+y-d)$$

Se $x < v(i)$ então $v(i) = x$

$$v(i)= -v(i)$$

A coleção $v(1), v(2), \dots, v(1.000)$ forma uma amostra aleatória de mil pontos da variável V , tamanho do reservatório. A ordenação da amostra resulta numa razoável aproximação para a distribuição de probabilidades de V . Para uma efetiva corrida do programa, obteve-se média e desvio padrão amostrais de V respectivamente iguais a 109 e 47. A seguir, alguns elementos da amostra ordenada.

Nº ordem (i)	v(i)	Nº ordem (i)	v(i)	Nº ordem (i)	v(i)
1	28	401	90	801	140
51	52	451	96	851	154
101	59	501	99	901	173
151	65	551	104	941	190
201	71	601	111	951	202
251	76	651	117	961	210
301	81	701	125	1.000	386
351	86	751	133		

É interessante observar que o resultado $P(V \leq 100) \cong 0,5$, que é uma estimativa para β , já havia sido antecipado na solução do exemplo 5, através do emprego da Cadeia de Markov. Da ordenação acima extrai-se que o valor de v_α associado com $\alpha= 0,95$ é aproximadamente igual a 202.

O método de Monte Carlo serve também para prever a operação de um sistema de recursos hídricos, quando as estruturas hidráulicas já estão dimensionadas e os armazenamentos iniciais são dados pelas condições atuais do sistema.

Exemplo 10

Suponha um reservatório com capacidade de armazenamento de 60 unidades volumétricas, com um estoque inicial de dez unidades (condição presente), demandas nos próximos três anos respectivamente iguais a 35, 40 e 45 unidades e demais condições idênticas ao exemplo 9. Imagine ainda que haja interesse em prever o estado do armazenamento daqui a três anos. Ignorando o método de Monte Carlo, uma solução aproximada para o problema poderia ser facilmente alcançável substituindo-se o problema estocástico inicial por um problema determinístico. Tipicamente, as futuras e desconhecidas afluições seriam substituídas pelos correspondentes valores esperados. Para o caso específico, o valor esperado da afluição é 50:

Ano	Volume inicial	Afluição	Defluência	Volume final
1	10	+ 50	- 35	25
2	25	+ 50	- 40	35
3	35	+ 50	- 45	40

Logo, o armazenamento ao final do terceiro ano poderia ser previsto, segundo este método, em 40 unidades volumétricas. A dificuldade da aproximação consiste na indefinição da precisão da resposta e na incapacidade de prover uma distribuição de probabilidades para a variável aleatória "armazenamento ao final do terceiro ano", que permite, por exemplo, o cálculo da probabilidade de que o reservatório esteja vazio.

O problema seria resolvido, segundo o método de Monte Carlo, pela implementação do seguinte algoritmo:

$K = 1.000, m = 3, \mu = 50, \sigma = 30, d(1) = 35, d(2) = 40, d(3) = 45$
 $x_{max} = 60$

Repita $i = 1, k$

$x = 10$
 Repita $j = 1, m$

Sorteie uma variável normal padrão z

$y = \mu + \sigma z$
 $x = \max(0, \min(x_{max}, x + y - d(j)))$
 $v(i) = x$

A coleção $v(1), v(2), \dots, v(1.000)$ forma uma amostra aleatória de mil pontos da variável aleatória "armazenamento ao final do terceiro ano". Para uma efetiva execução do programa, obteve-se média e desvio padrão amostrais de v respectivamente iguais a 34,5 e 23,2, o que permite afirmar, com grande segurança, que o verdadeiro valor esperado de v deve estar contido no intervalo $34,5 \pm 2 \times 23,2 / \sqrt{1.000}$, isto é, entre 33 e 36. É interessante observar que a previsão obtida com o "substituto determinístico" (40) situa-se fora do intervalo. Isso ocorre pela mesma razão que faz com que o valor esperado de uma função de uma variável aleatória não seja, em geral, igual à função do valor esperado. Isto é, se $Y = g(x)$, em geral $E(Y) \neq g(E(x))$.

Nessa mesma execução do programa ocorreram 151 valores de $v(i)$ idênticos a zero. Isto é, a probabilidade de que o reservatório esteja vazio ao final do terceiro ano é da ordem de 0,15.

3.2 Número de simulações

No exemplo 9, a quantidade de simulações escolhida foi 1.000, supondo-se que a distribuição de probabilidade empírica de V , derivada a partir de mil pontos amostrais, já fosse suficientemente próxima da distribuição teórica. O grau de aproximação pode ser medido de duas maneiras.

- a) Para um tamanho de reservatório v , arbitrado, a probabilidade $\alpha = P(V \leq v)$ é estimada contando-se quantos elementos da amostra $\{v(1), v(2), \dots, v(k)\}$ são menores que v . Seja a variável aleatória W tal que

$$W = 1 \Leftrightarrow V \leq v$$

e

$$W = 0 \Leftrightarrow V > v$$

É simples mostrar que

$$E(W) = \alpha \text{ e } \text{var}(W) = \alpha(1-\alpha) \tag{3.1}$$

seja $\hat{\alpha}$ o estimador de α definido por

$$\hat{\alpha} = \frac{1}{K} \sum_{k=1}^K W(k) \tag{3.2}$$

Logo,

$$E(\hat{\alpha}) = \alpha \text{ e } \text{var}(\hat{\alpha}) = \frac{\alpha(1-\alpha)}{K} \tag{3.3}$$

Quando K é grande, a distribuição de probabilidade de $\hat{\alpha}$ tende a ser normal, devido ao teorema do limite central. O Quadro 48 fornece valores para o coeficiente de variação de $\hat{\alpha}$ ($\sqrt{\text{var}(\hat{\alpha})}/E(\hat{\alpha})$) para diversos valores de α e K .

Quadro 48. Coeficiente de variação de $\hat{\alpha}$.

$\alpha \backslash K$	10	100	1.000	10.000	100.000
0,05	—	0,44	0,14	0,04	0,01
0,10	0,95	0,30	0,09	0,03	0,01
0,50	0,32	0,10	0,03	0,01	0,00
0,90	0,11	0,03	0,01	0,00	0,00
0,95	0,07	0,02	0,01	0,00	0,00

Exemplo 11

No exemplo 9, estimou-se que $P(V \leq 100)$ é aproximadamente 0,50. Com base no Quadro 48, seria razoável afirmar que existe uma probabilidade de 95% para que o desconhecido número esteja situado no intervalo $0,50 (1 \pm 1,96 \cdot 0,03)$. Isto é, entre 0,47 e 0,53.

b) Para uma probabilidade $\alpha = P(V \leq v_\alpha)$, arbitrada, um intervalo de confiança para V_α pode ser construído a partir da amostra ordenada $v(1), \dots, v(K)$.

Assim como para dois eventos A e B quaisquer, definidos no mesmo espaço amostral vale a identidade $P(A) = P(A \cap B) + P(A \cap \bar{B})$; onde \bar{B} é o evento complementar de B, tem-se:

$$P(V(j) \leq v_\alpha) = P(V(j) \leq v_\alpha \leq V(u)) + P(V(u) \leq v_\alpha) \quad (3.4)$$

onde j e u são inteiros $1 \leq j \leq u \leq K$

Logo,

$$P(V(j) \leq v_\alpha \leq V(u)) = P(V(j) \leq v_\alpha) - P(V(u) \leq v_\alpha) \quad (3.5)$$

Mas,

$$\begin{aligned} P(V(j) \leq v_\alpha) &= P(\text{pelo menos } j \text{ observações sejam inferiores a } v_\alpha, \text{ em } K \text{ sorteios}) \\ &= P(\text{exatamente } j \text{ observações sejam inferiores a } v_\alpha, \text{ em } K \text{ sorteios}) \\ &\quad + P(\text{exatamente } j+1 \text{ observações sejam inferiores a } v_\alpha, \text{ em } K \text{ sorteios}) \end{aligned}$$

+

.

.

$$+ P(\text{exatamente } K \text{ observações sejam inferiores a } v_\alpha, \text{ em } K \text{ sorteios})$$

$$= \sum_{i=j}^K \binom{K}{i} \alpha^i (1-\alpha)^{K-i} \quad (3.6)$$

Analogamente

$$P(V(u) \leq v_\alpha) = \sum_{i=u}^K \binom{K}{i} \alpha^i (1-\alpha)^{K-i} \quad (3.7)$$

Logo

$$P(V(j) \leq v_\alpha \leq V(u)) = \sum_{i=j}^{u-1} \binom{K}{i} \alpha^i (1-\alpha)^{K-i} \quad (3.8)$$

Para valores de K grandes, a equação 3.8 pode ser aproximada por: (a distribuição binomial converge para a normal)

$$P(V(j) \leq v_\alpha \leq V(u)) \cong \Phi\left(\frac{u - K\alpha}{\sqrt{K\alpha(1-\alpha)}}\right) - \Phi\left(\frac{j - K\alpha}{\sqrt{K\alpha(1-\alpha)}}\right) \quad (3.9)$$

Exemplo 12

No exemplo 9, o valor de v_α , para $\alpha = 0,95$, foi estimado em 202, já que $v(951) = 202$ e $K = 1.000$. Naquele exemplo viu-se também que $v(941) = 190$ e $v(961) = 210$. A probabilidade de que o verdadeiro valor de v_α esteja situado no intervalo [190,210] pode ser calculado através do uso da equação 3.9.

$$P(190 \leq v_\alpha \leq 210) \cong \Phi\left(\frac{961 - 950}{\sqrt{950 \cdot 0,05}}\right) - \Phi\left(\frac{941 - 950}{\sqrt{950 \cdot 0,05}}\right) = 0,85$$

Portanto, o intervalo de confiança [190,210] está associado ao grau de confiança de 85%.

3.3 Estudos de sensibilidade

No exemplo 9, tanto o modelo de simulação (Rippl) quanto o modelo estocástico para as afluições anuais (vazões independentes) foram extremamente simples. Na realidade, o exemplo serviu apenas para ilustrar uma maneira alternativa de se resolver um problema que, a rigor, já havia sido resolvido no item 4.2, através da Cadeia de Markov. O método de Monte Carlo, no entanto, é particularmente apropriado para a solução de problemas mais complexos, em que os métodos do item 4.2 não são facilmente aplicáveis. Por exemplo, quando as afluições anuais são dependentes entre si ou quando se dimensiona não apenas um reservatório e sim um sistema de reservatórios.

Exemplo 13

Suponha que as afluições anuais ao reservatório do exemplo 9 sejam dependentes entre si, com coeficiente de autocorrelação (item 4.4) igual a ρ . Sejam, por exemplo, $\rho = 0,3$. A solução deste problema pelo método de Monte Carlo é praticamente equivalente, em termos de dificuldade computacional, à solução do problema original, em que as afluições eram independentes entre si. Em contraposição, o leitor recordará que no item 4.2 foi feita menção de que a solução pelo método da Cadeia de Markov exigiria a manipulação de matrizes de transição muito grandes. O seguinte algoritmo seria utilizado.

$$K= 1.000, m=50, \mu=50, \sigma=30, d=40, \rho=0,3$$

```

Repita i = 1, K
  v(i) = 0
  x = 0

  Sorteie uma variável normal padrão, z
  y =  $\mu + \sigma * z$ 

  Repita j=1, m
    Sorteie uma variável normal padrão, z
    y =  $\mu + \rho (y - \mu) + \sigma * \sqrt{1 - \rho^2} * z$ 
    x =  $\min(0, x + y - d)$ 
    Se  $x < v(i)$ , então  $v(i) = x$ 
  v(i) = -v(i)
    
```

O leitor notará que, quando $\rho = 0$, o algoritmo acima e o do exemplo 9 são equivalentes. A equação utilizada para definir y (nona linha do algoritmo) será derivada no item 4.4. Como no exemplo 9, a coleção $v(1), v(2), \dots, v(1.000)$ forma uma amostra aleatória de mil pontos da variável aleatória V , tamanho do reservatório. Para uma efetiva corrida do programa, obteve-se média e desvio padrão amostrais respectivamente iguais a 150 e 78. A seguir, alguns elementos da amostra ordenada.

Nº ordem	v(i)	Nº ordem	v(i)	Nº ordem	v(i)
1	19	351	110	751	187
51	57	401	117	801	204
101	67	451	125	851	227
151	78	501	134	901	257
201	86	551	144	941	290
251	96	601	153	951	300
281	100	651	163	961	315
301	104	701	173	1.000	588

É interessante observar o efeito causado pela consideração da dependência da afluição em anos consecutivos, mesmo quando o coeficiente de autocorrelação é pequeno (0,3): a probabilidade de sucesso para um reservatório de tamanho igual a 100 cai de 0,50 para 0,28, e o tamanho de um reservatório associado a uma probabilidade de sucesso de 0,95 sobe de 202 para 300.

Estudos de sensibilidade, tais como a avaliação do efeito quando ρ varia de 0,0 para 0,3, são facilmente executáveis pelo método de Monte Carlo. Tudo se passa como se o computador fosse um laboratório experimental.

3.4 Escolha do modelo estocástico

Identificação do modelo

A disponibilidade de muitas séries sintéticas pode dar a impressão de que é possível ter confiança absoluta nos resultados encontrados com o seu emprego. É importante ressaltar, no entanto, que a geração de séries sintéticas não cria informação alguma. Existem até casos em que é preferível utilizar apenas a série histórica (Staschus e Kelman, 1987).

O processo de seleção de um modelo e de estimação de seus parâmetros é feito, como se verá nos próximos capítulos, a partir de inferência com a série histórica. Portanto, a utilização do modelo estocástico de vazões deve ser encarada simplesmente como uma das maneiras de se manipular a informação existente. Nenhuma nova informação é criada, a não ser pela adoção de hipóteses subjacentes ao modelo estocástico. Essas hipóteses devem ser exaustivamente testadas antes que se aceite qualquer modelo. Infelizmente, nunca se pode ter absoluta certeza quanto à identidade do verdadeiro modelo utilizado pela natureza.

A primeira etapa na escolha do modelo estocástico, a identificação do modelo, é muito dependente da experiência do hidrólogo. Trata-se de escolher o modelo que seja mais indicado para um particular rio ou bacia hidrográfica. Para facilitar a discussão, imagine que apenas os dois modelos empregados nos exemplos 9 e 13 estejam sendo considerados para ajuste à série histórica de um rio hipotético. Esses modelos serão descritos com mais detalhes no item 4.4.

modelo 1:

$$Y_{t+1} = \mu + \sigma Z_{t+1}$$

modelo 2:

$$Y_{t+1} = \mu + \rho (Y_t - \mu) + \sigma \sqrt{1 - \rho^2} Z_{t+1}$$

onde Y_t = vazão no ano t

$$\mu = E(Y_t)$$

$$\sigma = \text{var}(Y_t)$$

$$\rho = \text{corr}(Y_t, Y_{t+1})$$

Z_{t+1} - "ruído" normal padrão. Isto é, $Z_{t+1} \sim N(0,1)$, $\forall t$ e Z_i independente de Z_j para $i \neq j$.

A etapa de identificação neste exemplo consiste em decidir qual das duas formulações é a mais indicada. Conforme anteriormente mencionado, o modelo 1 é um caso particular do modelo 2, para $\rho = 0$.

Como não existem razões físicas para um valor de ρ negativo, esta hipótese não é sequer cogitada. Por outro lado, valores positivos para ρ são perfeitamente aceitáveis, devido ao armazenamento de água no subsolo.

As características estocásticas do reservatório existente no subsolo podem ser descritas pelos métodos empregados no item 4.2. A analogia com reservatório superficial é útil para enfatizar que, mesmo quando as afluições ao reservatório sejam independentes em anos consecutivos, os estados de armazenamento poderão ser dependentes, desde que a capacidade do reservatório permita regularização plurianual. Sabe-se que o escoamento básico varia diretamente com o estado do armazenamento subterrâneo. Este é um conceito adotado por um grande número de modelos de transformação de chuva em vazão, que pode ser verificado empiricamente pela simples observação do grande número de nascentes que secam após uma prolongada estiagem.

Conseqüentemente o escoamento básico poderá ser dependente em anos consecutivos, mesmo que os processos meteorológicos não sejam dotados de suficiente "memória" para fazer com que a precipitação anual, e também o total infiltrado, sejam variáveis dependentes. Como o deflúvio anual pode ser conceitualmente separado em escoamento superficial e básico, pode-se dizer que a escolha do modelo matemático mais plausível, entre os dois candidatos acima especificados, depende da característica geológica (capacidade do armazenamento subterrâneo) da bacia hidrográfica estudada.

Idealmente, as informações sobre as características do solo, subsolo e clima deveriam ser consideradas não apenas para a identificação do modelo estocástico mais apropriado, como também para a estimativa dos parâmetros. A hidrologia, no entanto, não tem ainda uma metodologia capaz de atingir, na prática, esse objetivo, apesar de inúmeros esforços feitos nessa direção (por exemplo, Koch, 1985). Com base apenas na série histórica de vazões anuais, a discriminação entre os dois modelos pode ser feita através do seguinte teste de hipótese.

hipótese nula: $\rho = 0$
hipótese alternativa: $\rho > 0$

Se a hipótese nula puder ser rejeitada, o modelo 1 deve ser também rejeitado, possivelmente em favor do modelo 2. O coeficiente de correlação amostral $\hat{\rho}$ é a estatística indicada para a realização do teste (item 4.4).

Pode-se mostrar que $\hat{\rho}$ tem distribuição aproximadamente normal, com valor esperado igual a zero (caso a hipótese nula esteja correta) e desvio padrão igual a $(n)^{-1/2}$. Conseqüentemente, a hipótese nula deverá ser rejeitada com um nível de confiança igual a α (tipicamente 0,95), e com ela também o modelo 1, caso $\hat{\rho} > z_{\alpha} n^{-1/2}$, onde z_{α} é o quantil de uma variável normal padrão.

Como se sabe, a não rejeição da hipótese nula não equivale a sua aceitação. Ou seja, é desconhecida a probabilidade de engano (erro tipo II), caso o modelo 1 seja aceito com base na não rejeição da hipótese nula. Por outro lado, caso a hipótese nula seja rejeitada, a probabilidade de se cometer um engano (erro tipo I) é de apenas $1-\alpha$. Isso quer dizer que o teste só fornece um resultado afirmativo caso resulte na rejeição do modelo 1. Caso contrário, ambos os modelos permanecem na competição.

Um critério possível para discriminar entre dois ou mais modelos competidores é verificar qual a verossimilhança atribuída por cada um deles à série histórica: deve-se

escolher o modelo para o qual a série histórica seja uma típica realização do processo estocástico. Convém lembrar que o método da máxima verossimilhança para estimativa de parâmetros baseia-se num princípio análogo. Esse procedimento não pode ser empregado quando os modelos competidores tiverem número diferente de parâmetros, porque o modelo com mais parâmetros tenderá a se ajustar melhor aos dados, e, conseqüentemente, produzir uma maior verossimilhança, sem que necessariamente seja melhor. Na realidade devem-se utilizar modelos com o menor número possível de parâmetros, para se evitar que flutuações amostrais da série histórica sejam consideradas propriedades de população. Esse é o chamado "princípio da parcimônia" (Box e Jenkins, 1976).

O critério de Akaike é uma medida que procura balancear a meta de parcimônia de parâmetros com o objetivo de se obter um bom ajuste, com base na função de verossimilhança:

$A = -2 \ln(\text{máxima verossimilhança}) + 2(\text{número de parâmetros})$.

Segundo o critério, deve ser selecionado o modelo com o menor valor de A . Recomenda-se a prudência de que este critério não seja utilizado isoladamente. Afinal, é possível que dois modelos com valores de A muito próximos resultem em decisões de engenharia bastante distintas. Ao selecionar um modelo, o hidrólogo deve ter em mente as incertezas na estimativa dos parâmetros e, principalmente, os resultados dos testes de validação, assunto tratado mais adiante. Caso não haja uma clara definição em favor de um dos modelos competidores, a decisão final deverá seguir a tradição da engenharia de adotar a solução que fique em favor da segurança.

Box e Jenkins (1976) apresentaram diversos procedimentos mais elaborados para a identificação de modelos aplicáveis a séries temporais pertencentes à família Arma (item 4.4). Hipel e McLeod (1986) ampliaram a metodologia para aplicação específica a séries hidrológicas. Recomenda-se ao leitor que consulte estas referências, uma vez que o assunto não será tratado em profundidade aqui.

Estimativa dos Parâmetros

Mesmo que o verdadeiro modelo fosse conhecido, restaria ainda estimar os parâmetros a partir da única informação disponível: a série histórica.

Na prática, grandezas amostrais, calculadas freqüentemente a partir de registros curtos, são utilizadas em lugar dos valores de população. A geração de uma série sintética suficientemente longa faz com que as grandezas amostrais da série sintética convirjam para as grandezas amostrais da série histórica e não para as grandezas populacionais, como seria desejável. Isto é, os erros amostrais da série histórica são perpetuados na série sintética. Esse defeito do método de Monte Carlo é chamado de tendenciosidade operacional.

Uma maneira de levar em consideração a incerteza quanto ao valor numérico dos parâmetros consiste em reconhecer que os estimadores dos parâmetros são também variáveis aleatórias. A geração de k séries sintéticas se processa segundo o algoritmo:

Repita $i = 1, a$
Sorteie um vetor de parâmetros (seja θ_i) da distribuição de probabilidades do estimador não tendencioso $\hat{\theta}$, seja $F_{\hat{\theta}}(\cdot)$

Repita $j = 1, b$
Sorteie uma série sintética de vazões, dado o parâmetro θ_i

onde $k = a \times b$

Exemplo 14

Suponha as afliências anuais do exemplo 9, $Y \sim N(E(Y) = 50, \text{var}(Y) = 30^2)$. Seja um modelo hipotético cujo único parâmetro é $\theta = E(Y)$. Desconhecendo-se $E(Y)$, θ seria estimado por $\hat{\theta} = \bar{y} = \frac{1}{n} \sum y_i$. É fácil mostrar que $\hat{\theta} \sim N(E(\hat{\theta}) = 50, \text{var}(\hat{\theta}) = 30^2/n)$, onde n é o número de anos do registro histórico. Portanto, $(\hat{\theta} - 50) \sqrt{n}/30$ tem distribuição normal padrão. Na prática, não se conhecem $E(Y)$ (caso contrário, não haveria incerteza quanto ao parâmetro θ) nem $\text{var}(Y)$. Nesta circunstância, $(\hat{\theta} - \bar{y}) \sqrt{n}/s$, onde s é o desvio padrão amostral, tem distribuição t-student. Obtém-se um θ_i sorteando-se uma variável t_i e resolvendo-se a equação $\theta_i = \bar{y} + t_i s / \sqrt{n}$.

