

GERAÇÃO DE VAZÕES DIÁRIAS I: COMPARAÇÃO ENTRE DOIS MODELOS NÃO-PARAMÉTRICOS

Fabricio Muller¹ & Eloy Kaviski²

RESUMO --- Yakowitz (1979) defendeu o ponto de vista de que modelos não-paramétricos tendem a gerar séries de vazões diárias que preservam melhor a forma dos hidrogramas do que modelos paramétricos, resultado corroborado por Müller (1997, 1998). Tendo isto em mente, este artigo compara dois modelos não-paramétricos de geração de vazões diárias, o modelo de Yakowitz (1979) e o Diana (Kelman *et al.*, 1983). A estação escolhida para a comparação foi a de Rio dos Patos, no Estado do Paraná. Os dois modelos apresentaram bons resultados em termos de estatísticas de longo período, e o modelo de Yakowitz apresentou resultados mais consistentes para estatísticas de vazões máximas, possivelmente por causa da equação introduzida por Barbosa (1981). As autocorrelações diárias de lags 1 a 4 mostraram um resultado melhor para o modelo Diana do que para o Yakowitz, enquanto que esta situação se inverte nos lags maiores. Em termos puramente visuais, os hidrogramas do modelo Diana parecem um pouco mais consistentes com os do histórico, resultado causado provavelmente pelo melhor ajuste das autocorrelações mais baixas.

ABSTRACT --- Yakowitz (1979) defended that not-parametric models tend to generate series of daily outflows that preserve better the form of the hydrographs than parametric models, resulted corroborated by Müller (1997, 1998). Having this in mind, this article compares two non-parametric models of generation of daily outflows, the model of Yakowitz (1979) and the Diana (Kelman *et al.*, 1983). The gauging site chosen for the comparison was of Rio dos Patos, in the State of the Paraná, Brazil. Both models had presented good results in terms of long period statistics, and the Yakowitz model presented results more consistent for maximum outflows statistics, possibly because of the equation introduced by Barbosa (1981). The daily autocorrelations of lags up to 4 had shown better results for the Diana model then for the Yakowitz model, whereas this situation inverts in higher lags. In purely visual terms, the hydrographs of the Diana model seem a little more consistent with the ones of the historical series, result caused probably by the better adjustment of lowest autocorrelations.

Palavras-chave: séries temporais, geração de vazões diárias, modelos não-paramétricos.

1) Engenheiro civil e diretor da Tramo Sociedade Civil-Estruturas. Curitiba/PR. e-mail: fabricio_muller@yahoo.com

2) Professor do Departamento de Hidráulica e Saneamento da UFPR. e-mail: eloy.dhs@ufpr.br

1 - INTRODUÇÃO

Séries sintéticas de vazões diárias podem ser utilizadas em hidrologia em, por exemplo, simulação de vazões máximas ou refinamento de decisões obtidas com simulação mensal.

Yakowitz (1979) defendeu o ponto de vista de que modelos não-paramétricos tendem a gerar séries de vazões diárias que preservam melhor a forma dos hidrogramas do que modelos paramétricos, resultado corroborado por Müller (1997,1998).

Este trabalho faz uma comparação entre dois modelos não-paramétricos de geração de vazões diárias, o modelo de Yakowitz (1979) e o Diana (Kelman et al, 1983) na estação fluviométrica de Rio dos Patos.

2 – DESCRIÇÃO DOS MODELOS

2.1 – Modelo de Yakowitz (1979)

O modelo de Yakowitz (1979) é baseado na distribuição de probabilidades empírica de \mathbf{q}_t , condicionada nos valores imediatamente anteriores, $\mathbf{q}_t = [q_{t-1}, q_{t-2}, \dots, q_{t-N}]$, onde N é a memória arbitrada para o processo. Para isto, o autor definiu um conjunto finito de vetores representativos, $\mathbf{q}_{N+1}, \mathbf{q}_{N+2}, \dots, \mathbf{q}_m$, de tal forma que qualquer vetor composto por N vazões geradas $\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_m$, pudesse ser classificado numa das m categorias de acordo com a distância entre \mathbf{Q} a cada um dos \mathbf{q}_i , $i=N+1, \dots, m$. (A notação utilizada para este item é semelhante igual àquela utilizada por Yakowitz (1979), em que símbolos em negrito significam vetores.)