Raramente a caracterização de $F_{\theta}(\cdot)$ é tão trivial quanto no exemplo acima. Quase sempre procedimentos mais sofisticados seriam necessários. Uma maneira de evitar essa complicação consiste em sortear θ_i com a ajuda do modelo estocástico, de acordo com o algoritmo:

Ajuste o modelo estocástico à série histórica, obtendo θ_h

Repita $i = 1, a$

Gere uma série sintética de comprimento igual ao da série histórica, dado θ_h . Essa será uma série "filha"

Ajuste o modelo à série filha. Obtenha θ_i

Repita $j = 1, b$

Gere uma série sintética de comprimento não necessariamente igual ao da série histórica. Essa será uma série "neta"

Com essa abordagem, a incerteza quanto aos parâmetros de população é traduzida pela variação nos parâmetros das séries filhas, uma vez que cada série filha tem o mesmo comprimento da série histórica.

As equações de estimativa dos parâmetros serão discutidas nos itens 4.4, 4.5 e 4.6, para cada modelo específico. O método dos momentos será utilizado preponderantemente, por ser, em geral, mais simples. Entretanto, sempre que conveniente, referências serão feitas ao método da máxima verossimilhança.

Convém que o leitor tenha em mente que a incerteza quanto à escolha do modelo e a estimativa dos parâmetros é mais regra do que exceção em hidrologia. Por exemplo, no item 4.2 a solução de diversos problemas pelo método da Cadeia de Markov passou pela definição do processo estocástico afliência anual, que, na prática, também deve ser inferido a partir da série histórica.

Verificação do modelo

É necessário confirmar que as hipóteses subjacentes ao modelo estocástico passível de adoção sejam satisfeitas, antes que efetivamente se utilize o modelo. Por exemplo, suponha que o modelo 2 seja o candidato. Existem duas hipóteses associadas a ele:

- a) Z_t tem distribuição normal padrão. Para testar esta hipótese, é preciso, antes de mais nada, obter um conjunto de $n-1$ observações (z_2, z_3, \dots, z_n), através da solução da equação do modelo para Z :

$$Z_{t+1} = [y_{t+1} - \mu - \rho(y_t - \mu)] / (\sigma \sqrt{1 - \rho^2})$$

O teste de hipótese propriamente dito pode ser feito de diversas maneiras, como, por exemplo, através do teste de qui-quadrado ou de Kolmogorov-Smirnov, embora este não seja rigorosamente apropriado, visto que a distribuição "teórica" é, na realidade, ajustada aos dados, (a média e a variância amostrais de Z são, por construção, idênticamente iguais a 0 e 1).

- b) Z_i independente de Z_j , para $i \neq j$. Para testar esta hipótese, é necessário que se calculem os coeficientes de correlação amostrais $\text{corr}(Z_t, Z_{t+k})$. O gráfico de $\text{corr}(Z_t, Z_{t+k})$ contra k é chamado de correlograma (item 4.4). Se a hipótese de independência estiver correta, o correlograma tenderá a ser uma linha horizontal, perto de zero.

O teste de hipótese propriamente dito pode ser feito de diversas maneiras, como, por exemplo, através do teste de Portmanteau (item 4.4).

Se qualquer das hipóteses subjacentes ao modelo puder ser rejeitada, o modelo deverá também ser rejeitado.

Validação do modelo

A confiança que se possa ter num modelo estocástico de vazões depende da capacidade que ele tenha de preservar, nas séries sintéticas, algumas propriedades observadas na série histórica. Uma propriedade pode ser constituída por um único valor, como, por exemplo, a vazão média observada em n anos. Ou pode ser um conjunto de valores, como, por exemplo, a distribuição empírica de probabilidades da máxima vazão observada em cada ano. Diz-se que um modelo preserva uma determinada propriedade quando não se pode distinguir estatisticamente a série histórica da sintética, com base nas observações desta propriedade nas duas séries.

Quando alguma propriedade é utilizada para a determinação de um parâmetro do modelo, esta propriedade é automaticamente preservada, por construção. Nesse caso, comparar a observação da propriedade da série sintética com seu correspondente da série histórica serve apenas para verificar a adequação do programa de computador utilizado, e não para validar o modelo (Kelman e Pereira, 1977, Stedinger e Taylor, 1982).

A seleção do conjunto de propriedade em que se vai centrar atenção depende da aplicação que se pretenda dar às séries sintéticas. Por exemplo, quando o planejamento da expansão do sistema hidrelétrico é feito com auxílio de séries sintéticas mensais, a atenção deve estar voltada primordialmente para a preservação dos volumes afluentes durante as secas (Kelman e Pereira, 1977). Quando o cálculo do volume de amortecimento de cheias é feito com o auxílio de séries sintéticas diárias, a atenção deve estar voltada para a preservação dos picos e volumes das cheias excepcionais.

Suponha que, a partir da série histórica e da série sintética produzida por algum modelo estocástico, seja possível extrair amostras aleatórias de alguma variável cujas propriedades haja interesse em preservar, respectivamente $\{X_h(1), X_h(2), \dots, X_h(j)\}$ e $\{x_s(1), x_s(2), \dots, x_s(k)\}$. Como a série sintética é, em geral, mais longa do que a histórica, é de se supor que $k > j$. Por exemplo, se X for a máxima vazão diária num período de um ano, e admitindo-se que o modelo de vazões diárias sob escrutínio não use esta variável como parâmetro, j e k serão, respectivamente, o número de anos na série

histórica e sintética. Verifica-se a validação do modelo testando-se a hipótese de que ambas as amostras tenham sido sorteadas de uma mesma população. Se a hipótese nula for rejeitada, o modelo deverá também ser rejeitado. Como, em princípio, não se sabe qual a forma da distribuição de probabilidades de X , os métodos não paramétricos são particularmente indicados (Bradley, 1968). Por exemplo, o teste de Smirnov pede pelo cálculo da maior diferença entre as distribuições empíricas provenientes das duas amostras:

$$d = \max_X |n_h(x) / j - n_s(x) / k|$$

onde $n_h(x)$ e $n_s(x)$ são respectivamente iguais ao número de elementos na amostra histórica e sintética menores ou iguais a x . O modelo deverá ser rejeitado com um nível de confiança de 95%, caso $d > 1,358 \sqrt{(j+k) / (jk)}$.

Suponha que, a partir da série histórica, seja possível extrair-se apenas uma observação da variável cujas propriedades haja interesse em preservar, x_h . Por exemplo, X pode ser o tamanho do reservatório capaz de garantir um determinado deflúvio anual durante n anos (duração da série histórica). O modelo estocástico candidato pode ser utilizado para produzir r séries sintéticas, cada uma delas também de comprimento n , resultando na amostra $\{x_s=(1), x_s=(2), \dots, x_s=(r)\}$ (Figura 115). Uma proporção muito grande ou muito pequena de elementos do conjunto que sejam menores do que x_h indica que a série histórica é atípica para o modelo considerado, o que sugere a rejeição do modelo. Mais formalmente, o que se está estimando é a probabilidade $\beta = P(X = < x_h)$ (Figura 115).

O leitor notará um paradoxo neste procedimento: um dos objetivos da hidrologia estocástica é o de estimar a probabilidade β . No entanto, quando a estimativa for menor do que 0,05, ou maior do que 0,95, que são valores legítimos, o teste de validação indicará a rejeição do modelo. A analogia com o conceito de erro tipo I em testes de hipóteses é evidente: em 10% das vezes em que o modelo for adequado, o teste de validação recomendará a rejeição. É exatamente nestes 10% de casos que o uso do modelo estocástico tem mais chances de causar importantes conseqüências no processo decisório. Por isso, é conveniente que se avalie um modelo estocástico em relação a diversas variáveis aleatórias e não apenas especificamente em relação à variável que fornece resposta a um problema de engenharia. É com base no conjunto dos resultados dos testes de validação que se deve tomar uma decisão quanto à rejeição ou não do modelo estocástico.

Uma variante do procedimento de validação foi sugerida por Thompstone *et al* (1987), que recomendaram a adoção de amostras decompostas. Nessa abordagem, ajusta-se o modelo estocástico apenas aos primeiros n_1 anos da série histórica, reservando-se os n_2 anos remanescentes ($n_1+n_2 = n$) para o cálculo de x_h . O teste é feito com r séries sintéticas de comprimento igual a n_2 anos. Essa variante resulta em resultados mais confiáveis do que os obtidos ajustando-se o modelo estocástico à totalidade da série histórica.

4. MODELAGEM DE VAZÕES ANUAIS

4.1 Estacionariedade

Os estudos do sistema de recursos hídricos não utilizam, em geral, intervalos de discretização anual. No entanto, é indispensável que se aborde a questão da modelagem de vazões anuais, tanto por fornecer uma base teórica sobre a qual se irá construir as demais modelagens, como porque uma das alternativas de se produzir séries sintéticas mensais necessita que já estejam disponíveis as séries sintéticas anuais.

A Figura 116 mostra a hidrógrafa do rio Paraíba do Sul em Barra do Piraf com intervalo de discretização anual. Como se vê, as vazões anuais (na realidade, vazão média ao longo de um ano) flutuam entre 177 e 515m³/s. Diversos pesquisadores têm procurado descobrir, nessas flutuações, regularidades que teriam origem nas relações entre as variáveis climático-hidrológicas e as atividades do Sol. A ocorrência das manchas solares em intervalos regulares permitiria prever, segundo esses pesquisadores, ciclos de anos secos e anos úmidos. Outras tentativas de se descobrirem periodicidades nas vazões anuais estiveram em voga na primeira metade do século, quando pesquisadores procuravam ajustar as séries hidrológicas observadas com o emprego das Séries de Fourier. Esta técnica consiste em se ajustar uma série temporal por uma soma de funções periódicas (senos e co-senos) de amplitude e frequência distintas. Para uma particular série hidrológica, algumas dessas funções periódicas serão mais relevantes do que outras para a obtenção de um bom ajuste. As correspondentes frequências são chamadas de periodicidades escondidas, uma vez que não foram identificadas causas físicas que as expliquem. A técnica de se fazerem previsões a partir da extrapolação desses sinais periódicos foi bastante utilizada. Porém, em apenas alguns casos essas previsões têm significado estatístico. Em geral, o método carece de consistência teórica (Yevjevich, 1979).

Ocorreram no passado notáveis variações climáticas, conforme o atestam as diversas eras glaciais. Existem fatores astronômicos causadores de alterações na distribuição de energia solar sobre a superfície da Terra, com periodicidades conhecidas. Só que essas variações se processam numa enorme escala de tempo, absolutamente irrelevante para os estudos de planejamento que se possa fazer para os próximos 50 anos, baseados em dados coletados nos últimos 50 anos (Yevjevich, 1979). Neste capítulo, a partir de agora, a série de vazões anuais será considerada estacionária (as leis probabilísticas que governam o processo não se alteram com o tempo), ressalvada a intervenção humana sobre a Natureza.

Uma série anual pode ser não estacionária devido à alteração no uso da terra, a modificações na seção transversal ou mudanças de régua. A melhor maneira de se determinar se uma série é ou não estacionária é saber da eventual existência de algum desses fatores. Caso isso não seja possível, é preciso examinar a série histórica em busca de algum indício. A observação da Figura 116 não revela irregularidade alguma para a série de vazões anuais em Barra do Piraf. Essa avaliação subjetiva é, em geral, suficiente. Uma confirmação aproximada pode, no entanto, ser obtida da seguinte maneira: divide-se a série anual em duas subséries (Quadro 49). Para cada subsérie, computam-se o valor de média e o desvio padrão amostral, μ_j e σ_j , $j = 1, 2$. Em seguida, aplica-se o teste t de Student para verificar a hipótese de igualdade entre as duas médias (ver, por exemplo, Gomide, 1976).

Exemplo 15

Para o caso do rio Paraíba do Sul, pode-se calcular a estatística t como:

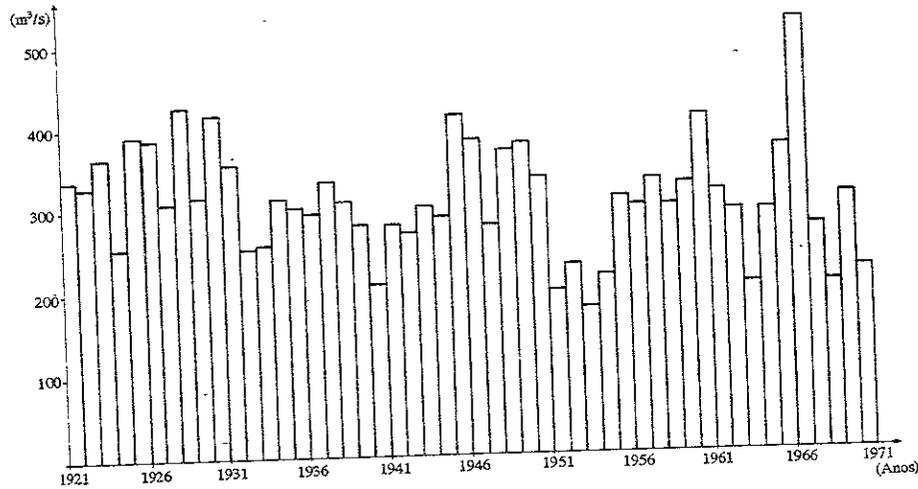


Figura 116. Hidrógrafo de vazões médias anuais do Rio Paraíba do Sul em Barra do Piauí (1921 - 1971)

$$t = \frac{(\hat{\mu}_1 - \hat{\mu}_2) \sqrt{\frac{m}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} = 0,62$$

onde m é o tamanho de cada subamostra (no caso $m = 25$). Como o valor crítico, para o nível de significância de 95% (48 graus de liberdade), é de 2,00, não se pode rejeitar a hipótese da igualdade das duas médias. Um teste análogo poderia ser feito para as duas variâncias, $\hat{\sigma}_1^2$ e $\hat{\sigma}_2^2$, empregando-se a estatística F. Rigorosamente, tanto o teste t quanto o teste F são aplicáveis apenas quando as variáveis sob escrutínio, no caso vazões médias anuais, são normalmente distribuídas. No entanto, os testes, em geral, são passíveis de aplicação, mesmo quando a condição de normalidade não for atendida. Alternativamente, podem-se aplicar testes não paramétricos (Bradley, 1968), para os quais exigência alguma é feita com relação à distribuição de probabilidades.

Distribuição Marginal de Probabilidades

As vazões do rio Paraíba do Sul podem ser ordenadas de forma a produzir a distribuição acumulada empírica, mostrada na Figura 116. Para a i -ésima vazão q_i (em ordem crescente), associa-se a ordenada dada pela posição de locação. No exemplo, adotou-se a posição $i/n+1$, mas outras opções poderia ser selecionadas. Um primeiro gerador de séries sintéticas de vazões anuais poderia ser concebido, partindo-se do seguinte teorema: se Q é uma variável aleatória qualquer e $F_Q(q)$ sua distribuição de probabilidades acumulada, $F_Q(q) = \text{Prob}(Q < q)$, então $U = F_Q(Q)$ tem distribuição de probabilidades uniforme (convém lembrar que se adota a convenção de utilizar letras maiúsculas para denominar variáveis aleatórias e minúsculas para denominar as observações das variáveis aleatórias). Números aleatórios com distribuição uniforme seriam sorteados, U_1, U_2, U_3, \dots . Para cada valor u_j se associa uma vazão anual q_j , conforme Figura 117, formando-se assim a série q_1, q_2, q_3, \dots , que poderá ser tão longa quanto se queira. Alguma hipótese será necessária para extrapolar a distribuição acumulada empírica, uma vez que sempre é possível sortear-se um valor para U maior do que $n/n+1$. Alternativamente, pode-se ajustar uma expressão analítica à distribuição acumulada empírica, garantindo-se uma extrapolação automática. Isso equivale a se adotar uma particular distribuição de probabilidades para vazão anual. Diversas distribuições têm sido sugeridas na literatura, sem que nenhuma possa ser indicada para aplicação generalizada. Markovic (1965) fez um estudo com 446 séries de vazões anuais, comparando as diversas distribuições com base na qualidade de ajuste, avaliada pelo qui-quadrado. Sua conclusão foi de que "todas as distribuições são aplicáveis e nenhuma tem ajuste melhor do que outra em relação à distribuição empírica de precipitação anual ou vazão anual". O Quadro 50 apresenta as principais propriedades das distribuições mais utilizadas.

A distribuição normal é freqüentemente adotada por ser bastante conhecida e porque a vazão anual resulta do somatório de um grande número de variáveis aleatórias (365 vazões diárias). O teorema do limite central garante que, sob condições bastante gerais, a soma de um grande número de variáveis aleatórias independentes e igualmente dispostas, tem distribuição normal, não importa qual seja a distribuição de probabilidades de cada parcela. Na realidade, a aplicabilidade do teorema do limite central para o caso em pauta é bastante questionável, visto que as parcelas não são

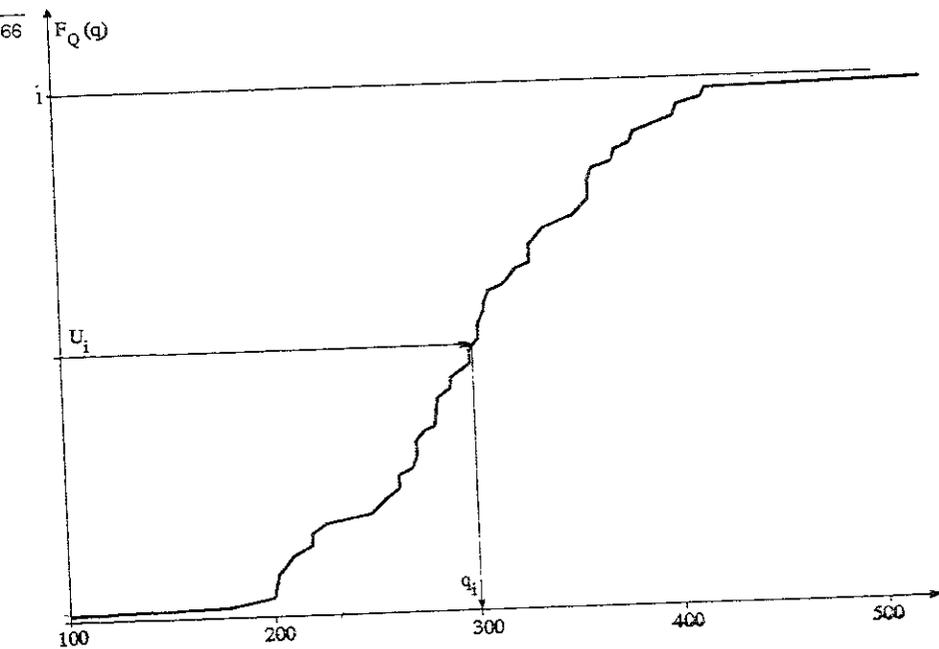


Figura 117. Distribuição acumulada empírica das vazões anuais do Rio Paraíba do Sul em Barra do Piauí

Quadro 49. Série de vazões anuais do rio Paraíba do Sul, em Barra do Piraí, RJ.

Subsérie 1		Subsérie 2		Observações	
Ano	Vazão (m ³ /s)	Ano	Vazão (m ³ /s)		
21 - 22	336	46 - 47	406	1) Média amostral $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n q_i$	
22 - 23	329	47 - 48	380		
23 - 24	361	48 - 49	271		
24 - 25	247	49 - 50	361		
25 - 26	387	50 - 51	376		
26 - 27	385	51 - 52	331		
27 - 28	306	52 - 53	197		
28 - 29	421	53 - 54	223		
29 - 30	313	54 - 55	177		
30 - 31	416	55 - 56	214		
31 - 32	355	56 - 57	303	2) Desvio padrão amostral. $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n (q_i - \hat{\mu})^2}$	
32 - 33	253	57 - 58	293		
33 - 34	258	58 - 59	319		
34 - 35	306	59 - 60	282		
35 - 36	299	60 - 61	401		
36 - 37	292	61 - 62	308		
37 - 38	329	62 - 63	286		
38 - 39	304	63 - 64	201		
39 - 40	276	64 - 65	285		
40 - 41	206	65 - 66	362		
41 - 42	274	66 - 67	515	3) Coeficiente de variação amostral $\hat{C}_v = \frac{\hat{\sigma}}{\hat{\mu}}$	
42 - 43	265	67 - 68	265		
43 - 44	298	68 - 69	200		
44 - 45	273	69 - 70	303		
45 - 46	284	70 - 71	216		
					4) Coeficiente de assimetria amostral $\hat{\gamma} = \frac{1}{\hat{\sigma}^3} \cdot \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (q_i - \hat{\mu})^3$
				5) Coeficiente de autocorrelação amostral $\hat{\rho} = \frac{\sum_{i=1}^n (q_i - \hat{\mu})(q_{i+1} - \hat{\mu})}{n \sum_{i=1}^n (q_i - \hat{\mu})^2}$	

n = número de anos

	Subsérie 1	Subsérie 2	Série completa
μ	311	299	305
σ	53	79	67
C_v	0,17	0,26	0,22
γ	0,45	0,63	0,49
ρ	0,22	0,29	0,26

Quadro 50. Distribuição de probabilidades

- Normal (N) $-\infty < q < +\infty$
 $f_Q(q) = \frac{1}{\sqrt{2\pi} \sigma} \cdot \exp \left[-\frac{1}{2} \left(\frac{q - \mu}{\sigma} \right)^2 \right]$
 média = μ
 variância = σ^2
 assimetria = 0
- Log-Normal de Dois Parâmetros (LN2), $q > 0$
 $f_Q(q) = \frac{1}{q \sqrt{2\pi} \sigma_n} \cdot \exp \left[-\frac{1}{2} \left(\frac{\ln(q) - \mu_n}{\sigma_n} \right)^2 \right]$
 média = $\exp \left[\frac{\sigma_n^2}{2} + \mu_n \right] = \mu$
 variância = $\exp [2(\sigma_n^2 + \mu_n)] - \exp [\sigma_n^2 + 2\mu_n] = \sigma^2$
 assimetria = $\left(\frac{\sigma}{n}\right)^3 + 3\left(\frac{\sigma}{n}\right) = \gamma$
 onde
 $\mu_n = E(\ln(q))$
 $\sigma_n^2 = E[(\ln(q) - E(\ln(q)))^2]$
- Log-Normal de Três Parâmetros (LN3), $q > a$
 $f_Q(q) = \frac{1}{(q-a) \sqrt{2\pi} \sigma_n} \exp \left[-\frac{1}{2} \left(\frac{\ln(q-a) - \mu_n}{\sigma_n} \right)^2 \right]$
 média = $a + \exp \left[\frac{\sigma_n^2}{2} + \mu_n \right] = \mu$
 variância = $\exp [2(\sigma_n^2 + \mu_n)] - \exp [\sigma_n^2 + 2\mu_n] = \sigma^2$
 assimetria = $\frac{(\exp 3(\sigma_n^2) - 3 \exp(\sigma_n^2) + 2)}{(\exp(\sigma_n^2) - 1)^{3/2}} = \gamma$
 onde
 $\mu_n = E[\ln(q-a)]$
 $\sigma_n^2 = E[(\ln(q-a) - \mu_n)^2]$
- Gama de Três Parâmetros (G3), $q > a$
 $f_Q(q) = \frac{\lambda^\beta}{\Gamma(\beta)} (x-a)^{\beta-1} \exp(-\lambda(x-a))$
 média = $a + \beta/\lambda = \mu$
 variância = $\beta/\lambda^2 = \sigma^2$
 assimetria = $2/\sqrt{\beta} = \gamma$

independentes entre si, nem são igualmente distribuídas. A distribuição normal apresenta o inconveniente de ser definida para valores negativos de q ; ou seja, existe uma probabilidade de que seja gerada uma vazão anual negativa. No caso do rio Paraíba do Sul, essa probabilidade é menor do que 0,000005; ou seja, pode-se esperar apenas um valor negativo em 200 mil anos. Os parâmetros μ e σ , que aparecem no Quadro 50 são estimados respectivamente pela média e desvio padrão amostrais $\hat{\mu}$ e $\hat{\sigma}$, cujas expressões aparecem no Quadro 49. O sorteio de números aleatórios normalmente distribuídos é tarefa simples e usualmente procedimentos para esse fim estão disponíveis em qualquer sistema computacional.

A série sintética $q^*_1, q^*_2, q^*_3, \dots, q^*_m$, assim obtida, será tal que $\lim_{m \rightarrow \infty} \hat{\mu} = \mu$ e $\lim_{m \rightarrow \infty} \hat{\sigma} = \sigma$. Isto é, a média e o desvio padrão amostrais da série sintética tenderão aos correspondentes valores da série histórica, à medida que o tamanho da série sintética cresça. Diz-se que esse modelo preserva esses dois momentos. Conforme mencionado no item 4.3, isso não significa que o modelo irá reproduzir a média e o desvio padrão populacionais, μ e σ . Na realidade, esses valores serão sempre desconhecidos.

A distribuição log-normal de dois parâmetros é definida unicamente para valores positivos de q . Portanto, é impossível gerar uma vazão anual negativa. Se Q tem distribuição log-normal de dois parâmetros, então $\log Q$ tem distribuição normal; ou seja, a distribuição log-normal de dois parâmetros é indicada para a hipótese de que a vazão anual seja o resultado de um grande número de fatores que interagem de forma multiplicativa. Uma possível maneira de se estimarem os parâmetros dessa distribuição seria repetir o procedimento adotado no caso da distribuição normal para a série $l_n, q_1, l_n, q_2, l_n, q_3, \dots$. Dessa forma, os parâmetros são estimados como

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n l_n q_t \quad (4.1)$$

$$\hat{\sigma}_n = \sqrt{\frac{1}{n} \sum_{t=1}^n (l_n q_t - \hat{\mu}_n)^2} \quad (4.2)$$

Números aleatórios normalmente distribuídos seriam, neste caso, gerados com média e desvio padrão respectivamente iguais a $\hat{\mu}_n$ e $\hat{\sigma}_n$. $l_n q^*_1, l_n q^*_2, \dots$, que, quando exponenciados, resultariam na série sintética $q^*_1, q^*_2, q^*_3, \dots$. Ocorre que, no caso em pauta, $\lim_{m \rightarrow \infty} \hat{\mu} \neq \mu$ e $\lim_{m \rightarrow \infty} \hat{\sigma} \neq \sigma$ ou seja, a preservação dos dois primeiros momentos na série dos logaritmos não garante a preservação desses momentos na série original. As relações entre esses dois conjuntos de parâmetros são expressas pelas equações que aparecem no Quadro 50, que, quando explicitadas para μ_n e σ_n , resultam nas seguintes equações:

$$\sigma_n^2 = \ln(\psi) \quad (4.3)$$

$$\mu_n = \frac{1}{2} \ln(\sigma^2 / (\psi^2 - \psi)) \quad (4.4)$$

$$\psi = 1 + (\sigma/\mu)^2 \quad (4.5)$$

Usualmente, o objetivo é preservar os momentos da série histórica e não os dos logaritmos. Neste caso, é necessário resolver o sistema de equações (4.3-4.5), substituindo-se os valores de μ e σ , respectivamente, por $\hat{\mu}$ e $\hat{\sigma}$. É comum, no entanto, que ambas as abordagens de estimativa de μ_n e σ_n resultem em valores numéricos muito próximos. Além disso, Stedinger (1981) demonstrou que em alguns casos, é preferível preservar os momentos da série de logaritmos para aproximar os momentos de população (desconhecidos) das vazões.

Exemplo 16

No caso do rio Paraíba do Sul, em Barra do Piraí, o emprego das equações (4.3-4.5) resulta em $\mu_n = 5.6952$ e $\sigma_n = 0.2271$, ao passo que a estimativa a partir da série dos logaritmos, em $\mu_n = 5.6952$ e $\sigma_n = 0.2235$.

A Figura 118 foi apresentada por Klemes (1978), com base em dados coletados por Yevjevich (1963). Ela mostra a relação entre os coeficientes de assimetria (γ) e de variação (σ/μ) para 140 rios espalhados pelo mundo. Como se pode notar, a maior parte dos pontos fica localizada entre o eixo vertical (assimetria nula; possivelmente distribuição normal) e a reta $\gamma = (\sigma/\mu)^3 + 3(\sigma/\mu)$, que é a relação pertinente à distribuição log-normal de dois parâmetros. Uma possível explicação para esse fato foi dada originalmente por Chow (1954) e repetida por Klemes (1978). Para esses autores, os fenômenos hidrológicos podem ser interpretados como sendo o resultado de um conjunto não enumerável de causas físicas que interagem tanto de forma multiplicativa como aditiva. Uma vez que os efeitos multiplicativos estão associados à distribuição log-normal de dois parâmetros (LN2) e os efeitos aditivos estão associados à distribuição normal (N), pode-se esperar que a verdadeira distribuição esteja situada entre os dois tipos limites (N e LN2). A distribuição log-normal de três parâmetros (LN3) (Quadro 50) pode satisfazer essa propriedade: quando o parâmetro "a" (deslocamento) tende para $-\infty$, LN3 tende para N; quando a = 0, a distribuição LN3 equivale a LN2.

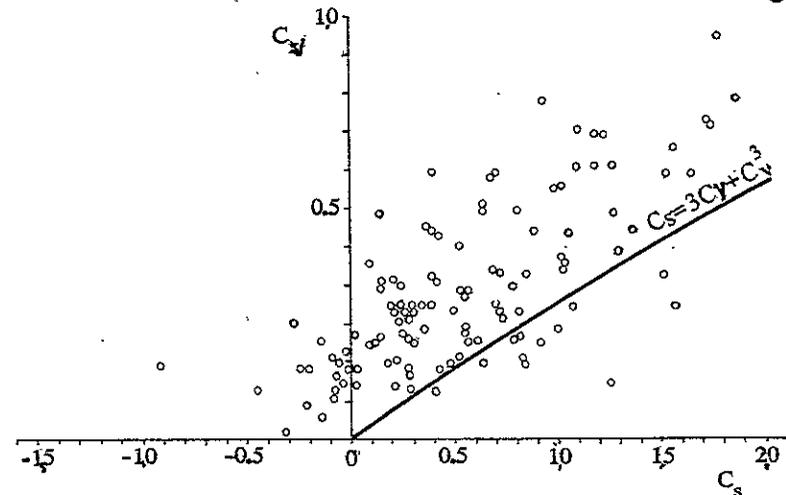


Figura 118. Relação entre o coeficiente de variação e o coeficiente de assimetria para séries de vazões anuais de 140 rios de todo o mundo

A distribuição LN3 é definida para valores de $q > a$. Se $a < 0$, existe a probabilidade de que se obtenha alguma vazão anual negativa. A estimativa dos parâmetros pode ser feita tanto com base em valores amostrais da média, mediana e desvio padrão, como em média, desvio padrão e assimetria. Burges, Lettenmaier e Bates (1975) demonstraram a superioridade da segunda alternativa. Portanto, se o propósito é preservar os momentos da série histórica, e não os de séries de logaritmos, os três parâmetros, μ_n , σ_n e a , devem ser estimados resolvendo-se as equações abaixo, onde μ , σ e γ devem ser substituídos pelos respectivos valores amostrais, $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\gamma}$

$$\delta = \left[\left(1 + \frac{\gamma^2}{2} \right) + \left(\gamma^2 + \frac{\gamma^4}{4} \right)^{1/2} \right]^{1/3} + \left[\left(1 + \frac{\gamma^2}{2} \right) - \left(\gamma^2 + \frac{\gamma^4}{4} \right)^{1/2} \right]^{1/3} - 1 \quad (4.6)$$

onde

$$\gamma > 0$$

$$\sigma_n^2 = L_n(\delta) \quad (4.7)$$

$$\mu_n = \frac{1}{2} L_n(\sigma^2/(\delta^2 - \delta)) \quad (4.8)$$

$$a = \mu - [\sigma^2/(\delta - 1)]^{1/2} \quad (4.9)$$

Uma série sintética com distribuição LN3 é gerada obtendo-se, primeiro, uma série com distribuição normal e, com média e desvio padrão respectivamente iguais a μ_n e σ_n . A série normal é, então, exponenciada e somada ao parâmetro a .

Na União Soviética utiliza-se a distribuição gama de três parâmetros (G3) que pode ser justificada com base num estudo chuva-deflúvio (Klimes, 1978). Os parâmetros são estimados resolvendo-se as seguintes equações, onde μ , σ e γ devem ser substituídos pelos respectivos valores amostrais, $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\gamma}$.

$$\beta = 4/\gamma^2 \quad (4.10)$$

$$\lambda = 2/(\gamma\sigma) \quad (4.11)$$

$$a = \mu - 2\sigma/\gamma \quad (4.12)$$

A geração de uma série sintética com distribuição próxima da G3 pode ser conseguida a partir da transformação de Wilson-Hilferty válida quando $|\gamma| \leq 3$:

$$Q = \mu + \sigma \left(\frac{2}{\gamma} \left(1 - \frac{\gamma^2}{36} + \frac{\gamma W}{6} \right)^3 - \frac{2}{\gamma} \right) \quad (4.13)$$

onde W é uma variável aleatória normal padrão.

A Figura 119 mostra as quatro distribuições acumuladas de probabilidades, quando ajustadas para as vazões do rio Paraíba do Sul. Como se vê, todas elas parecem aderir convenientemente à distribuição acumulada empírica. Nenhuma distribuição deve ser considerada para modelagem se, nessa fase, não for aprovada em um teste de adequação de ajuste.

4.3 Persistência

A Figura 120 mostra o gráfico dos pares de pontos (q_t, q_{t+1}) do rio Paraíba do Sul, onde q_t é a vazão anual do ano t . Pode-se perceber que, se o ano t foi seco (q_t pequeno), o ano $t+1$ tenderá também a ser seco. É vice-versa. A esse fenômeno dá-se o nome de persistência ou de dependência temporal. O modelo visto no item anterior não poderia ser adequado porque adota a hipótese de aleatoriedade para a série anual do rio Paraíba do Sul. Isto é, adota a hipótese de que a distribuição de probabilidade de Q_{t+1} independente dos valores observados no passado ($q_t, q_{t-1}, q_{t-2}, \dots$). Na realidade, a Figura 120 mostra que talvez essa hipótese seja falsa, uma vez que Q_{t+1} parece depender pelo menos de q_t . Com base nessa constatação, pode-se sugerir o seguinte processo estocástico para modelar as séries anuais:

$$Q_{t+1} = \mu + \rho (Q_t - \mu) + \sigma \sqrt{1 - \rho^2} Z_{t+1} \quad (4.14)$$

onde

$$E(Z_{t+1}) = 0, \text{ var}(Z_{t+1}) = 1;$$

Q_t e Z_{t+1} são independentes e

Z_t e Z_τ são independentes, $t \neq \tau$

aplicação do operador $E(\cdot)$, valor esperado, em ambos os lados da equação (4.14) resulta em

$$E(Q_{t+1}) = \mu + \rho (E(Q_t) - \mu) + \sigma \sqrt{1 - \rho^2} E(Z_{t+1}) = \rho E(Q_t) + \mu(1 - \rho) \quad (4.15)$$

como o processo é estacionário, $E(Q_{t+1}) = E(Q_t)$ e, portanto,

$$E(Q_{t+1}) = \mu \quad (4.16)$$

Analogamente, a aplicação do operador $\text{var}(\cdot)$, variância, resulta em

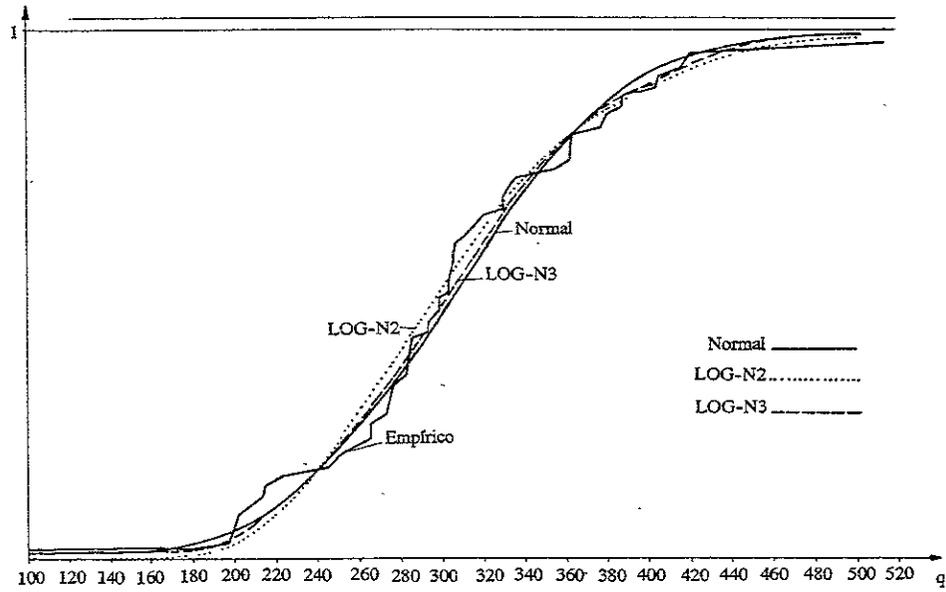


Figura 119. Ajuste das distribuições acumuladas teóricas para vazões anuais do Rio Paraíba do Sul em Barra do Piauí

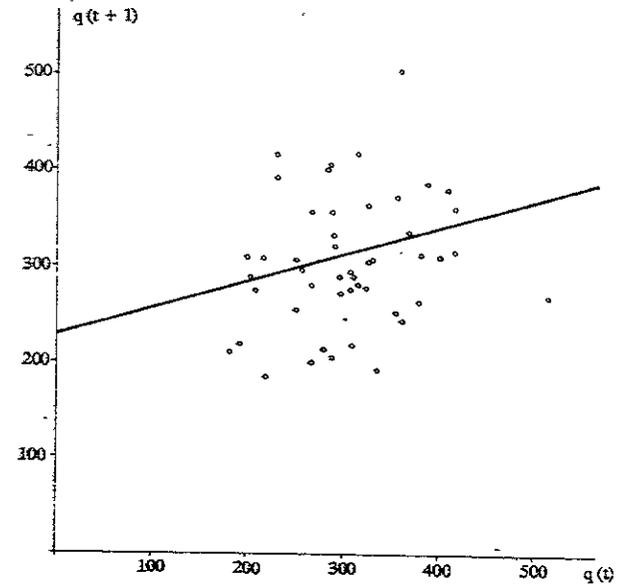


Figura 120. Relação entre vazões de anos consecutivos para o Rio Paraíba do Sul em Barra do Piauí

$$\text{var}(Q_{t+1}) = \rho^2 \text{var}(Q_t) + \sigma^2 (1-\rho^2) \text{var}(Z_{t+1}) + 2\sigma\rho\sqrt{1-\rho^2} \text{cov}(Q_t, Z_{t+1}) \quad (4.17)$$

onde $\text{cov}(Q_t, Z_{t+1})$ significa covariância entre Q_t e Z_{t+1} , que é nula porque essas duas variáveis aleatórias são independentes. Portanto, lançando-se mão de $\text{var}(Q_{t+1}) = \text{var}(Q_t)$, tem-se que

$$\text{var}(Q_{t+1}) = \sigma^2 \quad (4.18)$$

Por um processo semelhante, seria possível derivar a relação entre a assimetria de Q e a assimetria de Z , respectivamente γ e γ_Z

$$\gamma_Z = \gamma \frac{(1-\rho^2)^{3/2}}{(1-\rho^2)} \quad (4.19)$$

A regressão de Q_{t+1} em q_t , isto é, a expressão do valor esperado da vazão no ano $t+1$, conhecida a vazão no ano t , é dada por

$$E(Q_{t+1} | Q_t = q_t) = E(Q_{t+1} | q_t) = \mu + \rho(q_t - \mu) + \sigma\sqrt{1-\rho^2} E(Z_{t+1}) = \mu + \rho(q_t - \mu) \quad (4.20)$$

A variância condicionada, às vezes também chamada de variância não explicada pela regressão ou variância residual, é dada por:

$$\text{var}(Q_{t+1} | q_t) = \sigma^2 (1-\rho^2) \quad (4.21)$$

Multiplicando ambos os lados da equação por Q_t e aplicando-se o operador $E(\cdot)$, resulta

$$E(Q_{t+1} Q_t) = \mu E(Q_t) + \rho(E(Q_t^2) - \mu E(Q_t)) + \sigma\sqrt{1-\rho^2} E(Z_{t+1} Q_t) \quad (4.22)$$

Utilizando-se $\text{var}(Q_t) = E(Q_t^2) - E^2(Q_t)$, as relações (4.16) e (4.18) e a independência entre Q_t e Z_{t+1} , resulta

$$E(Q_{t+1} Q_t) = \mu^2 + \rho(\sigma^2 + \mu^2 - \mu^2) + \sigma\sqrt{1-\rho^2} \mu E(Z_{t+1}) = \mu^2 + \rho\sigma^2 \quad (4.23)$$

Logo, a correlação entre Q_{t+1} e Q_t , também chamada de autocorrelação, é dada por

$$\text{corr}(Q_{t+1}, Q_t) = \frac{E(Q_{t+1} Q_t) - E(Q_t) E(Q_{t+1})}{\sqrt{\text{var}(Q_t) \text{var}(Q_{t+1})}} = \frac{\mu^2 + \rho\sigma^2 - \mu^2}{\sigma^2} = \rho \quad (4.24)$$

O modelo expresso pela equação (4.14) implica que a vazão, num ano qualquer Q_{t+1} , é uma variável aleatória cuja distribuição de probabilidade marginal tem média μ e desvio padrão σ (ambos valores de população). Entretanto, quando a vazão, no ano imediatamente anterior q_t , é conhecida, então Q_{t+1} continua sendo uma variável aleatória, só que com média situada na reta que aparece na Figura 117 e desvio padrão (medida de dispersão) menor, dado por $\sigma\sqrt{1-\rho^2}$ (o coeficiente de autocorrelação situa-se entre -1 e $+1$). Isto é, a distribuição de probabilidades condicionada tem média $\mu + \rho(q_t - \mu)$ e desvio padrão $\sigma\sqrt{1-\rho^2}$. Convém observar que, se $\rho = 0$, não há persistência alguma e o processo estocástico é, na realidade, independente, assunto já examinado. Por outro lado, se $\rho = 1$ a variância residual é nula, o que significa que todos os pontos na Figura 119 cairiam rigorosamente sobre a reta de regressão. Ou seja, a série anual q_1, q_2, q_3, \dots seria formada por valores iguais.

Os parâmetros do modelo μ, σ e ρ podem ser estimados a partir da média, desvio padrão e autocorrelações amostrais, $\hat{\mu}, \hat{\sigma}$ e $\hat{\rho}$, cujas expressões se encontram no Quadro 49. A geração de uma série sintética se faz da seguinte maneira (exemplo 12):

- escolha um valor para q_0 e faça $t = 0$
- $t = t+1$
- sorteie um número aleatório Z_{t+1} de média zero e desvio padrão unitário
- $q_{t+1} = \hat{\mu} + \hat{\rho}(q_t - \hat{\mu}) + \hat{\sigma}\sqrt{1-\hat{\rho}^2} z_{t+1}$
- vá para (b)

Se a distribuição de probabilidade de Z_{t+1} for normal padrão, então, tanto a distribuição marginal de Q_{t+1} como a distribuição condicionada de Q_{t+1} - conhecido q_t - serão também normais. Dessa forma, é possível ampliar um dos modelos descritos no item anterior - quando se assumia que a série histórica anual é uma realização de um processo aleatório, com distribuição marginal normal - para o modelo descrito acima, em que se conserva a mesma hipótese quanto à distribuição marginal, mas se inclui o efeito da persistência.

Quanto ao valor inicial q_0 , há duas possibilidades:

- caso se pretenda produzir diversas séries sintéticas para fazer simulações com relação ao futuro imediato, q_0 deve ser igual à última observação da série histórica;
- caso o estudo seja sobre o futuro longínquo, q_1 deverá ser sorteado de uma distribuição normal de média $\hat{\mu}$ e desvio padrão $\hat{\sigma}$, para cada série sintética.

Caso se pretenda adotar um modelo que considere a persistência e simultaneamente produza vazões com distribuição marginal de probabilidade LN3, modifica-se a equação (4.14) da seguinte forma:

$$\log(Q_{t+1} - a) = \mu_n + \rho_n (\log(Q_t - a) - \mu_n) + \sigma_n \sqrt{1-\rho_n^2} z_{t+1} \quad (4.25)$$

onde Z_{t+1} - variável aleatória com distribuição normal padrão.

A estimativa de ρ_n pode ser feita calculando-se o coeficiente de autocorrelação amostral, segundo expressão contida no Quadro 49, para a série $\{\log(q_1 - \hat{a}), \log(q_2 - \hat{a}), \dots, \log(q_m - \hat{a})\}$, sendo \hat{a} determinado pela resolução do sistema de equações 4.6 a 4.9. Neste caso, analogamente ao que foi visto no item anterior, $\lim_{m \rightarrow \infty} \rho^* \neq \rho$. Isto é, as séries sintéticas não preservarão o coeficiente de autocorrelação amostral da série histórica.