Em outras palavras, neste modelo a vazão num dia $j+1$ é obtida das N vazões nos dias anteriores, onde N é a ordem do modelo. Por exemplo, se a ordem do modelo for igual a dois, adota-se para início da simulação duas vazões, Q_{j-1} e Q_j aleatoriamente. A partir destas duas vazões o modelo não-paramétrico, inicialmente procura em qual "estado representativo" previamente obtido melhor se encaixa o vetor contendo as vazões Q_{j-1} e Q_j . A cada "estado representativo" corresponde uma função de distribuição empírica e , a partir desta, o modelo de Yakowitz gera a vazão no próximo dia, ou seja, Q_{j+1} . Obtidas as vazões nos dias j e $j+1$ gera-se a vazão no dia $j+2$, e assim por diante. Maiores detalhes do modelo serão apresentados a seguir.

No modelo de Yakowitz a maior importância está na obtenção da função empírica acumulada $F[g(Q_{j+1})|Q_j]$ onde Q_j é um vetor contendo as N vazões anteriores a Q_{j+1} , e N é a ordem do modelo. Isto é feito seguindo-se os passos:

1. Escolhe-se a ordem do modelo;

2. Dividem-se as vazões do histórico em $h-N+1$ grupos de N valores consecutivos, os vetores \mathbf{q}_j , $j=1..h-N+1$, onde h é o número de dias do histórico. Por exemplo, se a ordem do modelo for 3, o vetor \mathbf{q}_j começando no dia 1 de janeiro conterà as vazões dos dias 1, 2 e 3 de janeiro;

3. Determinam-se os "estados representativos" \mathbf{c}_v , $v=1, \dots, m$ que são m vetores de dimensão N (onde m é estimado por tentativas) que representam o conjunto de vetores \mathbf{q}_j . Logo, \mathbf{c}_v são N -uplas tais que:

$$J(\mathbf{c}_1, \dots, \mathbf{c}_m) = \sum_{j=N}^h \left[\min_{1 \leq v \leq m} \left\| \mathbf{q}_j - \mathbf{c}_v \right\|^2 \right] \quad (1)$$

seja mínimo, isto é, cada vetor \mathbf{c}_v , $v=1, \dots, m$ é aquele que melhor representa um grupo de vetores \mathbf{q}_j mais próximos dele (o sinal $\| \|$ representa distância euclidiana). Para resolver este problema de otimização, Yakowitz (1979) sugere o algoritmo para análise de agrupamentos *K-Means*, proposto por MacQueen (Kaviski, 1992). O conjunto de vetores \mathbf{q}_j representados por \mathbf{c}_v é chamado S_v . Deste modo, ter-se-á m conjuntos S_v , cada um composto por $B(v)$ vetores \mathbf{q}_j . O número $B(v)$, para cada conjunto S_v , é calculado pelo processo de otimização utilizado. Daqui por diante, os vetores \mathbf{q}_j representados por \mathbf{c}_v serão chamados $\mathbf{Q}_{k,v}$, para $k=1, \dots, B(v)$ e $v=1, \dots, m$.