Alternativamente pode-se obter esta propriedade estimando-se ρ_n por

$$\rho_n = \frac{\ln(1 + \rho(\delta - 1))}{\ln \delta} \quad (4.26)$$

onde δ e ρ devem ser substituídos pelos respectivos valores amostrais e $\delta = \exp(\sigma_n^2)$. Caso se pretenda adotar um modelo que considere a persistência e simultaneamente produzir vazões com distribuição marginal de probabilidades que seja aproximadamente G3, utiliza-se a transformação de Wilson Hilferty - equação (4.13) - para gerar Z_{t+1} , a ser utilizado na equação (4.14):

$$Z = \left(\frac{2}{\gamma_z} \left(1 - \frac{\gamma_z^2}{36} + \frac{\gamma_z W}{6} \right)^3 - \frac{2}{\gamma_z} \right) \quad (4.27)$$

onde W é uma variável aleatória normal padrão e γ_z é dado pela equação (4.19). Um método mais preciso, porém mais trabalhoso, é apresentado por Obeysekera e Yevjevich (1986).

4.4 Transformação para normal

A preservação do coeficiente de autocorrelação amostral da série histórica não é necessariamente a melhor maneira de se conseguir uma série sintética com coeficiente de autocorrelação o mais próximo possível do valor de população. Em alguns casos, a tendenciosidade operacional é menor quando se preserva o coeficiente das vazões transformadas (Stedinger, 1981). Uma maneira de tomar partido deste fato e de simplificar a modelagem de vazões anuais consiste em inicialmente transformar a série q_1, q_2, \dots , numa série y_1, y_2, \dots com distribuição marginal normal padrão. Idealmente a transformação é dada por

$$y_t = \Phi^{-1}(F_Q(q_t)) \quad (4.28)$$

onde $\Phi(\cdot)$ - distribuição acumulada de probabilidades normal padrão

$F_Q(\cdot)$ - distribuição acumulada de probabilidades de Q

Na prática as transformações mais adotadas em hidrologia são as listadas no

Quadro 51. Transformações para uma variável normal padrão

Distribuição da vazão	y_t
Normal (dois parâmetros)	$(q_t - \mu)/\sigma$
Log-normal (dois parâmetros)	$(\ln q_t - \mu_n)/\sigma_n$
Log-normal (três parâmetros)	$(\ln(q_t - \hat{a}) - \mu_n)/\sigma_n$
Gama (três parâmetros) (*)	$6(\gamma/2(q_t - \mu)/\sigma + 2/\gamma)^{1/3} - 1 + \gamma^2/36/\gamma$
Potência de normal (quatro parâmetros)	$(w_t - \mu_b) / \sigma_b$

(*) transformação aproximada

A variável w_t , que aparece na última transformação do Quadro, é definida como

$$W_t = \lambda^{-1} ((q_t - \hat{a})^\lambda - 1), \quad \lambda \neq 0 \quad (4.29)$$

μ_b e σ_b são respectivamente, a média e o desvio padrão amostrais de w_t .

A equação acima é conhecida por transformação de Box e Cox (1964). Pode-se mostrar que w_t tende a $\ln(q_t - \hat{a})$ quando λ converge para zero. Isto é, a distribuição potência de normal pode ser vista como uma generalização da distribuição log-normal de três parâmetros.

A escolha da transformação não é uma tarefa simples. Uma recomendação de caráter geral é minimizar o número de parâmetros para evitar a perpetuação, nas séries sintéticas, de características da série histórica que não traduzam propriedades de população e sim variações amostrais. Nesse sentido, convém que se proceda de cima para baixo, na lista de transformações do Quadro 51. Inicialmente tenta-se a distribuição normal.

Se houver uma rejeição de algum teste de ajuste, tenta-se a log-normal de dois parâmetros. Se houver uma rejeição, tenta-se a log-normal de três parâmetros, e assim sucessivamente. Convém enfatizar, no entanto, que esta orientação de caráter geral deve subordinar às considerações sobre as propriedades da série sintética (ver "Escolha do modelo estocástico"), que podem indicar a preferência de uma particular transformação, mesmo quando existir uma transformação localizada mais alta na hierarquia do Quadro 51.

4.5 Modelos Arma

Considere-se a modelagem da persistência das vazões anuais, pelo processo Y_t , através da equação

$$Y_{t+1} = \Phi Y_t + V_{t+1} \quad (4.30)$$

onde Φ é o coeficiente de autocorrelação do processo Y_t e $V_{t+1} \sim N(0, 1 - \Phi^2)$; V_{t+1} independente de V_t, V_{t-1}, \dots , e de $Y_t, Y_{t-1}, Y_{t-2}, \dots$.

O modelo expresso pela equação (4.30) é chamado auto-regressivo linear de ordem 1, AR(1). O termo deriva do fato de que $E(Y_{t+1} | Y_t)$ é uma função linear de Y_t . Convém observar que o processo estocástico original Q_t não é necessariamente linear, a não ser que Q_t já seja normalmente distribuído. Por exemplo, se Q_t tiver distribuição LN2 e/ou LN3, $E(Q_{t+1} | Q_t = q_t)$ será uma função não-linear de q_t . Além disso, $\text{var}(Q_{t+1} | q_t)$ será uma função crescente de q_t . Conseqüentemente, a incerteza sobre a vazão anual que suceda à de um ano seco é menor que a incerteza sobre a vazão que suceda à de um ano úmido. Em bacias em que o escoamento básico seja um componente importante do deflúvio total, esta propriedade é particularmente atraente.

O modelo expresso pela expressão (4.30) pode ser interpretado como uma decomposição de Y_{t+1} em duas parcelas: uma completamente depende de Y_t , dado por ΦY_t ; e outra independente de Y_t , dado por V_{t+1} . O significado estatístico da parcela dependente pode ser verificado testando-se a hipótese de que Φ seja nulo. Analisando-se a partir desta hipótese, o estimador Φ (autocorrelação amostral, Quadro 49) tem distribuição aproximadamente normal com valor esperado nulo e variância n^{-1} .

Conseqüentemente, se $\hat{\Phi} > 1.28/\sqrt{n}$, pode-se rejeitar a hipótese nula (isto é, aceitar a parcela dependente) com nível de significância de 10%. Entretanto, se $\hat{\Phi} \leq 1.28/\sqrt{n}$, a decisão quanto à consideração ou não da parcela dependente deve ser feita levando-se em consideração o uso que se pretenda dar às séries sintéticas, assunto examinado quando se tratou da escolha do modelo estocástico.

Se o modelo da equação (4.30) estiver correto, $V_{t+1} = Y_{t+1} - \Phi Y_t$ será independente de Y_{t-1} . Essa condição pode ser facilmente verificada, calculando-se o coeficiente de correlação amostral $\text{corr}(V_{t+1}, Y_{t-1})$. Se $\text{corr}(V_{t+1}, Y_{t-1}) > 1.28/\sqrt{n}$ convém substituir V_{t+1} por V'_{t+1} e expressar V'_{t+1} como uma soma de um componente dependente de Y_{t-1} e outro independente de Y_{t-1} , dado por V_{t+1} . Isto é,

$$Y_{t+1} = \phi_1 Y_t + V'_{t+1} \quad (4.31)$$

e
$$V'_{t+1} = \phi_2 Y_{t-1} + V_{t+1} \quad (4.32)$$

Logo,

$$Y_{t+1} = \phi_1 Y_t + \phi_2 Y_{t-1} + V_{t+1} \quad (4.33)$$

onde

$$\phi_2 = \text{corr}(Y_{t+1}, Y_{t-1} | Y_t)$$

O modelo expresso pela equação (4.32) é linear e auto-regressivo de ordem 2, AR(2), já que Y_{t+1} depende de Y_t e de Y_{t-1} . Generalizando, um modelo AR(p) é dado por

$$Y_{t+1} = \sum_{i=1}^p \phi_i Y_{t+1-i} + V_{t+1} \quad (4.34)$$

onde V_{t+1} independente de V_t, V_{t-1}, \dots e de Y_t, Y_{t-1}, \dots

A multiplicação de ambos os lados da equação (4.34) por Y_{t+1-j} e a aplicação do operador valor esperado resulta em

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \text{ para } j = 1, 2, \dots, p \quad (4.35)$$

onde

$$\rho_j = \rho_{-j} = \text{corr}(Y_{t+1}, Y_{t+1-j})$$

Os coeficientes $\phi_i, i = 1, 2, \dots, p$ dependem da ordem do modelo p . Por exemplo, se $p = 1$, tem-se $\phi_1 = \rho_1$, e se $p = 2$, o valor de ϕ_1 é obtido resolvendo-se o seguinte sistema de equações:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \quad (4.36)$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \quad (4.37)$$

Ou seja, $\phi_1 = \rho_1$, se $p = 1$, e $\phi_1 = \rho_1(1 - \rho_2)/(1 - \rho_1^2)$, se $p = 2$. Para explicitar essa dependência, as equações (4.35) são reescritas mostrando-se a dependência de cada ϕ_j na ordem $p, \phi_{p,j}$

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{p,1} \\ \phi_{p,2} \\ \vdots \\ \vdots \\ \phi_{p,p} \end{bmatrix} \quad (4.38)$$

Pode-se mostrar que a condição para que o modelo AR(p) seja estacionário é que todas as raízes do polinômio (4.39) estejam situadas dentro do círculo unitário.

$$x^p - \phi_{p,1} x^{p-1} - \phi_{p,2} x^{p-2} \dots - \phi_{p,p} \quad (4.39)$$

Por exemplo, se $p = 1$, a condição de estacionariedade é dada por $\phi_{p,p} = \rho_1 < 1$.

Já, se $p = 2$, a condição de estacionariedade é dada por $(\phi_{2,1} \pm \sqrt{\phi_{2,1}^2 + 4\phi_{2,2}})^2 < 1$ que equivale a

$$-1 < \phi_{2,2} < 1, \phi_{2,1} + \phi_{2,2} < 1 \text{ e } \phi_{2,2} - \phi_{2,1} < 1 \quad (4.40)$$

Foi visto que, no modelo AR(2), $\phi_{2,2}$ é o coeficiente de autocorrelação parcial entre Y_{t+1} e Y_{t-1} , conhecido o valor de $Y_t, \text{corr}(Y_{t+1}, Y_{t-1} | Y_t)$ - equação (4.33) -, onde $\phi_{2,2}$ é chamado de ϕ_2 . Generalizando,

$$\phi_{p,p} = \text{corr}(Y_{t+1}, Y_{t+1-p} | Y_t, Y_{t-1}, \dots, Y_{t+2-p}) \quad (4.41)$$

A aplicação da regra de Cramer na equação de Yule-Walker (4.38) conduz a

$$\phi_{k,k} = A/B \quad (4.42)$$

onde A e B são os seguintes determinantes

$$A = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \rho_1 & \rho_k & \cdot \end{vmatrix} \quad (4.43)$$

$$B = \begin{vmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-3} & \rho_{k-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \rho_1 & 1 & \cdot \end{vmatrix} \quad (4.44)$$

É evidente que, para um processo AR(p), o determinante A é nulo para k > p, uma vez que uma das linhas será uma combinação linear das demais. Conseqüentemente, um método para se determinar a ordem p do modelo consiste em generalizar o procedimento adotado quando se rejeitou o modelo AR(0) – processo independente – em favor de AR(1). Isto é, dado que o modelo é pelo menos de ordem p-1, calcula-se o coeficiente $\Phi_{p,p}$ e testa-se a hipótese de que esse coeficiente seja nulo. Se $\Phi_{p,p}$ for suficientemente grande para rejeitar a hipótese, o processo é pelo menos de ordem p. E assim sucessivamente. Conforme anteriormente mencionado, a não rejeição da hipótese nula não significa necessariamente que o modelo de ordem mais elevada deva ser desconsiderado. Essa decisão deve levar em consideração o uso que se fará das séries sintéticas.

Se o modelo da equação (4.30) estiver correto, $V_{t+1} = Y_{t+1} - \Phi Y_t$ será independente de V_t . Essa condição pode ser facilmente verificada, calculando-se o coeficiente de correlação amostral, $c\hat{o}r r(V_{t+1}, V_t)$. Se $|c\hat{o}r r(V_{t+1}, V_t)| > 1.28/\sqrt{n}$, convém substituir V_{t+1} por V''_{t+1} e expressar V''_{t+1} como a soma de um componente dependente de V_t e com outro independente de V_t , dado por V_{t+1} . Isto é,

$$Y_{t+1} = \Phi_1 Y_t + V''_{t+1} \quad (4.45)$$

e

$$V''_{t+1} = \theta_1 V_t + V_{t+1} \quad (4.46)$$

Portanto,

$$Y_{t+1} = \Phi_1 Y_t + \theta_1 V_t + V_{t+1} \quad (4.47)$$

V_{t+1} independente de V_t, V_{t-1} e de $Y_t, Y_{t-1}, Y_{t-2}, \dots$

O modelo expresso pela equação (4.47) é conhecido como Arma (1,1). Trata-se de um caso particular do modelo Arma (p,q), que foi minuciosamente estudado por Box e Jenkins (1976) e que tem encontrado aplicação em diversas disciplinas, principalmente quando utilizado para previsão. A aplicação dos modelos Arma às séries hidrológicas tem sido feita por diversos autores (por exemplo, Braga, 1983). A forma geral desse modelo é dada por:

$$Y_{t+1} = \sum_{i=1}^p \phi_i Y_{t+1-i} + \sum_{j=1}^q \theta_j V_{t+1-j} + V_{t+1} \quad (4.48)$$

Na modelagem de vazões anuais, raramente utiliza-se algum modelo que seja de ordem superior a Arma (2,1). Para esse modelo, tem-se o seguinte correlograma (função de ρ_k em k):

$$\rho_0 = 1 \quad (4.49)$$

$$\rho_1 = A/B \quad (4.50)$$

$$\rho_k = \Phi_1 \rho_{k-1} + \Phi_2 \rho_{k-2}, \quad k \geq 2 \quad (4.51)$$

onde A e B são os seguintes determinantes:

$$A = \begin{vmatrix} (1 - \phi_1^2) & (1 - \phi_1 \theta_1 + \theta_1^2) \\ -\phi_1 & -\theta_1 \end{vmatrix} \quad (4.52)$$

$$B = \begin{vmatrix} (1 - \phi_1 \theta_1 + \theta_1^2) & -\phi_1 (1 + \phi_2) \\ -\theta_1 & (1 - \phi_2) \end{vmatrix} \quad (4.53)$$

Os correlogramos de modelos Arma (p,q) de ordem mais baixa são obtidos zerando-se os correspondentes parâmetros das equações (4.49) a (4.53).

Pode-se estimar os parâmetros de um modelo Arma (p,q) igualando as expressões do autocorrelograma do modelo, que são funções dos parâmetros, com os correspondentes coeficientes de autocorrelação amostral, definidos por:

$$\hat{\rho}_j = \frac{1}{n} \sum_{t=1}^{n-j} Y_{t+j} Y_t \quad (4.54)$$

Quando $q = 0$, basta que se substituam os coeficientes pelas estimativas nas equações de Yule-Walker, equação (4.39). Quando $q = 1$, é necessário resolver um sistema não-linear com duas equações – caso Arma (1,1) – ou três equações – caso Arma (2,1). O emprego desse método de estimativa faz com que a série sintética produzida pelo modelo preserve os $p + q$ primeiros coeficientes da autocorrelação amostral da série histórica.

Alternativamente, podem-se estimar os parâmetros que minimizem a soma dos quadrados dos resíduos (componentes independentes), o que, no caso, equivale ao método da máxima verossimilhança:

$$S = \sum_t v_t^2 \tag{4.55}$$

Para um processo AR (p), ou equivalente Arma (p,0), equação (4.34), os coeficientes que minimizam a expressão (4.55) são dados pela equação (4.56), cuja derivação pode ser encontrada nos textos de regressão multivariada linear

$$\hat{\phi} = (B^T B)^{-1} B^T A \tag{4.56}$$

onde

$$A = \begin{bmatrix} Y_{P+1} \\ Y_{P+2} \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} \tag{4.57}$$

$$B = \begin{bmatrix} Y_P & Y_{P-1} & \dots & Y_1 \\ Y_{P+1} & Y_P & \dots & Y_2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ Y_{n-1} & Y_{n-2} & \dots & Y_{n-P} \end{bmatrix} \tag{4.58}$$

$$\hat{\phi} = \begin{bmatrix} \hat{\phi} \\ 1 \\ \hat{\phi} \\ 2 \\ \cdot \\ \cdot \\ \hat{\phi} \\ P \end{bmatrix}$$

A variância do componente independente é estimada por

$$\hat{\sigma}_v^2 = \frac{1}{n-p} \sum_{t=P+1}^n (y_t - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \dots - \hat{\phi}_p y_{t-p})^2$$

A analogia entre o modelo Arma (p,0) e a regressão multivariada linear, equações (4.56 – 4.60), facilita a inclusão da incerteza dos parâmetros (item 4.3) na modelagem. Pode-se mostrar que $(n-p) \hat{\sigma}_v^2 / \tau_v^2$ tem distribuição qui-quadrado com $(n-p)$ graus de liberdade e $\hat{\phi}$ tem distribuição multivariada normal com média ϕ e matriz de covariância $\tau_v^2 (B^T B)^{-1}$. Conseqüentemente, a geração de cada série sintética, para esta particular opção de modelagem, seguiria as seguintes etapas:

- a) Sorteie uma variável qui-quadrado com $n-p$ graus de liberdade (χ^2), que é igual à soma dos quadrados de $(n-p)$ variáveis normal padrão, independentes entre si

- b) Assuma $\hat{\sigma}_v^2 = (n-p) \hat{\sigma}_v^2 / \chi^2$, onde $\hat{\sigma}_v^2$ é calculado pela equação (4.60)
- c) Sorteie um vetor normal com média $\hat{\mu}$ e matriz de covariância $\hat{\sigma}_v^2 (B^T B)^{-1}$. Os procedimentos necessários para a execução dessa tarefa serão apresentados logo adiante. Seja \hat{v} o vetor sorteado.
- d) Gere série sintética de vazões com parâmetros $\hat{\sigma}_v^2$ e $\hat{\mu}$.

Quando $q \geq 1$, a estimativa dos parâmetros tem que ser feita iterativamente. Inicialmente, assume-se um conjunto de parâmetros $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ que permite a determinação de v_{p+1}, v_{p+2}, \dots , através da equação.

$$v_t = y_t - \sum_{i=1}^p \hat{\phi}_i y_{t-1} - \sum_{j=1}^q \hat{\theta}_j v_{t-j} \quad t = p+1, p+2, \dots$$

Os valores iniciais v_1, v_2, \dots, v_p são desconhecidos. Assume-se que eles são iguais ao valor esperado, isto é, zero. O conjunto v_{p+1}, v_{p+2}, \dots permite que se calcule a função objetivo equação (4.55) – e suas derivadas parciais em relação aos parâmetros de θ e ϕ . Conseqüentemente, um conjunto de parâmetros pode ser calculado o que resulta numa nova seqüência v_{p+1}, v_{p+2}, \dots , e assim sucessivamente, até que algum critério de convergência seja satisfeito. O leitor interessado em mais informações sobre a aplicação do método de máxima verossimilhança para estimar os parâmetros de um modelo Arma (p,q) deve consultar, por exemplo, Mcleod e Sales (1983). Qualquer modelo Arma (p,q) que se cogite adotar precisa satisfazer a hipótese de independência dos v_t . Se essa hipótese for verdadeira, $\text{corr}(v_t, v_{t+\tau}), \tau < 0$ terá distribuição aproximadamente normal com valor esperado nulo e variância igual a $1/n$. Conseqüentemente, qualquer valor de $\text{corr}(v_t, v_{t+\tau}) \sqrt{n}$ superior, em módulo, a 1.28 deve funcionar como um alerta contra a escolha do modelo. Equivalentemente, Box e Jenkins (1970) sugeriram a teste de Portmanteau, que considera todas as autocorrelações amostrais simultaneamente através da expressão

$$\chi^2 = (n-p) \sum_{i=1}^k [\text{corr}(v_t, v_{t+i})]^2$$

Pode-se mostrar que χ^2 tem aproximadamente distribuição qui-quadrado, com $k-p$ graus de liberdade. O teste consiste em comparar o valor calculado com o valor crítico tabelado, para o nível de significância selecionado. O valor de k deve ser, pelo menos, igual a $18 \sqrt{6}$.

4.6 Modelagem multivariada

Dois aproveitamentos hídricos próximos são, em geral, atingidos simultaneamente por baixas (ou altas) vazões anuais, ocasionando secas (ou cheias) regionais. Nesse caso, $\text{corr}(Q_t^a, Q_t^b) > 0$, onde Q_t^a e Q_t^b são as vazões anuais nos locais a e b. Por outro lado, locais afastados, mas submetidos à mesma fonte de umidade, podem ser hidrologicamente complementares, no sentido de que um ano úmido em a tenderá a estar associado com um ano seco em b, e vice-versa. Nesse caso, espera-se que $\text{corr}(Q_t^a, Q_t^b) < 0$.

Q_t^b é o resultado de uma complexa sucessão de eventos hidrológicos que ocorrem durante o ano t na área de drenagem a montante do local b. Dependendo da aplicação que se cogite para as séries sintéticas, pode haver interesse em utilizar a totalidade da área de drenagem ou apenas sua porção "incremental". Por exemplo, se o local b for situado a jusante do local a, a série histórica Q_t^b será definida, nesta segunda alternativa, como a diferença entre as vazões totais em b e em a.

Nas páginas anteriores, em que se tratou da modelagem univariada, foi visto que o caso mais simples é quando $\text{corr}(Q_t, Q_{t+1}) = 0$, isto é, quando as vazões anuais são *serial* ou *temporalmente* independentes. Nesse caso, a modelagem da dependência *espacial* é feita através da seguinte equação

$$\underline{Y}_t = B \underline{V}_t \quad (4.63)$$

onde $\underline{Y}_t^T = [Y_t^a, Y_t^b, \dots]$ é um vetor de dimensão m, igual ao número de locais sob modelagem. Y_t^a é uma variável normal padrão, correspondente à vazão Q_t^a no ano t e no local a (o tema normalização já foi tratado neste capítulo). \underline{V}_t é um vetor também de dimensão m, de componentes independentes e com distribuição normal padrão. B é uma matriz parâmetro, a ser estimada, de dimensão $m \times m$.

Multiplicando-se a equação acima por \underline{Y}_t^T , tem-se

$$\underline{Y}_t \underline{Y}_t^T = B \underline{V}_t \underline{V}_t^T B^T \quad (4.64)$$

Lembrando que o valor esperado de uma matriz é a matriz dos valores esperados:

$$E(\underline{Y}_t \underline{Y}_t^T) = B E(\underline{V}_t \underline{V}_t^T) B^T \quad (4.65)$$

O elemento na linha a, coluna b da matriz $E(\underline{Y}_t \underline{Y}_t^T)$ é $E(Y_t^a Y_t^b)$. Como Y_t^a e Y_t^b são normais padrão, $E(\underline{Y}_t \underline{Y}_t^T) = \text{corr}(\underline{Y}_t, \underline{Y}_t)$. Logo, a matriz $R = E(\underline{Y}_t \underline{Y}_t^T)$, composta por correlações espaciais, é simétrica e tem os elementos da diagonal iguais a 1. Analogamente, os componentes de $E(\underline{V}_t \underline{V}_t^T)$ são também correlações e, como por construção estes elementos são independentes entre si, $E(\underline{V}_t \underline{V}_t^T)$ é a matriz identidade de dimensão mxm. Portanto, a equação acima pode ser reescrita como

$$BB^T = R \tag{4.66}$$

A estimativa da matriz B é feita substituindo-se o lado direito da equação acima pela estimativa da matriz R. Cada elemento dessa matriz é o coeficiente de correlação amostral entre concomitantes observações y_t^a e y_t^b , $t = 1, 2, \dots$

Existem muitas matrizes B que satisfazem a equação acima. Uma maneira trivial de se encontrar uma dessas soluções é arbitrar que B seja triangular.

Exemplo 17

Se o número de locais sob modelagem for $m=3$, a equação (4.66) seria escrita como

$$\begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} b_{11} & b_{21} & b_{31} \\ 0 & b_{22} & b_{32} \\ 0 & 0 & b_{33} \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}$$

ou

$$\begin{aligned} b_{11}^2 &= r_{11} \rightarrow b_{11} = r_{11}^{1/2} \\ b_{11} b_{21} &= r_{12} \rightarrow b_{21} = r_{12}/b_{11} \\ b_{11} b_{31} &= r_{13} \rightarrow b_{31} = r_{13}/b_{11} \\ b_{21}^2 + b_{22}^2 &= r_{22} \rightarrow b_{22} = (r_{22} - b_{21}^2)^{1/2} \\ b_{21} b_{31} + b_{22} b_{32} &= r_{23} \rightarrow b_{32} = (r_{23} - b_{21} b_{31})/b_{22} \\ b_{31}^2 + b_{32}^2 + b_{33}^2 &= r_{33} \rightarrow b_{33} = (r_{33} - b_{31}^2 - b_{32}^2)^{1/2} \end{aligned}$$

B pode também ser calculada por decomposição espectral:

$$B = CD \tag{4.67}$$

onde

C = matriz mxm, em que cada coluna é um autovetor de R;

D = matriz mxm com a diagonal composta por raízes quadradas dos autovalores de R, ordenados decrescentemente, e elementos fora da diagonal nulos.

Qualquer que seja o método de cálculo de B, é necessário que a matriz R seja positivo-definida para que haja uma solução. Quando as séries não são todas de mesmo comprimento ou as séries históricas contêm lacunas, R pode resultar não positivo-definida, sendo conveniente, no caso, a utilização do método da máxima verossimilhança (Kuczera, 1987), em contraposição ao método dos momentos, que conduziu à equação (4.66).

No caso mais geral, em que Q_t e Q_{t+1} são dependentes (e, portanto, também Y_t e Y_{t+1}), é preciso adotar um modelo que considere a dependência temporal. Alcança-se esse objetivo generalizando o modelo AR(1) para o caso multivariado, pioneiramente aplicado à hidrologia por Matalas (1967):

$$\underline{Y}_{t+1} = A \underline{Y}_t + B \underline{V}_{t+1} \tag{4.68}$$

Os vetores \underline{Y}_{t+1} , \underline{Y}_t e \underline{V}_{t+1} são de dimensão m (número de locais sob modelagem) e as matrizes parâmetro A e B são de dimensão mxm. As matrizes A e B podem ser estimadas através do método dos momentos, multiplicando-se ambos os lados por \underline{Y}_t^T e achando-se os valores esperados:

$$E(\underline{Y}_{t+1} \underline{Y}_t^T) = AE(\underline{Y}_t \underline{Y}_t^T) + BE(\underline{V}_{t+1} \underline{Y}_t^T) \tag{4.69}$$

Como os componentes dos vetores \underline{V}_{t+1} e \underline{Y}_t são independentes entre si, $E(\underline{V}_{t+1} \underline{Y}_t^T) = 0$ é uma matriz com todos os elementos nulos. Portanto, a equação pode ser reescrita como

$$S = AR, \text{ ou} \tag{4.70}$$

$$A = SR^{-1}$$

onde S é uma matriz mxm cujo componente, na linha a e coluna b, é a correlação entre Y_{t+1}^a e Y_t^b . Conseqüentemente, a diagonal da matriz S é formada pelos coeficientes de autocorrelação. Além disso, S é não simétrica, ao contrário de R.

Multiplicando-se ambos os lados da equação (4.68) por \underline{Y}_t^T e achando-se os valores estimados, obtém-se:

$$E(\underline{Y}_{t+1} \underline{Y}_t^T) = AE(\underline{Y}_t \underline{Y}_t^T) \bar{A} + BE(\underline{V}_{t+1} \underline{Y}_t^T) \bar{A} + \tag{4.71}$$

$$+ BE(\underline{V}_{t+1} \underline{V}_t^T) B^T \bar{A} + A E(\underline{V}_t \underline{V}_t^T) R^T$$

Como $E(\underline{V}_t, \underline{Y}_t^a)$ e $E(\underline{Y}_t, \underline{V}_t^a)$ são matrizes nulas, e $E(\underline{V}_t, \underline{V}_t^T)$ é a matriz identidade, a equação acima pode ser reescrita como

$$R = ARA^T + BB^T \quad (4.72)$$

ou

$$BB^T = R - ARA^T \quad (4.73)$$

Substituindo-se o valor de A, dado pela equação (4.70), e levando-se em consideração que R é uma matriz simétrica, a equação acima pode ser reescrita como

$$BB^T = R - SR^{-1}S^T \quad (4.74)$$

As matrizes parâmetro A e B podem ser facilmente estimadas substituindo-se as matrizes R e S pelas correspondentes matrizes de correlação amostral R e S nas equações (4.70) e (4.73) ou (4.74).

Se o lado direito da equação (4.74) não for uma matriz positiva definida, convém que se utilize o método da máxima verossimilhança (Kuczera, 1987).

Kelman *et al.*, (1981) e Pereira *et al.*, (1984) generalizaram a equação (4.67) para o caso em que se pretende gerar séries sintéticas para um conjunto de locais, uma bacia hidrográfica por exemplo, sem que seja necessário gerar novas séries sintéticas para outros locais que já disponham dessas séries:

$$\underline{Y}_{t+1} = A\underline{Y}_t + B\underline{V}_t + F\underline{W}_{t+1} \quad (4.75)$$

onde

F – matriz $m \times k$ (m é o número de locais gerados presentemente e k é o número de locais já gerados);

\underline{W}_{t+1} – vetor $k \times 1$ de vazões sintéticas anteriormente gerados para os k locais.

Os demais termos da equação acima foram anteriormente definidos. A estimativa dos parâmetros A, B e F é feita de maneira a preservar a dependência entre vazões nos novos e nos antigos locais. Essa modelagem é particularmente útil para o planejamento da expansão de um sistema hidrelétrico, quando se faz necessária a simulação do sistema com e sem a adição de novas usinas.

Matalas (1967) sugeriu que, se a preservação do coeficiente de correlação amostral da série histórica entre \underline{Y}_{t+1}^a e \underline{Y}_t^b não for particularmente relevante, é possível obter-se um processo de estimativa mais simples, arbitrando-se que todos os elementos fora da diagonal de A são nulos. Os elementos da diagonal são feitos iguais aos coeficientes de autocorrelação amostral. A matriz B é calculada, neste caso, através da equação (4.73) e não através da equação (4.74), já que só a diagonal da matriz S precisa ser estimada. As séries sintéticas produzidas por esse modelo terão coeficientes de autocorrelação (em cada local) e coeficientes de correlação espacial que tenderão aos correspondentes valores amostrais das séries históricas, na medida em que o

número de anos da série sintética crescer. Por outro lado, $\text{corr}(\underline{Y}_{t+1}^a, \underline{Y}_t^b)$ tenderá ao produto de $\text{corr}(\underline{Y}_{t+1}^a, \underline{Y}_t^a)$ com $\text{corr}(\underline{Y}_t^a, \underline{Y}_t^b)$, conforme pode ser visto através da equação (4.69).

É interessante observar que a adoção de uma matriz A diagonal equivale à modelagem do processo multivariado em duas etapas. Em primeiro lugar, o modelo AR(1) univariado é ajustado à série histórica de cada local, permitindo a determinação de componentes temporalmente independentes. Em segundo lugar, a dependência espacial entre $(\underline{Y}_t^a, \underline{Y}_t^b)$, para quaisquer locais a e b é indiretamente levada em consideração ao se preservar a correlação espacial entre os componentes temporalmente independentes.

De uma maneira mais geral, a série histórica de cada local pode ser ajustada por um modelo Arma(p,q) univariado, sendo que a ordem (p,q) pode variar de local para local. Essa abordagem é chamada de modelo Carma (Salas *et al.*, 1980) e é interessante porque evita as complicações vinculadas à escolha da ordem do modelo e à estimativa dos parâmetros que resultariam da extensão do modelo AR(1) multivariado para o caso mais geral de Arma(p,q) multivariado (Hipel, 1985).

Para cada local $k = a, b, c, \dots$ em que se tenha ajustado um modelo Arma (p_k, q_k), a série temporalmente independente é determinada por

$$z_t^k = y_t^k - \sum_{i=1}^{p_k} \phi_i^k y_{t-i}^k - \sum_{j=1}^{q_k} \theta_j^k z_{t-j}^k \quad (4.76)$$

A dependência espacial entre Z_t^a e Z_t^b , para os locais a e b quaisquer, é modelada por

$$Z_t = B V_t \quad (4.77)$$

Conforme anteriormente visto, a matriz B é estimada resolvendo-se a equação

$$BB^T = \hat{U} \quad (4.78)$$

onde \hat{U} é a estimativa da matriz de correlação $E(\underline{Z}_t, \underline{Z}_t^T)$. O elemento da linha a e coluna b de \hat{U} é o coeficiente de correlação amostral entre as concomitantes observações Z_t^a e Z_t^b , $t = 1, 2, \dots$

Camacho (1985) observou que o ajuste separado dos modelos univariados de cada série desperdiça informações contidas no histórico e apresenta um algoritmo para as estimativas concomitantes de máxima verossimilhança.

4.7 Modelos de Longa Memória

Hurst (1956) estudou séries temporais bastante longas de diversos processos naturais, centrando atenção na estatística "amplitude ajustada", definida como:

$$a_n = \max_t \left[\sum_{\tau=1}^t (q_\tau - \bar{q}) \right] - \min_t \left[\sum_{\tau=1}^t (q_\tau - \bar{q}) \right] \quad (4.79)$$

Pode-se mostrar que a_n é o tamanho que deve ter um reservatório, caso o volume anual regularizado seja igual ao volume anual afluyente médio ao longo de n anos (\bar{y}) e caso a seqüência futura de afluências seja $y_1, y_2, \dots, y_n, y_1, y_2, \dots, y_n, y_1, y_2, \dots$

Hurst fez o gráfico de $\log(a_n/s)$, onde s é o desvio padrão amostral de Q , contra $\log n$, e descobriu um bom ajuste para uma reta com coeficiente angular em torno de $H = 0,7$. Esse resultado permitiu inferir que

$$E(A_n/\sigma) \approx n^H \quad (4.80)$$

onde H é o chamado coeficiente de Hurst.

Paralelamente, o próprio Hurst e outros autores, relacionados por Gomide (1975), chegaram à conclusão de que todos os modelos estocásticos até então cogitados para representar séries de vazões anuais possuíam a propriedade de

$$\lim_{n \rightarrow \infty} E(A_n/\sigma) \approx n^{0,5} \quad (4.81)$$

Como o resultado (4.80) foi baseado em séries com valores elevados de n , a discrepância entre a tendência empírica expressa pela equação (4.80) e o resultado teórico expresso pela equação (4.81), chamada de "fenômeno de Hurst", causou desconforto na comunidade hidrológica durante alguns anos. Modelos estocásticos foram desenvolvidos especificamente para reproduzir o coeficiente de Hurst, notadamente o *Fast Fractional Gaussian Noise* (Mandelbrot, 1971, e McLeod e Hipel, 1978) e o *Broken Line* (Mejia et al, 1972). Nesses modelos, a correlação entre Q_t e Q_{t+6} não decai tão rapidamente como nos módulos Arma, e por isso são chamados de longa memória.

Posteriormente, diversas explicações foram formuladas para o fenômeno de Hurst, sendo as mais importantes:

- a) o efeito transiente – para valores finitos de n , da ordem de grandeza dos utilizados por Hurst, o coeficiente H é maior do que 0,5 apesar de que, quando n tende para infinito, H tende para 0,5 (Gomide 1975, Lettenmaier e Burges, 1977).
- b) o efeito da não estacionariedade – séries naturais que exibam não estacionariedade, devido, por exemplo, à mudança da régua quando se medem níveis ao longo de vários séculos (rio Nilo, por exemplo) apresentarão valores de H que não tendem para 0,5 (Klemes, 1974).

Comparações entre as propriedades das séries sintéticas geradas por modelos de longa memória (preservação do coeficiente de Hurst) e os de curta memória (Arma), para estudos relacionados à confiabilidade de um sistema hidrelétrico (*California Energy*

Commission, 1981) e ao desempenho de um reservatório de regularização (Klemes et al 1981) mostraram não haver diferença prática significativa entre as duas abordagens.

Como o fenômeno de Hurst não parece ser relevante em termos de aplicação, o assunto não será desenvolvido neste livro. A menção serve apenas para satisfazer uma curiosidade histórica.

5. MODELAGEM DE VAZÕES MENSAIS

5.1 Sazonalidade

Os modelos de simulação e/ou otimização empregados para o planejamento e a operação de sistemas de recursos hídricos utilizam usualmente séries de vazões com intervalo de discretização mensal.

Ao contrário do processo anual, vazões mensais não podem ser consideradas como originárias de um processo estacionário. A Figura 120 mostra a evolução da média das vazões, para cada mês do ano, do rio Paraíba do Sul em Barra do Pirai. Um simples exame comprova o caráter sazonal dessa estatística. O mesmo se dá com o desvio padrão e coeficiente de autocorrelação, respectivamente nas Figuras 120 e 121.

5.2 Modelo PAR(1)

O modelo estacionário expresso pela equação (4.14) não pode mais ser empregado porque, para vazões mensais, a igualdade $E(Q_t) = E(Q_{t+1})$ por exemplo, não vale. Uma adaptação da equação (4.14) pode, no entanto, ser conseguida através do chamado modelo de Thomas-Fiering ou PAR(1), ou AR(1) periódico.

$$M_{\tau+1,t} = \mu_{\tau+1} + \frac{\rho_{\tau} \sigma_{\tau+1}}{\sigma_{\tau}} (M_{\tau,t} - \mu_{\tau}) + \sigma_{\tau+1} \sqrt{1-\rho_{\tau}^2} Z_{\tau+1,t} \quad (5.1)$$

onde

$M_{\tau,t}$ é a vazão no mês τ , ano t
 $\tau = 1, 2, \dots, 12$ é o índice do mês
 $t = 1, 2, \dots$ é o índice do ano
 $Z_{\tau,t}$ é o componente independente padrão

Adota-se a convenção "circular" de que, se $\tau = 12$, então

$$M_{\tau+1,t} = M_{1,t+1}, \mu_{\tau+1} = \mu_1 \text{ e } \sigma_{\tau+1} = \sigma_1$$

A manipulação da equação (5.1), nos mesmos moldes da discussão efetuada após a apresentação da equação (4.14), revelaria uma série de resultados análogos aos do item 4.4 deste capítulo.

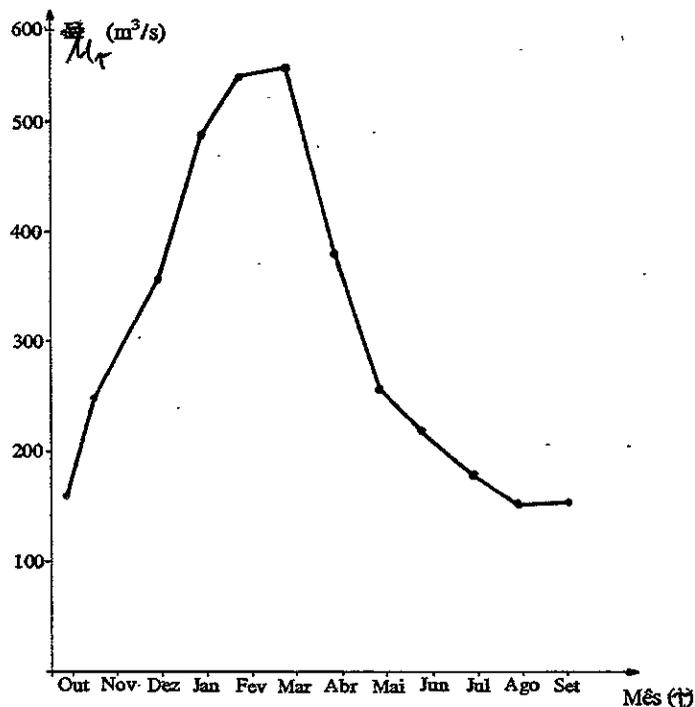


Figura 121. Variação sazonal do método da média das vazões mensais do rio Paraíba do Sul em Barra do Piauí

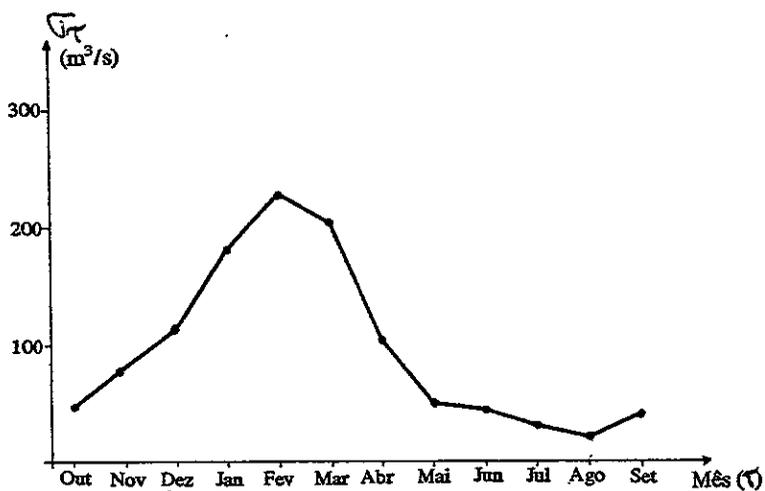


Figura 122. Variação sazonal do desvio padrão das vazões mensais do Rio Paraíba do Sul em Barra do Piauí

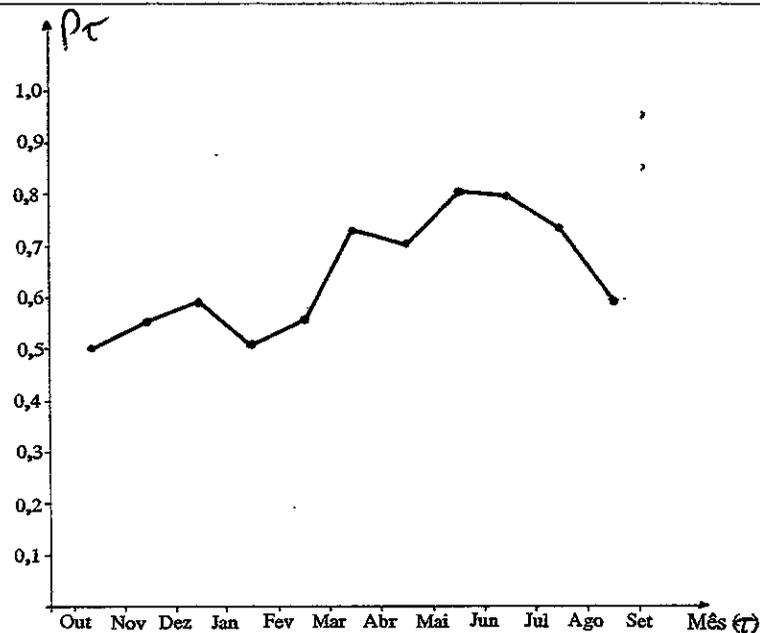


Figura 123. Variação sazonal do coeficiente de correlação de vazões meses consecutivos do Rio Paraíba do Sul em Barra do Piauí

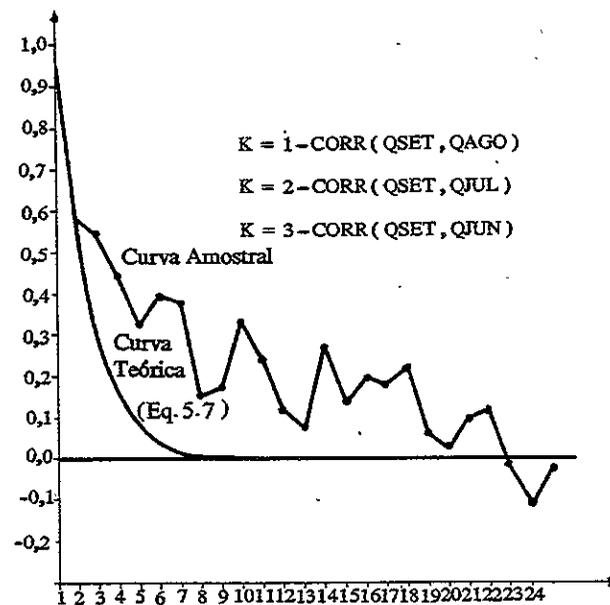


Figura 124. Correlação entre vazões mensais do Rio Paraíba do Sul em Barra do Piauí

$$E(M_{\tau, t}) = \mu_{\tau} \quad (5.2)$$

$$\text{var}(M_{\tau, t}) = \sigma_{\tau}^2 \quad (5.3)$$

$$E(M_{\tau+1, t} | M_{\tau, t} = m_{\tau, t}) = \mu_{\tau+1} + \frac{\rho_{\tau} \sigma_{\tau} + 1}{\sigma_{\tau}} (m_{\tau, t} - \mu_{\tau}) \quad (5.4)$$

$$\text{var}(M_{\tau+1, t} | M_{\tau, t} = m_{\tau, t}) = \sigma_{\tau+1}^2 (1 - \rho_{\tau}^2) \quad (5.5)$$

$$\text{corr}(M_{\tau+1, t}, M_{\tau, t}) = \rho_{\tau} \quad (5.6)$$

$$\text{corr}(M_{\tau+1, t}, M_{\alpha, t-1}) = \frac{\rho_{\tau} \rho_{\tau-1} \dots \rho_1 \rho_2 \rho_1 \rho_{10} \dots}{\text{produtos de K fatores}} \quad (5.7)$$

396 onde

$$K = 121 + \tau - \alpha + 1$$

A extensão do modelo expresso pela equação (5.1) para o caso multivariado poderia ser conseguida, de forma aproximada, pela preservação das correlações entre pares de ruídos $(Z_{\tau, t}^a, Z_{\tau, t}^b)$ correspondentes aos locais a e b, conforme explicado no texto que antecede a equação (4.76).