4. O passo seguinte é definir os conjuntos Y_v ($v=1, \dots, m$) tais que

$$Y_v = \left\{ \mathbf{q}_v, \quad j+1: \left\| \mathbf{q}_j - \mathbf{c}_v \right\| \leq \left\| \mathbf{q}_j - \mathbf{c}_i \right\|, \quad \text{para todo } i \leq m \right\} \quad (2)$$

onde Y_v é a vazão do dia subsequente aos vetores \mathbf{q}_j mais próximos de \mathbf{c}_v . A cada vetor $\mathbf{Q}_{k,v}$, $k=1, \dots, B(v)$ que compõe um conjunto S_v corresponderá uma vazão $q_{k,j+1}$ no conjunto Y_v . Foi visto acima que os vetores $\mathbf{Q}_{k,v}$, $k=1, \dots, B(v)$, que compõem um conjunto S_v são originários da série de vetores \mathbf{q}_j , $j=1, \dots, h$; para a obtenção de um valor de $q_{k,j+1}$ pertencente a um conjunto Y_v é necessário que o seu correspondente valor de $\mathbf{Q}_{k,v}$ seja "igualado" ao vetor \mathbf{q}_j original - como o vetor \mathbf{q}_j tem N componentes, o valor de $q_{k,j+1}$ será tomado como igual ao próximo valor $(N+1)$ encontrado na série \mathbf{q}_j histórica. Por exemplo, seja $N = 3$ e um vetor $\mathbf{Q}_{k,v}$ dentro de um conjunto S_v seja composto pelas vazões originárias \mathbf{q}_j , vetor este composto pelas vazões dos dias 1, 2 e 3 de janeiro de 1975. O correspondente valor da vazão $q_{k,j+1}$ que representará o vetor $\mathbf{Q}_{k,v}$ no conjunto Y_v é a próxima vazão do calendário, ou seja, a vazão do dia 4 de janeiro de 1975.

5. Obtidos os m conjuntos Y_v , as m funções empíricas $G_v [q]$ dos conjunto Y_v são calculadas através da técnica da posição de plotagem. As $B(v)$ vazões $q_{k,j+1}$, que compõem um conjunto Y_v são ordenadas decendentemente e a cada vazão corresponderá uma probabilidade, dada por:

$$G_{\nu}(q) = \begin{cases} \frac{2i-1}{2B(\nu)}, & \text{se } q \leq q_{\max} \\ 1 - e^{-\lambda}, & \text{se } q > q_{\max} \end{cases} \quad (3)$$

onde

$$q = q_{k,j+1}$$

q_{\max} = maior valor encontrado em cada conjunto Y_{ν} , $\nu=1, \dots, m$;

i = número de ordem descendente da série $q_{k,j+1}$: para $i=1$, corresponderá $q = q_{\max}$; e

$$\lambda = -\frac{1}{q_{\max}} \ln \left[\frac{1}{2B(\nu)} \right]$$

A segunda equação em (3) foi introduzida por Barbosa (1981) para que o modelo possibilitasse a obtenção de vazões maiores do que o histórico. A função discreta $G_{\nu}(q)$ deve ser transformada numa função $F_{\nu}(q)$, contínua através de *splines*, ou através de interpolação linear. Como para cada função $F_{\nu}(q)$, $\nu=1, \dots, m$ corresponde um conjunto Y_{ν} e um vetor representativo \mathbf{c}_{ν} a função $F_{\nu}(q)$ pode simplesmente ser denominada $F_{\nu}(q|\mathbf{c}_{\nu})$.

6. O próximo passo é a geração propriamente dita. As etapas para geração são as seguintes:

a) Gerar por algum método as N primeiras vazões da série simulada;

b) Através da função $F_{\nu}(q|\mathbf{c}_{\nu})$, gerar a próxima vazão. Isto é feito da seguinte maneira. Seja o vetor \mathbf{v} o vetor composto pelas N vazões previamente geradas. Deve-se procurar nos m vetores representativos \mathbf{c}_{ν} qual destes vetores é o mais aproximado do vetor \mathbf{v} , calculando-se para $\nu=1, \dots, m$ qual o vetor \mathbf{c}_{ν} tal que $\|\mathbf{v} - \mathbf{c}_{\nu}\|^2$ resulte mínimo. A próxima vazão q simulada é obtida fazendo-se $F_{\nu}(q|\mathbf{c}_{\nu}) = F_{\nu}(q|\mathbf{v})$: um valor aleatório uniforme U entre 0 e 1 é gerado e a próxima vazão é calculada através de $q = F_{\nu}^{-1}(U)$. Se o valor de U gerado for menor que $[1-1/B(\nu)]$, deve-se utilizar a primeira equação em (3) na simulação; caso contrário, a segunda. É bom lembrar que para cada valor U gerado corresponderá um correspondente valor de q simulado, tendo em vista que a função discreta $G_{\nu}(q)$ em (3) foi transformada numa função contínua $F_{\nu}(q)$ através de *splines* ou através de interpolação linear.