A Figura 124 mostra o gráfico de correlação da vazão de setembro com agosto (K=1), de setembro com julho (K=2), e assim por diante. Como se vê, o ajuste deixa muito a desejar, na medida em que a curva teórica – calculada pela equação (5.7) – decai muito mais rapidamente do que a amostral. Resultados semelhantes seriam encontrados caso se centrasse a análise em qualquer outro mês que não setembro, e para uma grande variedade de rios. Conseqüentemente, as séries sintéticas geradas pelo modelo de Thomas-Fiering tenderão a exibir menos persistência do que as correspondentes séries históricas. Mais ainda, chamando Q_t de vazão média anual no ano t, tem-se

$$Q_t = 12^{-1} \sum_{\tau} M_{\tau, t} \quad (5.8)$$

A variância da vazão sintética anual será

$$\text{var}(Q_t) = 144^{-1} \left\{ \sum_{\tau=1}^{12} \sigma_{\tau}^2 + 2 \sum_{\tau=1}^{11} \sum_{j=\tau+1}^{12} (\rho_{\tau} \rho_{\tau+i} \dots \rho_j) \sigma_{\tau} \sigma_j \right\} \quad (5.9)$$

que tenderá a ser inferior à verdadeira variância da vazão anual, porque o termo $(\rho_{\tau} \rho_{\tau+1} \dots \rho_j)$ tende a subestimar a verdadeira correlação entre $M_{\tau, t}$ e $M_{j, t}$.

Um argumento nas mesmas linhas conduziria à conclusão de que também a autocorrelação em escala anual da série sintética tenderá a ser inferior à correspondente autocorrelação da série histórica. O efeito da subestimação da variância e da autocorrelação em escala anual terá maior ou menor importância dependendo da aplicação que se pretenda dar às séries sintéticas. Em estudos com horizontes de planejamento curto, de alguns meses, ou quando o reservatório sempre enche ou esvazia pelo menos uma vez por ano, o efeito é de menor importância. Por outro lado, quando o estudo exigir a simulação da operação de reservatórios capazes de exercer uma regularização plurianual, o efeito é de maior importância. Os diversos modelos apresentados a seguir procuram sanar esse problema através de duas alternativas:

- a) adoção de um modelo Arma(p,q) periódico, chamado Parma(p,q) para vazões mensais;
- b) adoção de um modelo que gere inicialmente as vazões anuais e depois desagregue cada valor anual em componentes mensais.

5.3 Modelo Parma(p,q)

397

Modelo PAR(p)

O modelo PAR(1), expresso pela equação (5.1) pode ser visto como uma forma sintética de se representar um conjunto de 12 regressões, uma para cada mês do ano. Em cada uma dessas regressões, assume-se como variável previsora a vazão no mês imediatamente antecedente. Pode-se generalizar o conceito assumindo como variáveis predictoras as vazões nos p meses imediatamente antecedentes. A rigor, não há razão para impor para todas as 12 regressões um mesmo número de variáveis predictoras. O modelo resultante é chamado de PAR(p). Para um mês τ qualquer, tem-se

$$M_{\tau, t} = a_{\tau} Q_{\tau} + \sum_{i=1}^{p_{\tau}} a_{i\tau} M_{\tau-i, t} + Z_{\tau, t} \quad (5.10)$$

A seleção do número de variáveis predictoras, p_{τ} e a estimativa dos parâmetros $a_{i\tau}$ podem ser feitas através de procedimentos usualmente empregados na técnica de regressão. Uma amostra aleatória $\{Z_{\tau, 1}, Z_{\tau, 2}, \dots, Z_{\tau, n}\}$ é obtida para cada mês τ a partir da solução da equação acima para $Z_{\tau, t}$, substituindo-se os valores de $M_{\tau, t}$ pelas observações na série histórica.

Para efeito de geração de uma série sintética, deve-se proceder a sorteios dos ruídos $Z_{\tau, t}$, $\tau = 1, 2, \dots, 12$, seja através de uma distribuição de probabilidades que tenha sido ajustada à amostra aleatória obtida com os dados históricos, seja através da

distribuição empírica de $Z_{\tau,t}$ (o que equivale a assumir que os ruídos históricos são igualmente prováveis).

Modelo Arma(p,q) aplicado à série transformada

Nesta alternativa, aplica-se para cada mês uma transformação que resulte numa variável normal padrão. Dessa maneira, obtém-se uma série normalizada e estacionária. Possíveis transformações estão listadas no Quadro 51, não sendo necessário escolher o mesmo tipo de transformação para todos os meses. Um modelo Arma(p,q) é subsequentemente ajustado à série transformada, adotando-se as técnicas discutidas no item 4.4. A extensão para o caso multivariado se dá também através dos procedimentos discutidos naquele item.

Esse modelo admite implicitamente que a dependência entre as variáveis transformadas é função apenas da defasagem entre elas. Nas séries geradas pelo modelo, a correlação entre as variáveis correspondentes, por exemplo, aos meses de janeiro e março (defasagem igual a dois) será próxima da correlação correspondente aos meses de fevereiro e abril, idem em relação aos meses de março e junho, e assim sucessivamente.

Quando houver evidência estatística ou conceitual quanto à sazonalidade dos coeficientes de correlação da variável transformada, o modelo não deverá ser adotado. Por exemplo, suponha uma bacia hidrográfica com razoável capacidade de armazenamento no lençol freático e sujeita à uma época chuvosa, bem caracterizada, de outubro a abril. Admita que, durante a época chuvosa, a maior parte do deflúvio é de origem superficial e durante a época de estiagem (de maio a setembro), a maior parte do deflúvio provém do escoamento básico. Nessas circunstâncias, a correlação entre vazões de meses consecutivos (defasagem de um mês) da estação chuvosa deveria ser menor do que a correspondente correlação entre vazões de meses pertencentes ao período de estiagem. De fato, um deflúvio pequeno num mês da estação chuvosa é um indicativo precário da ocorrência ou não de precipitações abundantes no mês seguinte e, portanto, do que se pode esperar quanto ao escoamento superficial. Já um deflúvio pequeno num mês da estiagem é uma informação significativa sobre o estado do lençol freático, e, portanto, sobre o que se pode esperar quanto ao escoamento básico. A Figura 122 ilustra o caso. Thompson *et al* (1987) compararam o modelo PAR(p) com o modelo Arma(p,q) aplicado à série transformada e concluíram pela superioridade do primeiro.

Modelo Parma(p,q) aplicado à série transformada

Uma maneira de preservar o caráter sazonal da dependência entre vazões transformadas é considerar um conjunto de coeficientes do processo Arma(p,q) específico para cada mês do ano

$$X_{\tau+1,t} = \sum_{i=1}^{p_{\tau}+1} \phi_{\tau,i} X_{\tau+1-t,i} + \sum_{j=1}^{q_{\tau}+1} \theta_{\tau,j} Z_{\tau+1-t,j} + Z_{\tau+1,t} \tag{5.11}$$

onde

- $X_{\tau,t} = g_{\tau}(M_{\tau,t})$ tem distribuição normal padrão
- $X_{\tau,t} = X_{\tau-12,t+1}$ se $\tau > 12$
- $X_{\tau,t} = X_{\tau+12,t-1}$ se $\tau \leq 0$
- $Z_{\tau,t}$ = componente normal independente (ruído) com valor esperado nulo.



Salas *et al* (1982) derivam as equações de Yule-Walker para um processo Parma (p,q), ϕ_{τ} e q_{τ} constante. No caso particular de $p_{\tau} = q_{\tau} = 1, \forall \tau$, os 24 parâmetros: $\Phi_{\tau,1}$ e $\theta_{\tau,1}$ (ou simplesmente Φ_{τ} e θ_{τ}), $\tau = 1,12$, podem ser estimados por

$$\Phi_{\tau} = \rho_{\tau}(2)/\rho_{\tau-1}(1) \tag{5.12}$$

$$\theta_{\tau+1} = \frac{\Phi_{\tau} + (1 - \Phi_{\tau} \rho_{\tau}(1)) / (\Phi_{\tau} - \rho_{\tau}(1)) - (\Phi_{\tau+1} - \rho_{\tau+1}(1)) / ((\Phi_{\tau} - \rho_{\tau}(1)))}{(5.13)}$$

onde

$$\rho_{\tau}(i) = \text{corr}(X_{\tau,t}, X_{\tau-i,t})$$

Na equação (5.13), θ_{τ} depende de $\theta_{\tau+1}$. Conseqüentemente, 12 equações dessa forma precisam ser resolvidas simultaneamente. Vecchia (1983) derivou os estimadores de máxima verossimilhança, que, segundo Haltiner (1985), não apresentam vantagens sobre os estimadores dados pelas equações (5.12) e (5.13).

Como no caso da modelagem de vazões anuais, nenhum modelo deverá ser adotado se a hipótese de independência do ruído Z puder ser rejeitada, por exemplo, através do teste de Portmanteau – equação (4.62).

Modelagem multivariada

A extensão, para o caso multivariado, do modelo Parma(p,q) é trivial, bastando que se utilize a opção Carma (ver texto que antecede a equação (4.76)). Caso uma transformação tenha sido aplicada às vazões mensais nos locais a e b, $M_{\tau}(a)_t$ e $M_{\tau}(b)_t$, os correspondentes ruídos, $z_{\tau}(a)_t$ e $z_{\tau}(b)_t$, serão normais serialmente independentes mas especialmente dependentes. Essa dependência espacial é então modelada por

$$Z_{\tau,t} = B_{\tau} V_{\tau,t} \quad , \quad \tau = 1,2, \dots, 12 \tag{5.14}$$

onde $Z_{\tau,t}^T = [z_{\tau,t}^{(a)}, z_{\tau,t}^{(b)} \dots]$ é um vetor de dimensão m, igual ao número de locais sob modelagem, e $V_{\tau,t}$ é um vetor de dimensão m, de componentes independentes e com distribuição normal padrão.

A estimativa das 12 matrizes B_{τ} pode ser feita conforme abordagem já descrita no tópico *Modelagem variada*. Também conforme mencionado naquele tópico, $M_{\tau}(b)_t$ pode significar a vazão correspondente à área de drenagem total a montante do local b, ou apenas à área de drenagem incremental. Por exemplo, se o local b for situado a jusante do local a, a série histórica $M_{\tau}(b)_t$ será definida, optando-se pela representação incremental, como a diferença entre as vazões totais em b e a.

5.4 Modelo de desagregação

Modelo de Valencia e Schaake

Nesta alternativa, desagregam-se as vazões anuais, geradas por qualquer modelo, em vazões mensais (ou bimensais, trimestrais etc.). A versão original foi proposta por Valencia e Schaake (1973).

$$M_t = A Q_t + B Z_t \quad (5.15)$$

onde

$$M_t = \begin{bmatrix} M_{1,t} \\ M_{2,t} \\ \vdots \\ M_{12,t} \end{bmatrix} \quad (12 \times 1)$$

$M_{\tau,t}$ = vazão no ano t mês τ menos μ_τ

μ_τ = valor esperado da vazão no mês τ (note que $E(M_{\tau,t}) = 0$)

Q_t = vazão no ano t menos μ

$$\mu = \frac{1}{12} \sum_{\tau} \mu_\tau \quad (\text{note que } E(Q_t) = 0)$$

$$Z_t = \begin{bmatrix} Z_{1,t} \\ Z_{2,t} \\ \vdots \\ Z_{11,t} \end{bmatrix} \quad (11 \times 1)$$

$Z_{i,t}$ = componente independente padrão ou ruído.

Por construção, $E(Z_{i,t}) = 0$, $\text{var}(Z_{i,t}) = 1$ e $\text{corr}(Z_{i,t}, Z_{j,t}) = 0$, ($i \neq j$)

A = vetor de parâmetros

12x1

B = matriz de parâmetros

12x11

No trabalho original de Valencia e Schaake (1973), o vetor Z_t foi definido com 12 componentes em vez de 11. Entretanto, como a média das vazões mensais deve igualar a vazão anual, só existem 11 graus de liberdade. Isto é, a 12ª vazão é perfeitamente determinada uma vez que se conheçam as outras 11 e a vazão anual. Conseqüentemente, no esquema original de Valencia e Schaake a matriz B (de dimensão 12x12) resultava singular.

Faz-se a estimativa das matrizes A e B procurando preservar as covariâncias entre as vazões mensais, definidas pela matriz $E(M_t M_t^T)$ de dimensão 12x12, e as

covariâncias entre a vazão anual, e cada uma das vazões mensais, $E(M_t Q_t)$ de dimensão 12x1. Convém lembrar que todas as variáveis têm valor esperado nulo.

Multiplicando-se ambos os lados da equação (5.15) por Q_t , achando-se o valor esperado e omitindo-se o índice t (por simplicidade de notação), obtém-se:

$$E(MQ) = AE(QQ) + B^T E(ZQ) \quad (5.16)$$

Como Q é gerado antes de Z e os componentes de Z são sorteados independentemente, $E(ZQ)$ é o vetor nulo. Portanto,

$$A = E(MQ) E^{-1}(QQ) \quad (5.17)$$

Como $E(QQ)$ é a variância da vazão anual, uma outra maneira de escrever a equação acima é

$$A = (\sigma_Q^2)^{-1} E(MQ) \quad (5.18)$$

Multiplicando-se ambos os lados da equação (5.15) por M^T e achando-se o valor esperado, obtém-se

$$E(MM^T) = AE(QM^T) + BE(ZM^T) \quad (5.19)$$

Substituindo a equação (5.15) em (5.19), vem

$$E(MM^T) = AE(QQ) A^T + AE(QZ^T) B^T + BE(ZQ) A^T + BE(ZZ^T) B^T \quad (5.20)$$

Como $E(ZZ^T) = I$, tem-se que

$$BB^T = E(MM^T) - \sigma_Q^2 A A^T \quad (5.21)$$

Do mesmo modo, substituindo (5.17) em (5.21) obtém-se

$$BB^T = E(MM^T) - (\sigma_Q^2)^{-1} E(MQ) E^T(MQ) \quad (5.22)$$

A estimativa das matrizes A e B é feita substituindo-se as covariâncias de população nas equações (5.17) - ou (5.18) - e (5.21) - ou (5.22) - pelos correspondentes valores amostrais. Por esse método, a média das vazões mensais sintéticas será sempre igual à correspondente vazão anual sintética, como a seguir. Se

$$\hat{E}(MM^T)_{ij} = \frac{1}{n} \sum_{t=1}^n m_{i,t} m_{j,t} \quad (5.23)$$

$$\hat{E}(MQ)_i = \frac{1}{n} \sum_{t=1}^n m_{i,t} q_t \quad (5.24)$$

$$\hat{\sigma}_Q^2 = \frac{1}{n} \sum_{t=1}^n q_t^2 \quad (5.25)$$

the demonstrator

onde n é o número de anos da série histórica.

Seja o vetor linha com 12 elementos

$$D = [12^{-1}, 12^{-1}, \dots, 12^{-1}] \quad (5.26)$$

Da equação (5.18), tem-se que

$$D\hat{A} = (\hat{\sigma}_Q^2)^{-1} D\hat{E}(MQ) \quad (5.27)$$

Substituindo as equações (5.24) e (5.25) em (5.27), vem

$$D\hat{A} = 1 \quad (5.28)$$

resultado que será utilizado mais adiante.

Da equação (5.22), tem-se que

$$D\hat{B}\hat{B}^T D^T = D\hat{E}(MM^T)D^T - (\hat{\sigma}_Q^2)^{-1} D\hat{E}(MQ) \hat{E}^T(MQ) D^T \quad (5.29)$$

Substituindo as equações (5.23) e (5.25) em (5.29), obtém-se

$$D\hat{B}\hat{B}^T D^T = (\hat{\sigma}_Q^2) - (\hat{\sigma}_Q^2)^{-1} \hat{\sigma}_Q^2 \hat{\sigma}_Q^2 = 0 \quad (5.30)$$

Seja

$$\hat{C} = D\hat{B} = [\hat{C}_1, \hat{C}_2, \dots, \hat{C}_{11}] \quad (5.31)$$

1x11

A equação (5.30) pode ser reescrita como

$$D\hat{B}\hat{B}^T D^T = \sum_{i=1}^{11} C_i^2 = 0 \quad (5.32)$$

Portanto,

$$C_i = 0, \forall i \quad (5.33)$$

Multiplicando-se (5.15) por D, obtém-se

$$DM_t = D\hat{A} Q_t + D\hat{B}Q_t \quad (5.34)$$

Usando-se (5.28), (5.31) e (5.33) tem-se

$$DM_t = Q_t \quad (5.35)$$

A equação acima garante que as médias anuais das vazões mensais geradas pelo modelo serão exatamente iguais às vazões anuais.

De acordo com (5.21), B é calculada por:

$$\hat{B}\hat{B}^T = \hat{N} \quad (5.36)$$

onde N é uma matriz 12x12 obtida das matrizes de covariância históricas. Como DB = 0, N tem ordem 11 e é uma matriz positiva semidefinida. Por este motivo a equação (5.36) só pode ser resolvida por decomposição espectral de N:

$$\hat{B} = P\lambda^{1/2} \quad (5.37)$$

onde P é a matriz 12x11 de autovetores de N e λ é a matriz diagonal de autovalores não nulos ordenados em ordem decrescente.

Usando (5.18), a equação de estimativa para A é:

$$\hat{A} = (\hat{\sigma}_Q)^{-1} \hat{E}(MQ) \quad (5.38)$$

O modelo de Valencia e Schaake – equação (5.15) – não é capaz de reproduzir a correlação entre as vazões de dezembro de um ano e janeiro do ano seguinte (defasagem 1). Na realidade, essa correlação nas vazões sintéticas tenderá a ser igual à correlação entre as vazões janeiro do ano t e dezembro do ano t+1, uma defasagem de 23 meses (Kelman *et al*, 1979).

Mejia e Rousselle (1976) apresentaram uma extensão do modelo de Valencia e Schaake com o propósito de sanar essas inconveniências:

$$M_t = A Q_t + C M_{t-1}(k) + B Z_t \quad (5.39)$$

onde

$$M_{t-1}(k) = \begin{matrix} M_{12,t-1} \\ \vdots \\ M_{13-k,t-1} \end{matrix}$$

é o vetor das últimas k vazões do ano anterior e C = matriz de parâmetros 12xk

e os demais elementos da equação são definidos como na equação (5.15). A estimativa das matrizes A, B e C se faz de forma análoga ao procedimento adotado no modelo de Valencia e Schaake, procurando-se preservar as covariâncias amostrais: $\hat{E}(M_t M_t^T)$, $\hat{E}(M_t Q_t)$, $\hat{E}(M_{t-1}(k) Q_t)$, $\hat{E}(M_t M_{t-1}^T(k))$. O método de cálculo sugerido por Mejia e Rousselle não é exato (Kelman *et al*, 1979; Lane 1980), mas fornece resultados satisfatórios.

Caso $Z_{i,t}$ – equações (5.15), (5.34) e (5.39) – seja sorteada de uma distribuição normal padrão, tanto a vazão anual sintética quanto as vazões mensais sintéticas terão também distribuição normal. Conforme visto no item 4.4, esta propriedade pode ser aceitável para as vazões anuais. Entretanto, a distribuição de vazões mensais exige em



geral um coeficiente de assimetria suficientemente grande para que se rejeite a hipótese de normalidade. Existem pelo menos três alternativas para o tratamento desse problema:

- a) método não paramétrico;
- b) uso de funções que transformem vazões mensais em variáveis normalmente distribuídas;
- c) uso de distribuição assimétrica para o ruído Z.

Como a última alternativa apresenta alguns problemas numéricos (Kelman *et al.*, 1979; Todini, 1980), apenas as alternativas a) e b) serão descritas a seguir.

Método não paramétrico

Nesta abordagem, a preservação da distribuição marginal das vazões mensais se processa indiretamente através do uso de ruídos históricos na geração de vazões sintéticas (Kelman *et al.*, 1981; Pereira *et al.*, 1984). Ruídos históricos são calculados resolvendo-se a equação (5.15) ou (5.39), para Z_t , sendo que M_t e Q_t correspondem a valores observados da série histórica. Seja, por exemplo, a equação (5.15). Rearrmando-a, tem-se

$$B\hat{Z}_t = AQ_t - M_t \tag{5.40}$$

Se B fosse uma matriz quadrada, a solução da equação acima, para Z_t , seria obtida multiplicando em primeiro lugar ambos os lados da equação pela inversa de B. Como este não é o caso, convém inicialmente multiplicar por B^T , obtendo-se do lado esquerdo B^TB , esta sim uma matriz quadrada de dimensão 11 x 11. Conseqüentemente,

$$\hat{Z}_t = (B^TB)^{-1} B^T(AQ_t - M_t) \tag{5.41}$$

Através da equação anterior, 11 ruídos históricos são obtidos para cada um dos n anos da série histórica. Na fase de geração, um ano da série sintética é gerado sorteando-se cada um dos 11 ruídos, independentemente da correspondente coleção de n ruídos históricos. O Quadro 52 ilustra o processo: $Z_{i,t}$ representa o i-ésimo componente do vetor de ruídos históricos no ano t (i=1, 2, ..., 11; t=1, 2, ..., n). Esse ruído é indexado por I_t , que é sorteado de uma distribuição uniforme. Sorteios independentes de I_1, I_2, \dots, I_{11} produzem um vetor de ruídos Z_t a ser utilizado na geração de M_t . O número total de possíveis vetores de ruídos é n^{11} , que é uma população estritamente numerosa, tomando valores típicos de n entre 10 e 50.

O uso dos ruídos empíricos constitui um método não paramétrico, porque nenhuma distribuição (com parâmetros) é estipulada para Z_t . É comum encontrar-se métodos não paramétricos incapazes de gerar vazões sintéticas que sejam maiores ou menores do que, respectivamente, o máximo e o mínimo da série histórica. Este não é, felizmente o caso do método anteriormente proposto. Entretanto, neste esquema de geração, assim como no proposto por Valencia e Schaake, pode suceder que algumas das vazões sintéticas sejam negativas. Algumas maneiras de corrigir vazões negativas são discutidas logo adiante.

Quadro 52. Ruídos empíricos obtidos de n anos de séries históricas

Ano da série histórica	Primeiro componente	Segundo componente	(...)	Décimo-primeiro componente
1	$Z_{1,1}$	$Z_{2,1}$		$Z_{11,1}$
2	$Z_{1,2} \leftarrow I_1$	$Z_{2,2}$		$Z_{11,2}$
.	.	.	(...)	.
.	.	.		$\leftarrow I_{11}$
.	.	.		.
n	$Z_{1,n}$	$Z_{2,n}$	$\leftarrow I_2$	$Z_{11,n}$

Uso de variáveis transformadas

Stedinger *et al.*, (1985) argumentaram que um modelo de desagregação de vazões anuais em mensais necessita preservar apenas as seguintes propriedades:

- a) distribuição marginal das vazões mensais;
- b) correlação entre vazões mensais consecutivas;
- c) vazão anual previamente gerada.

O leitor deve notar que o modelo de desagregação de Valencia e Schaake é mais ambicioso, na medida em que procura preservar a correlação entre vazões mensais não necessariamente consecutivas. Se as vazões fossem normalmente distribuídas, um esquema de geração originalmente proposto por J.D.Salas (Grygier e Stedinger, 1987) poderia ser empregado para satisfazer os objetivos propostos por Stedinger *et al.* (1985).

$$M_{1,t} = a_1 + b_1 Q_t + C_1 V_{1,t} \tag{5.42}$$

$$M_{2,t} = a_2 + b_2 Q_t + C_2 V_{2,t} + d_2 M_{1,t} \tag{5.43}$$

$$M_{\tau,t} = a_{\tau} + b_{\tau} Q_t + C_{\tau} V_{\tau,t} + d_{\tau} M_{\tau-1,t} + l_{\tau} \sum_{i=1}^{\tau-1} M_{i,t}, \tau \geq 3 \tag{5.44}$$

onde $V_{\tau,t}$ é um ruído normal padrão.

É evidente que este modelo deve gerar vazões mensais cuja média é a vazão previamente gerada, uma vez que, para T=12, a equação degenera em $a_{\tau} = 0$, $b_{\tau} = 12$, $C_{\tau} = 0$, $d_{\tau} = 0$ e $l_{\tau} = -1$. Isto é



$$M_{12,t} = 12Q_t - \sum_{\tau=1}^{11} M_{\tau,t} \quad (5.45)$$

Stedinger *et al.*, (1985) propuseram utilizar um modelo com a forma das equações (5.42) a (5.44), porém substituindo a vazão anual Q_t e as mensais $M_{\tau,t}$ pelas correspondentes transformadas $Y_t = g(Q_t)$ (já utilizada no item 4.4) e $X_{\tau,t} = g_{\tau}(M_{\tau,t})$. Os autores propuseram ainda que o termo $\sum_{i=1}^{\tau-1} M_{i,t}$ deveria ser substituído pela sua aproximação de primeira ordem (ignorando um termo constante), $\sum_{i=1}^{\tau-1} w_i X_{i,t}$, onde

$$w_i = E \left(\frac{d M_{i,t}}{d X_{i,t}} \right) \quad (5.46)$$

Isto é, o modelo proposto por Stedinger *et al.*, (1985) é da forma

$$X_{1,t} = a_1 + b_1 Y_t + C_1 V_{1,t} \quad (5.47)$$

$$X_{2,t} = a_2 + b_2 Y_t + C_2 V_{2,t} + d_2 X_{1,t} \quad (5.48)$$

$$X_{\tau,t} = a_{\tau} + b_{\tau} Y_t + C_{\tau} V_{\tau,t} + d_{\tau} X_{\tau-1,t} + 1_{\tau} \sum_{i=1}^{\tau-1} w_i X_{i,t}, \quad \tau \geq 3 \quad (5.49)$$

Dependendo da transformação $g_{\tau}(\cdot)$ utilizada no mês τ , os coeficientes w_{τ} são calculados pela equação (5.46). O Quadro 53 fornece os resultados para as principais distribuições de probabilidades, usualmente adotadas em hidrologia para modelar vazões mensais.

As equações (5.47) a (5.49) são, na realidade, um conjunto de regressões de $X_{\tau,t}$ em $Y_t, X_{\tau-1,t}$ e $\sum_{i=1}^{\tau-1} w_i X_{i,t}$. Uma vez achada a série histórica correspondente a cada uma dessas variáveis, a estimativa dos parâmetros se processa segundo o procedimento usual adotado em regressão multivariada (Apêndice B). Como as transformações $g_{\tau}(\cdot)$ não são necessariamente lineares, não há garantia de que a média das vazões mensais sintéticas se iguale à vazão anual sintética previamente gerada. Ajustes das vazões mensais serão conseqüentemente desejáveis.

Uma alternativa de modelagem, dentro da opção pelo uso de variáveis transformadas, consiste em abandonar a vazão transformada $Y=g(Q)$, utilizada no item 4.4 para produzir série sintética anual, e definir Y como

$$y_t = 12^{-1} \sum_{\tau} X_{\tau,t} \quad (5.50)$$

Uma vez decidida a transformação $g_{\tau}(\cdot)$ de cada mês, obtém-se uma série temporal dos processos mensal e anual, respectivamente X e Y , que se relacionam analogamente à equação (5.15).

Quadro 53. Valores de W_{τ} para diferentes distribuições de probabilidades a serem ajustadas às vazões do mês τ

Distribuição de $M_{\tau,t}$	Número de parâmetros	$X_{\tau,t}$	$W_{\tau} = E(dM_{\tau,t}/dX_{\tau,t})$
Normal	2	$(M - \mu^M)/\sigma^M$	σ^M
Log-normal	2	$(\log(M) - \mu^N)/\sigma^N$	$\sigma^N \exp(\mu^N + (\sigma^N)^2)$
Log-normal	3	$(\log(M - a) - \mu^N)/\sigma^N$	$\sigma^N \exp(\mu^N + (\sigma^N)^2)$
Gamma	3	$6 (\gamma/2 ((M - \mu^M)/\sigma^M + 2\gamma)^{1/3} - 1 + \gamma^2/36)/\gamma$	$\sigma^M (1 - \gamma^2/36 + (\gamma/36)^2)$
Potencia de normal	4	$(W - \mu^B)/\sigma^B$	$(\sigma^B/\lambda)(\lambda\mu^B + 1)(1-\lambda)/\lambda (1 + (\sigma^B)^2(1-\lambda)(1-2\lambda)(\lambda\mu^B + 1)^{-2})$

- 1) $X_{\tau,t}$ é normal padrão
- 2) μ^M, μ^N e μ^B são notações simplificadas para μ_{τ}, μ_{τ}^N e μ_{τ}^B , que são as médias amostrais, respectivamente, de $M_{\tau,t}, \log(m_{\tau,t})$ e $W_{\tau,t}$.
- 3) $W_{\tau,t} = \lambda_{\tau}^{-1} (\lambda_{\tau}^{\tau} - 1)$, $\lambda_{\tau} \neq 0$
- 4) λ é notação simplificada para λ_{τ} , a assimetria amostral de $M_{\tau,t}$
- 5) O valor de W_{τ} para a distribuição potencia de normal foi obtida aproximadamente, por uma expansão em série de Taylor em torno de $F(X) = 0$.

$$X_t = AY_t + BZ_t \tag{5.51}$$

onde Z_t é um vetor 11×1 de ruídos. Como na equação (5.15), $E(Z_{1,t}) = 0$, $var(Z_t) = 1$ e $corr(Z_{1,t}, Z_{j,t}) = 0$ ($i \neq j$). Em acréscimo à definição de Z dada na equação (5.15), exige-se agora que $Z_{i,t}$ também seja normalmente distribuído.

A estimativa do vetor A e da matriz B é feita de acordo com o procedimento já descrito quando se tratou do modelo Valencia e Schaake. A geração de uma série sintética se faz trivialmente, uma vez que os componentes de Z_t são ruídos normais e independentes. Cada vazão mensal é obtida, usando-se a transformação inversa $M_{\tau,t} = g^{-1}(X_{\tau,t})$.

Quando se adota o modelo definido pelas equações (5.50) e (5.51), o processo a ser gerado em escala anual não é mais uma função da vazão anual e sim das vazões mensais.

Correção das vazões sintéticas

Conforme visto nos tópicos anteriores, dependendo da opção de modelagem, a média das vazões sintéticas mensais pode ser diferente da vazão sintética anual previamente gerada ou as vazões sintéticas mensais podem ser negativas. Em ambos os casos convém filtrar a série sintética com o objetivo de corrigir essas distorções. Kelman *et al.* (1981) e Pereira *et al.* (1984) sugeriram que vazões mensais corrigidas podem ser obtidas achando-se a solução para o seguinte problema de otimização:

$$\min \sum_{\tau,t} (M_{\tau,t}^* - M_{\tau,t})^2 / \sigma_{\tau}^2 \tag{5.52}$$

sujeito a $M_{\tau,t} \geq 0$ (5.53)

$$\sum_{\tau} M_{\tau,t}^* = Q_t \tag{5.54}$$

A função objetivo (5.52) procura fazer com que os ajustes sejam proporcionais ao desvio padrão da vazão em cada mês. Isto é, meses mais variáveis sofrem ajustes maiores. A restrição (5.53) impõe que as vazões sejam não negativas e a restrição (5.54) impõe que a média das vazões mensais seja igual à vazão anual. Um algoritmo eficiente para a solução do problema de otimização é descrito por Lauson e Hanson (1974). Convém observar que muitos valores negativos de $M_{\tau,t}$ resultam em valores $M_{\tau,t}^*$ nulos corrigidos, o que nem sempre é um resultado satisfatório.

Grygier e Stedinger (1987a) sugerem que, se as vazões mensais forem log-normalmente distribuídas, isto é, se

$$M_{\tau,t} = \exp(X_{\tau,t}), \forall \tau \tag{5.55}$$

não só $M_{\tau,t}$ será positivo como ficará simples conceber uma correção que resulte em valores também positivos:

$$M_{\tau,t}^* = \exp(\psi s_{\tau}) \exp(X_{\tau,t}) \tag{5.56}$$

onde s_{τ} é o desvio padrão de $X_{\tau,t}$. Nesta abordagem também os meses mais variáveis sofrem ajustes maiores. O valor de ψ é determinado por um método iterativo (*Newton-Raphson*, por exemplo), de forma a satisfazer a equação (5.54).

Algumas outras correções sugeridas na literatura são revistas por Grygier e Stedinger (1987a), entre elas o procedimento elementar de fazer com que as correções sejam proporcionais às vazões sintéticas $M_{\tau,t}$. Isto é

$$M_{\tau,t}^* = a M_{\tau,t} \tag{5.57}$$

$$a = Q_t / \sum_{\tau} M_{\tau,t} \tag{5.58}$$

O inconveniente desta abordagem é que as vazões em meses com pequena variabilidade são ajustadas tanto quanto as vazões em meses com grande variabilidade. Ademais, se $M_{\tau,t}$ for negativo, $M_{\tau,t}^*$ também o será.

Caso multivariado

Existem inúmeras possibilidades de se aplicar o método de desagregação para o caso multivariado. A mais óbvia é utilizar o próprio modelo de Valencia e Schaake, que, aliás, foi originalmente proposto para o caso multivariado. Basta que a vazão anual na equação (5.15), o escalar Q_t , seja substituído pelo vetor $Q_t^T = [Q_{t,1}, Q_{t,2}, \dots, Q_{t,k}]$, onde k é o número de locais sob modelagem. Naturalmente, o número de elementos nas matrizes M_t e A , assim como cada uma das dimensões da matriz B , seriam também multiplicados por k . É fácil perceber que o número de parâmetros do modelo, componentes do vetor A e da matriz B , pode se tornar muito grande, dependendo do valor k . Isto naturalmente não é desejável porque, de uma maneira geral, quanto mais parâmetros um modelo tiver mais as flutuações amostrais da série histórica serão preservadas nas séries sintéticas. Mesmo para valores moderados de k , é bem possível que o número total de parâmetros exceda o número de observações disponíveis na série histórica multivariada, o que obviamente implica num ajuste não confiável. As Figuras 125 e 126 mostram duas alternativas de modelagem que mantêm o número de parâmetros dentro de limites razoáveis.

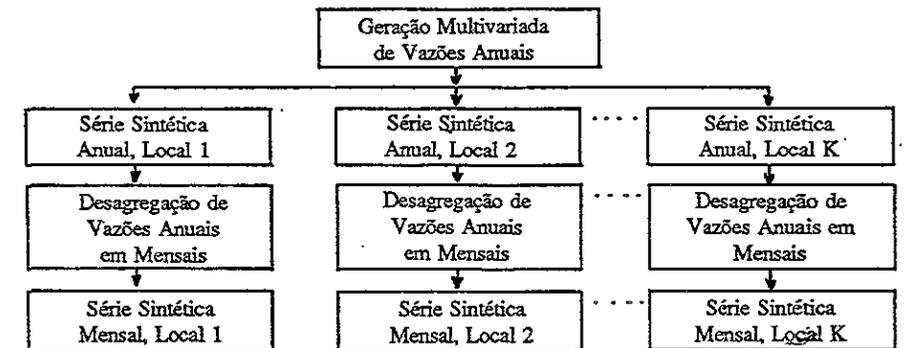


Figura 125. Alternativa I de geração multivariada de vazões mensais

$$\tilde{y}_t = \bar{y} + \sum_{k=1}^K (a_k \cos(2\pi k t/m) + b_k \sin(2\pi k t/m)) \quad (6.2)$$

$$\bar{y} = \frac{1}{m} \sum_{j=1}^m y_j \quad (6.3)$$

$$a_k = \frac{2}{m} \sum_{j=1}^m (y_j - \bar{y}) \cos(2\pi jk/m) \quad (6.4)$$

$$b_k = \frac{2}{m} \sum_{j=1}^m (y_j - \bar{y}) \sin(2\pi jk/m) \quad (6.5)$$

e \tilde{y}_j significa tanto \tilde{y}_j quanto $\tilde{\sigma}_j$. O "número de harmônicos" K necessita ser escolhido com alguma cautela. Quanto maior ele for, mais \tilde{y}_t se aproximará da "função patamar" (no caso em que não se utilizem frações, $m = 365$, quanto maior for K tanto mais \tilde{y}_t se aproximará de \bar{y}). Por outro lado, deseja-se utilizar \tilde{y}_t e não $\tilde{\sigma}_t$, exatamente para suavizar as flutuações que traduzam apenas variação amostral (princípio da parcimônia). Yevjevich (1972) descreve um método estatístico para escolher o melhor valor de K.

Tao (1973) tentou ajustar diversas expressões analíticas para as distribuições empíricas de Z_t de muitos rios norte-americanos, sendo mal sucedido devido à alta assimetria e curtose das distribuições empíricas. Este fato, por si só, já ilustra que quanto menor for o intervalo de discretização, mais difícil é modelar o correspondente processo estocástico. A complexidade do problema é adicionada pela existência de uma propriedade na série histórica que não estaria presente nas séries sintéticas produzidas pela equação (6.1): a ascensão da hidrógrafa é, em geral, mais abrupta do que a recessão. Processos cujas trajetórias exibam esta propriedade são chamados de "irreversíveis no tempo". Formalmente, um processo $x(t)$ é reversível, como é o caso de Arma (p,q), quando os conjuntos de variáveis aleatórias (X_1, X_2, \dots, X_k) e $(X_{n-1}, X_{n-2}, \dots, X_{n-k})$ identicamente distribuídos. Esse fato é possivelmente argumento suficiente para desestimular o emprego de modelos Arma (p,q) para o estudo de vazões diárias.

Modelo "shot noise"

Diversos pesquisadores têm procurado formular modelos estocásticos que possuam as propriedades observáveis em séries de vazões discretizadas em curtos intervalos, dia ou hora, através da propagação da precipitação através de um ou mais reservatórios lineares. A representação da bacia hidrográfica como uma cascata de reservatórios lineares permitiu a Nash (1957) a derivação da hidrógrafa unitária instantânea com forma matemática idêntica à função densidade de probabilidade gama. O ramo estocástico dessa linha de investigação desenvolveu-se basicamente a partir da década de 1970 (Bodo e Unny, 1987), através da modelagem estocástica da precipitação. O modelo *shot noise* (Bernier, 1970 e Weiss, 1977) é o mais conhecido membro da família, para aplicação em vazões diárias.

Suponha a bacia hidrográfica representada por um único reservatório linear

$$q(t) = b s(t) \quad (6.6)$$

onde

- q (t) – vazão instantânea (saída no reservatório linear)
- s (t) – armazenamento instantâneo
- b – coeficiente de armazenamento.

a equação de continuidade é

$$\frac{d(s(t))}{dt} = x(t) - q(t) \quad (6.7)$$

onde

x (t) – vazão de entrada instantânea no reservatório linear.

Substituindo (6.6) em (6.7)

$$\frac{dq(t)}{dt} + bq(t) = bx(t) \quad (6.8)$$

Integrando a equação (6.8) para um pulso unitário instantâneo $x(t) = 1, x(t) = 0, \forall t \neq 0$, resulta em

$$q_H(t) = \frac{\exp(-bt)}{b} \quad (6.9)$$

sendo $g_H(t)$ a hidrógrafa unitária instantânea.

Em geral, para uma entrada na bacia-reservatório-linear contínua no tempo, $x(t)$, tem-se a vazão de saída calculada pela equação de convolução (o sistema é linear):

$$q(t) = \int_0^t x(\tau) q_H(t-\tau) d\tau \quad (6.10)$$

Caso a entrada na bacia até o instante t ocorra como uma sucessão de n pulsos instantâneos de precipitação $x(t_1), x(t_2), \dots, x(t_n), t_1 < t_2 < \dots < t_n < t$, a integral (6.10) degenera no somatório

$$q(t) = \sum_{i=1}^n x(t_i) \exp(-b(t-t_i)) \quad (6.11)$$

Suponha que a ocorrência de pulsos seja um processo estocástico, chamado de processo Poisson, que satisfaz as seguintes propriedades:

- a) a probabilidade de ocorrência de um pulso num pequeno intervalo Δt é proporcional a Δt ;
- b) a probabilidade de ocorrência de mais de um pulso num pequeno intervalo Δt é nula; e

c) a ocorrência de um pulso num intervalo Δt é um evento independente do que aconteça fora do intervalo.

Seja W a variável aleatória que representa o tempo de espera entre dois pulsos.

Então:

$$f_W(w) = v \exp(-w) \quad (6.12)$$

ou

$$F_W(w) = 1 - \exp(-vw) \quad (6.13)$$

Ou seja, o tempo de espera num processo Poisson é uma variável aleatória com distribuição exponencial. Se o parâmetro v for conhecido, a geração de intervalos sintéticos w_1, w_2, \dots é tarefa bastante simples, porque $U = F_W(W)$ tem distribuição uniforme $(\theta, 1)$: números uniformemente distribuídos u_1, u_2, \dots seriam sorteados, e os correspondentes valores w_1, w_2, \dots calculados por

$$w_i = v^{-1} \ln(1 - u_i) \quad (6.14)$$

Seja $N(t)$ o número de pulsos no intervalo $(0, t)$. Obviamente, $N(t)$ é uma variável aleatória inteira, cuja distribuição de probabilidades (conhecida como Poisson) pode ser derivada da expressão

$$P(N(t) \geq n) = P(W_1 + W_2 + \dots + W_n < t) \quad (6.15)$$

A versão estocástica da equação (6.11) é

$$Q(t) = \sum_{i=1}^{N(t)} X(t_i) \exp(-b(t - T_i)) \quad (6.16)$$

onde

$$T_i = t_{i-1} + W_i, \quad t_0 = 0 \quad (6.17)$$

A utilização da equação (6.14) em modo de geração passa pela definição probabilística dos pulsos instantâneos de precipitação $x(t_1), x(t_2), \dots$. Diversos pesquisadores acharam que a altura da precipitação medida num intervalo de discretização muito pequeno é uma variável aleatória com distribuição exponencial (por exemplo, Todorovic e Woolhiser, 1975). Isto é

$$F_X(x) = 1 - \exp(-\theta x) \quad (6.18)$$

Se o parâmetro $\theta(\theta^{-1} = E(X))$ for conhecido, a geração das magnitudes dos pulsos sintéticos se faz de forma análoga à equação (6.14). É possível mostrar que a distribuição de probabilidades de $Q(t)$ é, nesse caso, gama.

As equações (6.13), (6.16), (6.17) e (6.18) definem completamente o processo estocástico *shot noise* e, conforme isto, a geração de uma realização contínua no tempo é bastante simples, desde que os parâmetros (b, v, θ) sejam conhecidos. A vazão média no dia i é

$$Q_i = \frac{1}{\delta} \int_{(i-1)\delta}^{i\delta} Q(s) ds \quad (6.19)$$

onde δ corresponde ao intervalo de um dia.

Weiss (1977) mostrou que Q_i tem valor esperado, variância e coeficiente de autocorrelação respectivamente iguais a:

$$E(Q) = \frac{v}{b\theta} \quad (6.20)$$

$$\text{var}(Q) = \frac{2v(b - (1 - e^{-b}))}{b^3 \theta^2} \quad (6.21)$$

$$\text{Corr}(Q_i, Q_{i+1}) = (1 - \exp(-b))^2 / (2(b - (1 - \exp(-b)))) \quad (6.22)$$

O uso do modelo, na prática, pode ser feito dividindo-se o ano em m frações quase estacionárias. Por exemplo, em $m = 26$ frações de 14 dias (para um ano de 364 dias). Calcula-se a média, desvio padrão e coeficiente de autocorrelação amostral para cada fração quase estacionária, o que permite a resolução das equações (6.20) a (6.22) para os parâmetros (b, v, θ) .

Alternativamente, Weiss também utilizou funções periódicas para representar a variação ao longo do ano da média e do desvio padrão das vazões diárias, segundo a equação (6.2), o que permite estimar as funções temporais \bar{v}_t e $\bar{\theta}_t$. Weiss assumiu um valor constante para b , por traduzir uma propriedade da bacia hidrográfica, e não uma propriedade do processo meteorológico.

A aplicação do modelo *shot noise* em rios ingleses não apresentou resultados satisfatórios (Weiss, 1977). Para melhorar os resultados, Weiss desenvolveu o modelo *shot noise* de Segunda Ordem, que segue uma abordagem similar à acima descrita, com a diferença de que a bacia hidrográfica é modelada por dois reservatórios em paralelo, um representando o escoamento superficial e outro representando o escoamento básico.