A geração, em outras palavras, se processa da seguinte forma: seja $\mathbf{v}_1=[q_3, q_2, q_1]$, o vetor contendo as três primeiras vazões simuladas q_3, q_2, q_1 . A vazão q_4 é simulada através de $F_{\nu}(q|\mathbf{v}_1)$. O valor de \mathbf{v}_2 será então dado por $[q_4, q_3, q_2]$. A vazão q_5 é obtida através de $F_{\nu}(q|\mathbf{v}_2)$, e assim por diante.

Yakowitz (1979) apresenta um exemplo simplificado que ajuda na compreensão do processo utilizado pelo modelo não-paramétrico na simulação de vazões diárias, transcrito no trabalho de Müller (1997).

2.1 – Modelo Diana

No modelo Diana de geração de séries sintéticas diárias (Kelman et al., 1983) a vazão é considerada como a soma de dois componentes:

$$Q(t) = U(t) + O(t) \quad , \quad t = 1, 2, \dots \quad (5)$$

onde $U(t)$ depende de fatores externos, tais como precipitações (podendo ser considerado um "pulso" do modelo Diana), e $O(t)$ representa o contínuo esvaziamento da água armazenada na bacia hidrográfica. (A notação apresentada aqui para descrever o modelo Diana é baseada naquela utilizada por Kelman et al., 1983.)

Na modelagem do processo $U(t)$, dois aspectos devem ser considerados: definir sua distribuição de probabilidade marginal e preservar a dependência entre valores sucessivos de $U(t)$. Para resolver estes aspectos, o modelo Diana mapeia o processo $U(t)$ através de uma transformação do tipo $U(t) = [W(t)]$. O processo $W(t)$ resulta de uma censura ao processo $Z(t)$, auto-regressivo de ordem 1 de distribuição normal. Já o processo $O(t)$, o contínuo esvaziamento da bacia hidrográfica, é igual a $Q(t-1)$ multiplicado por uma taxa de recessão.

Os passos para a geração de vazões diárias pelo modelo Diana são os seguintes:

1. Seja a série histórica de vazões diárias $q(\tau)$, para $\tau = 1, \dots, h$, onde h é o número de dias do histórico; necessita-se inicialmente determinar a função empírica de distribuição acumulada $F_{\Delta}[(\tau)]$, onde

$$\delta(\tau') = q(\tau) / q(\tau-1), \quad \text{se } \delta(\tau') < 1, \quad \tau' = 1, \dots, \Lambda \quad (6)$$

onde Λ = número de vezes no histórico em que $q(\tau) / q(\tau-1) < 1$.

Este passo serve para determinar uma estimativa inicial de λ , a taxa de recessão característica.

2. Cálculo da distribuição empírica acumulada $F_K[k(\tau'')]$, onde

$$k(\tau'') = q(\tau) / q(\tau-1) \quad \text{se } k(\tau'') < \lambda, \quad \tau'' = 1, \dots, m' \quad (7)$$

onde m' é o número de vezes no histórico em que $q(t) / q(t-1) < \lambda$.

Para calcular funções empíricas acumuladas nos passos 1 e 2, deve-se ordenar a amostra (de δ e de k) e então, para cada valor desta nova amostra ordenada, corresponderá um valor igual a i / N , onde i é a posição do valor na amostra ordenada, e N o tamanho da amostra.