Modelo Diana

Kelman (1976, 1980) também propôs um modelo com a propriedade da irreversibilidade temporal. Repartindo em três a vazão total em um dia t , Kelman procurou modelar o fenômeno segundo a representação esquemática da Figura 127.

$$Q_t = U_t + V_t + W_t \quad (6.23)$$

onde U_t representa o escoamento superficial, V_t representa o escoamento subsuperficial e W_t representa o escoamento básico. Essas duas últimas parcelas são supostas oriundas do esvaziamento de dois reservatórios fictícios, que representam as capacidades de armazenamento da bacia, respectivamente de pequena e de grande inércias.

Para modelar o processo U_t , Kelman adotou um procedimento também útil para modelar precipitação, que, assim como o escoamento superficial, é um processo intermitente. A abordagem é melhor compreendida se exposta dentro da seqüência de procedimento utilizada para sintetizar uma realização de U_t :

- i) Inicialmente números aleatórios independentes com distribuição normal padrão são produzidos, a_t .
- ii) A persistência que existe no escoamento superficial, assim como também na precipitação, devido, por exemplo, ao estacionamento de uma frente fria sobre a bacia, é introduzida por

$$z_t = m(i) + \frac{s(i)}{s(i-1)} r(i)[z_{t-1} - m(i-1)] + s(i) [1 - r^2(i)]^{1/2} a_t \quad (6.24)$$

onde $m(i)$, $s(i)$ e $r(i)$ são funções periódicas a serem estimadas, $i=1, 2, \dots, n$, - equações (6.2) a (6.5) - e n é o número de em que o ano é dividido.

- iii) O caráter intermitente do escoamento superficial (massa de probabilidade concentrada no valor de $U_t=0$) é obtida pela seguinte censura

$$Y_t = \max(0, Z_t) \quad (6.25)$$

- iv) Um grau de flexibilidade para o ajuste da distribuição marginal para os valores positivos é obtido por

$$U_t = Y_t^{d_t(i)} \quad (6.26)$$

As funções periódicas $m(i)$, $s(i)$, $r(i)$ e $d(i)$ podem ser estimadas por uma aproximação do método da máxima verossimilhança (Kelman, 1976), desde que uma série histórica para U_t esteja disponível. No caso de precipitação, isto não constitui um problema, visto que a série intermitente é medida diretamente. Já no caso de escoamento superficial, tudo o que se conhece é a série histórica correspondente à vazão total Q_t , sendo, portanto, necessária alguma hipótese que permita a definição de U_t . A opção feita foi

$$U_t = \max(0, Q_t - Q_{t-1}) \quad (6.27)$$

Esta equação revela que o valor de Q_t é simplesmente igual a $Q_{t-1} + U_t$, desde que U_t seja positivo. Se U_t for nulo, a hidrógrafa estará em recessão e Q_t será obtido somando-se o esvaziamento dos reservatórios. Admitindo que esses reservatórios são lineares (vazão de saída proporcional ao estado de armazenagem), tem-se

$$\begin{aligned} Q_t &= Q_{t0} [x \exp(-k_1(t - t_0)) + (1-x) \exp(-k_2(t-t_0))] \\ U_t &= 0 \end{aligned} \quad (6.28)$$

onde t_0 é o dia em que começa a recessão, k_1 e k_2 são as constantes de proporcionalidade para os dois reservatórios e x é uma variável aleatória com domínio entre 0 e 1, que define a repartição do escoamento entre os dois reservatórios.

Os valores de k_1 e k_2 , assim como uma amostra para x , podem ser obtidos minimizando-se a soma dos quadrados dos desvios entre a série histórica e a série

reconstituída pelo modelo. É conveniente trabalhar com a distribuição de probabilidades de x condicionada no valor de Q_{t0} . Isto porque se Q_{t0} for pequeno, aumentam as chances de que x esteja perto de 0, significando que a descarga básica será contribuição principal para a manutenção do escoamento total.

O modelo desenvolvido por Kelman vem sofrendo aperfeiçoamentos e extensões que foram substanciados num modelo chamado Diana (Kelman, Damazio e Costa, 1983, 1985). As principais diferenças entre o modelo Diana e o proposto originalmente por Kelman são:

- i) A equação (6.24) é substituída por

$$Z_t = r(i) z_{t-1} + \sqrt{1 - r^2(i)} a_t \quad (6.29)$$

isto é, assume-se que z_t é um processo padronizado.

- ii) A equação (6.25) é substituída por

$$Y_t = \max(\beta, Z_t) \quad (6.30)$$

$$\beta = \Phi^{-1}(p) \quad (6.31)$$

onde β define o intervalo de censura $(-\infty, \beta)$, $\Phi(\cdot)$ é a distribuição acumulada normal padrão e $p = P(U_t = 0)$.

- iii) A equação (6.26) é substituída por

$$F_i(U_t) = \Phi(Y_t) \quad (6.32)$$

onde $F_i(\cdot)$ é a distribuição acumulada empírica dos valores de $U_t > \theta$, para cada época i do ano, $i = 1, 2, \dots$

A equação (6.32) implica no mapeamento de cada y_t em algum U_t efetivamente medido na série histórica. Este enfoque não paramétrico parece razoável, levando-se em conta que, para cada época do ano i , a amostra de valores de $U_t > \theta$ é usualmente grande o bastante para representar a faixa de variação dessa variável aleatória.

- iv) A equação (6.27) é substituída por

$$U_t = \max(0, q_t - \Lambda q_{t-1}) \quad (6.33)$$

onde Λ é um número compreendido entre 0 e 1, chamado de taxa de recessão característica. Na equação (6.33) admite-se implicitamente a possibilidade de ocorrência de escoamento superficial, mesmo quando a hidrógrafa esteja baixando, o que é fisicamente possível.

- v) A representação da capacidade de armazenagem de bacia através de dois reservatórios lineares, que deu origem à equação (6.28), é substituída por outra que considera apenas um reservatório linear de comportamento estocástico. Assim, quando $U_t = 0$, q_t será uma fração $k_t \leq \lambda$ da vazão anterior.

$$Q_t = k_t q_{t-1}, k_t \leq \lambda; U_t = 0 \quad (6.34)$$

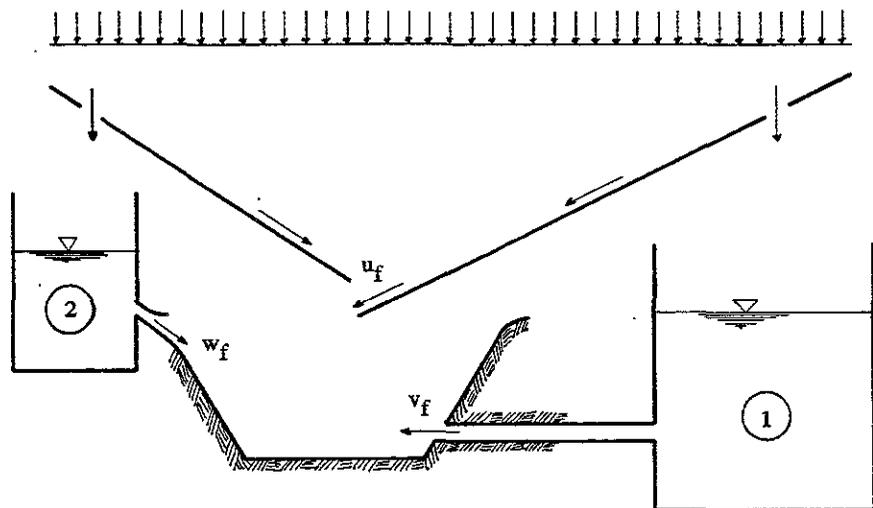


Figura 127. Representação esquemática dos componentes do modelo Kelman

Os valores de k_T são obtidos a partir da distribuição empírica dos valores de Q_t/Q_{t-1} , para valores desta razão inferior a λ , sem levar em consideração a época do ano.

O modelo Diana pode ser utilizado para a geração de vazões sintéticas em m postos que sejam dependentes entre si. Basta que a equação (6.29) seja substituída pela sua versão multivariada (item 4.4):

$$Z_t = F Z_{t-1} + G A_t \quad (6.35)$$

onde Z_t é um vetor de dimensão m , cujos componentes correspondem ao Z da equação (6.29); A_t é um vetor com m variáveis normais padronizadas independentes. O método de estimativa das matrizes F e G bem como dos demais parâmetros é descrita por Kelman, Damazio e Costa (1983).

O modelo não paramétrico

Yakowitz (1979) adotou um procedimento para a geração de séries sintéticas de vazões diárias baseado na utilização da distribuição de probabilidades empírica de q_t , condicionada nos valores imediatamente anteriores, $Q = (Q_{t-1}, Q_{t-2}, \dots, Q_{t-T})$, onde T é a memória arbitrada para o processo. Para isso, Yakowitz definiu um conjunto finito de vetores representativos $Q(1), Q(2), \dots, Q(m)$ de tal forma que qualquer vetor Q da série histórica ou gerada pudesse ser classificado numa das m categorias de acordo com a distância entre Q a cada um dos $Q(i)$, $i=1, m$. Assim, se

$$\|Q(i) - Q\| \leq \|Q(i) - Q\|, \forall i \quad (6.36)$$

diz-se que o vetor Q tem distância mínima para o j -ésimo vetor representativo. Neste caso, diz-se que a vazão do próximo dia - Q_t pertence ao conjunto C_j , $j = 1, m$.

Classificando-se todos os valores q_t da série histórica de acordo com o procedimento descrito, é possível caracterizar a distribuição de probabilidades empíricas associada a cada conjunto C_j , definida por

$$F_j(x) = \frac{k_j}{n_j} \quad (6.37)$$

onde k_j é o número de elementos do conjunto C_j menores ou iguais a x ; e n_j é o número total de elementos do conjunto C_j .

A geração de uma série sintética se processa sorteando-se sucessivamente variáveis aleatórias independentes uniformemente distribuídas entre 0 e 1, w_t , sendo Q_t dado por

$$Q_t = F_j^{-1}(w_t) \quad (6.38)$$

onde o valor de j é definido de acordo com a equação (6.36), isto é, depende de $Q_{t-1}, Q_{t-2}, \dots, Q_{t-T}$.

A calibração do modelo de Yakowitz é feita definindo-se valores para τ (memória do processo), para m (número de vetores representativos) e para os vetores $Q(1), Q(2), \dots, Q(m)$. Barbosa (1981) assumiu $\tau=1$ e fez extensas investigações quanto ao número de intervalos em que deveria ser dividido o domínio de Q , neste caso Q_{t-1} , levando em consideração principalmente o número de elementos resultantes em cada conjunto C_j .

Método indireto

No método indireto, modelam-se as precipitações diárias, que são utilizadas em modelos de transformação de chuva em vazão, ou em modelos de simulação cujo interesse primordial seja chuva e não vazão (produção agrícola, por exemplo). O método indireto pode ser vantajoso nas regiões em que se disponha de um registro pluviométrico de melhor qualidade do que o fluviométrico, o que é freqüente acontecer.

Precipitação pontual instantânea

No início deste item explicou como o processo de Poisson pode ser utilizado para representar a ocorrência de pulsos de precipitação $x(t)$, medidos, por exemplo, em mm. Essa abordagem é mais indicada para a modelagem da precipitação discretizada em intervalos bem curtos, 15 min ou uma hora, por exemplo. Entretanto, é possível utilizá-la também para o intervalo diário. O modelo tem dois parâmetros:

$\nu = E^{-1}(W)$, medido por exemplo em h^{-1} , onde W é o tempo de espera entre eventos. W é exponencialmente distribuído

$\theta = E^{-1}(X(t))$, medido por exemplo em mm^{-1} , onde $X(t)$ é a altura instantânea de precipitação, dado que um pulso ocorre em t . $X(t)$ é exponencialmente distribuído.

Para um t qualquer em que não ocorra um pulso, $X(t)$ é identicamente nulo. Seja X_i a precipitação ao longo do dia i , em mm:

$$X_i = \sum_{j=1}^N X(t_j) \quad \text{se } N > 0 \tag{6.39}$$

$$X_i = 0 \quad \text{se } N = 0 \tag{6.40}$$

onde t_1, t_2, \dots, t_N são os instantes de ocorrência dos pulsos dentro do dia i e N o número de pulsos dentro do dia i (N é variável aleatória inteira, com distribuição Poisson).

Pode-se mostrar que

$$E(X_i) = \theta^{-1} \nu \delta \tag{6.41}$$

$$\text{var}(X_i) = 2\theta^{-2} \nu \delta \tag{6.42}$$

$$\text{corr}(X_i, X_j) = 0, \forall i \neq j \tag{6.43}$$

onde $\delta = 24$ horas

A partir da média e da variância amostral da precipitação diária, podem-se estimar os parâmetros (ν, θ) do modelo, através da resolução das equações (6.41) e (6.42). A geração do processo Poisson, que possibilita a obtenção de uma série sintética de precipitação diária pelo uso das equações (6.39) e (6.40), já foi descrito em tópico específico.

Duas generalizações do citado processo para representação de pulsos de precipitação têm sido sugeridos na literatura, ambas capazes de introduzir persistência temporal ($\text{corr}(X_i, X_j) > \theta$) no processo. A primeira modela a precipitação através de uma série de pulsos retangulares de duração finita (não instantâneos). O intervalo entre os pulsos, a duração e a intensidade de cada pulso são variáveis aleatórias, usualmente supostas independentes e exponencialmente distribuídas. A segunda generalização assume que a precipitação é governada por dois processos estocásticos hierarquicamente superpostos. O processo macro é Poisson e serve para delimitar o início da ocorrência de um grupo de pulsos, relativamente próximos no tempo. O processo micro serve para definir o número de pulsos em cada grupo e o instante da ocorrência de cada pulso. Essa abordagem, conhecida como modelo Newman-Scott, procura reproduzir a maneira como os pulsos ocorrem dentro de tempestades. Valdes *et al.*, (1985) descrevem-na com mais detalhes, e comparam os modelos de precipitação pontual instantânea.

Precipitação pontual diária

Registros diários de precipitação têm um grande número de zeros. Eles podem ser vistos como realizações de um processo estocástico intermitente. O processo só poderia ser modelado ajustando-se uma distribuição de probabilidades mista: uma massa de probabilidade p ($0 \leq p \leq 1$) correspondente à precipitação nula ($x=0$), e uma distribuição contínua $(1-p) f_X(x)$ para $x > 0$. Ocorre, no entanto, que esta abordagem desconsidera uma característica relevante da série de precipitação diária: dias chuvosos tendem a ocorrer aos grupos, principalmente em regiões onde a maior parte da precipitação tem origem na chegada de frentes, que por vezes estacionam durante dias seguidos. Formalizando, seja a variável aleatória S_i definida como

$$S_t = 1, \text{ dia chuvoso}$$

$$S_t = 0, \text{ dia não chuvoso}$$

Pode-se definir a Cadeia de Markov de dois estados pela matriz de transição

$$\begin{bmatrix} 1 - P_{01} & P_{01} \\ 1 - P_{11} & P_{11} \end{bmatrix} \tag{6.45}$$

onde

$$P_{01} = p[S_{t+1} = 1 | S_t = 0] \tag{6.46}$$

$$P_{11} = p[S_{t+1} = 1 | S_t = 1] \tag{6.47}$$

Espera-se, devido à persistência de dias chuvosos, que $P_{11} > P_{01}$

Para os dias t em que $S_t = 1$, sabe-se que existe alguma precipitação. Sua magnitude deverá ser sorteada da distribuição de probabilidades da variável aleatória - altura de precipitação diária. Essa distribuição de probabilidades poderá ser a empírica ou então alguma distribuição paramétrica ajustada aos dados amostrais. A distribuição empírica só deverá ser utilizada quando o "polígono de influência" do posto pluviométrico corresponder a um área de drenagem com tempo de concentração de pelo menos alguns dias. Nesse caso, cheias mais severas do que as registradas no histórico poderão ser produzidas pela simples persistência da precipitação, sem que haja dia algum no período chuvoso em que a precipitação sintética ultrapasse a máxima do histórico. Em bacias menores, no entanto, cheias sintéticas mais extremas do que as observadas no passado estarão associadas necessariamente a extrapolações de precipitação diária, o que impõe o ajuste de alguma função analítica à correspondente distribuição de probabilidades empírica.

Na fase de estimativa, deve-se selecionar um período suficientemente curto para que se possa aceitar a hipótese de estacionariedade do processo estocástico "precipitação diária", por exemplo, o mês mais molhado do ano. As probabilidades p , p_{01} e p_{11} são estimadas por:

$$\hat{p} = \frac{\text{número de dias não chuvosos}}{\text{número de dias chuvosos} + \text{números de dias não chuvosos}} \quad (6.48)$$

$$\hat{p}_{01} = \frac{\text{número de dias não chuvosos seguidos por dias chuvosos}}{\text{número de dias não chuvosos}} \quad (6.49)$$

$$\hat{p}_{11} = \frac{\text{número de dias chuvosos seguidos por dias chuvosos}}{\text{número de dias chuvosos}} \quad (6.50)$$

O algoritmo de geração pode ser descrito como:

a) Sorteie um número uniforme U

$$b) U \begin{cases} \leq \hat{p} \rightarrow S_0 = 0 \\ > \hat{p} \rightarrow S_0 = 1 \end{cases}$$

c) $t = 0$

$$d) \text{ Sorteie um número uniforme } U \begin{cases} \leq \hat{p}_{01} \rightarrow S_{t+1} = 1 \\ > \hat{p}_{01} \rightarrow S_{t+1} = 0 \end{cases} \\ \text{ -- Se } S_t = 0 \text{ então } \\ \\ \text{ -- Se } S_t = 1 \text{ então } U \begin{cases} \leq \hat{p}_{11} \rightarrow S_{t+1} = 1 \\ > \hat{p}_{11} \rightarrow S_{t+1} = 0 \end{cases}$$

e) Se $S_{t+1} = 1 \rightarrow$ sorteie X

f) $t = t+1$

g) vá para d.

Caso haja interesse em utilizar o modelo por períodos mais extensos do que um mês, não é mais possível assumir estacionariedade. Nesse caso, como em modelos anteriores, utilizam-se as equações (6.2) a (6.5) para ajustar funções periódicas às estimativas $\hat{p}(j)$, $\hat{p}_{01}(j)$, $\hat{p}_{11}(j)$, $j = 1, 2, \dots, m$, sendo m o número de frações em que se divide o ano.

Woolhiser e Roldan (1986) estudaram a variabilidade espacial dos parâmetros de funções periódicas análogas à equação (6.2), ajustadas a estimativas $(\hat{p}(j), \hat{p}_{01}(j), \hat{p}_{11}(j))$ de cada posto pluviométrico de uma região. Esses autores conseguiram dessa maneira interpolar as características da precipitação pontual para um local desprovido de posto pluviométrico, a partir das informações de postos vizinhos.

Precipitação multivariada diária

A modelagem pontual de precipitação é útil quando se pretende estudar uma bacia hidrográfica pequena, em que o registro do que ocorre em apenas um posto pluviométrico já é suficiente para caracterizar, pelo menos de forma aproximada, a precipitação sobre a bacia. Entretanto, para bacias maiores é preciso levar em consideração a variação especial da chuva.

Idealmente seria interessante dispor de um modelo matemático que caracterizasse a precipitação como um processo contínuo no tempo e no espaço, como o proposto por Bras e Rodríguez-Iturbe (1976). Esse modelo, e alguns de seus derivados, possuem equações que representam a gênese, o deslocamento e o fim dos eventos chuvosos, com base em conceitos meteorológicos. Infelizmente não existem ainda técnicas confiáveis para a estimativa dos parâmetros (Valdes *et al*, 1985).

Uma alternativa simples é modelar de uma forma empírica, a precipitação medida em vários postos pluviométricos de uma região, procurando preservar a persistência temporal de precipitação em cada local, bem como a persistência espacial entre os vários postos.

Poder-se-ia pensar, inicialmente, em estender a abordagem utilizada para precipitação pontual diária. Por exemplo, se houvesse dois postos na bacia, seria necessário definir quatro estados para S_t , correspondentes aos arranjos: (chuva posto 1, chuva posto 2); (chuva posto 1, não chuva posto 2). No caso de r postos, o número de estados para S_t seria 2^r , o que implica na estimativa de uma matriz de transição com dimensão 2^r para cada fração em que se divide o ano. Isto é $2^{2^r} - 2^r$, probabilidade de transição. Suponha que a fração seja de 30 dias e o número de anos do registro histórico seja de 50 anos. Nesse caso, o número total de observações diárias (dependentes entre si)

é 1.500 r. Para $r \geq 7$, o número total de observações já seria menor do que o número de probabilidades de transição a serem estimadas, o que ilustra a impraticabilidade da abordagem.

Como os processos estocásticos multivariados gaussianos são mais fáceis de serem manipulados do que os não gaussianos, Kelman *et al*, (1985) propuseram a transformação preliminar de precipitação X numa variável normal padrão Y . Essa transformação pode ser facilmente efetivada assumindo-se que todos os valores de Y menores do que uma determinada constante α foram censurados, resultando em valores nulos para X :

$$y = \Phi^{-1}(F(x)), \text{ se } x > 0 \quad (6.51)$$

$$y \leq \alpha = \Phi^{-1}(j/k), \text{ se } x = 0 \quad (6.52)$$

onde

$\Phi(\cdot)$ – distribuição acumulada normal padrão j – número de dias com precipitação nula na série histórica, na fração do ano sob consideração

K – número total de dias, idem

$F(\cdot)$ – distribuição acumulada de precipitação.

As equações (6.51) e (6.52) são empregadas para mapear todas as precipitações numa série temporal multivariada e censurada $\{Y_{i,t}\}$, sendo $i = 1, 2, \dots, p$ o índice do posto e t o índice do dia. A modelagem do processo $\{Y_{i,t}\}$ pode ser feita de inúmeras maneiras. Kelman *et al*, (1985) sugeriram adotar a equação (4.68), abaixo reproduzida

$$Y_{t+1} = AY_t + Bw_{t+1} \quad (6.53)$$

As equações que definem as matrizes A e B foram descritas no item 4.4. A única peculiaridade neste caso é que as matrizes $E(Y_{t+1} Y_t^T)$ e $E(Y_t Y_t^T)$ precisam ser estimadas a partir de amostras censuradas do processo Y_t . Como os elementos dessas matrizes são coeficientes de correlação, essa estimativa pode ser feita com o auxílio de uma relação entre o coeficiente de correlação de uma amostra completa e o coeficiente de correlação de uma amostra truncada. Nesta última consideram-se apenas os elementos não censurados. Kelman *et al*, (1985) apresentam detalhes do método.

Uma vez estimadas as matrizes A e B , séries sintéticas de precipitação podem ser facilmente produzidas a partir do sorteio de valores V_{t+1} que, com o auxílio da equação (6.52), resultam em vetores Y_{t+1} . Através da transformação inversa definida por (6.51) e (6.52), que foi derivada para cada posto pluviométrico, obtém-se a precipitação pontual.