3. Calcula-se então a distribuição empírica acumulada $F_U[u(\tau)]$, onde

$$\begin{aligned} u(\tau) &= q(\tau) - q(\tau-1), & \text{se } q(\tau) / q(\tau-1) > \lambda \\ u(\tau) &= 0, & \text{se } q(\tau) / q(\tau-1) < \lambda \end{aligned} \quad (8)$$

A distribuição empírica acumulada é calculada através dos seguintes passos:

a. ordena-se a amostra $u(\tau)$, de modo que $u(1) \leq u(2) \leq \dots \leq u(h)$, onde h é o tamanho da amostra de vazões diárias do histórico, transformando-se a amostra original $u(\tau)$ na amostra $u(j)$, ordenada;

b. seja m' o número de zeros na amostra $u(j)$; logo, teremos que a probabilidade de que $u(j)$ seja igual a zero, ou seja, $P[u(j)=0]$, é igual a m'/h ;

c. para cada valor $u(j) > 0$ da amostra ordenada corresponderá um par $[D(j), E(j)]$, onde

$$\begin{aligned} D(j) &= (j-1) / h, & j &= m'+1, m'+2, \dots, h \\ E(j) &= j / h, & j &= m'+1, m'+2, \dots, h \end{aligned} \quad (9)$$

E para cada valor $u(j)$ [$u(j) > 0$] corresponderá então o valor médio

$$Y(j) = [D(j) + E(j)] / 2$$

A distribuição empírica acumulada de $u(t)$ é dada finalmente por

$$F[u(j)] = Y(j) \quad (10)$$

4. Seja a série $z(\tau)$, onde

$$z(t) = \rho z(t-1) + \sqrt{1-\rho^2} \varepsilon(\tau) \quad (11)$$

onde $\varepsilon(\tau)$ é uma variável aleatória normal com média nula e variância unitária. O processo $w(\tau)$ resulta em uma censura imposta ao processo $z(t)$, através de

$$\begin{aligned} w(\tau) &= z(\tau) & \text{se } z(\tau) > \beta \\ w(\tau) &= \beta & \text{se } z(\tau) \leq \beta \\ \beta &= \Phi^{-1}(m/h) \end{aligned} \quad (12)$$

onde $\Phi(\cdot)$ é a função de distribuição normal padrão acumulada. O valor de ρ , o coeficiente de correlação da população $z(\tau)$, em (11), é calculado através da distribuição histórica de $w(\tau)$, da seguinte maneira:

Seja o processo $u(\tau)$, histórico não ordenado. A série $\Psi(\tau)$, histórica não ordenada, é obtida através de $\Psi(\tau) = F_U[u(\tau)]$, isto é, relacionando-se um valor $\Psi(\tau)$ a cada valor $u(\tau)$ do histórico. Este valor de $\Psi(\tau) = F_U[u(\tau)]$ é obtido através da relação mostrada no passo 3. Para cada valor de $u(\tau)$

igual a zero o valor correspondente de $w(t)$ será igual a m' / h . A série $w(\tau)$ é obtida pela transformação

$$w(t) = \phi[\Psi(\tau)] \quad (13)$$

Para $\Psi(\tau) = m' / h$ (correspondente a $u(\tau) = 0$), $w(\tau) = \beta$. A estimativa inicial de ρ (o coeficiente de correlação na distribuição truncada) é igual a $\hat{\rho}'$, obtido utilizando-se todos os pares contíguos ($w(\tau) > \beta$, $w(\tau+1) > \beta$) em:

$$\hat{\rho}' = \frac{\sum [w(\tau) - \bar{w}_\tau] [w(\tau+1) - \bar{w}_{\tau+1}]}{\{\sum [w(\tau) - \bar{w}_\tau]^2 \sum [w(\tau+1) - \bar{w}_{\tau+1}]^2\}^{1/2}} \quad (14)$$

$$\bar{w}_\tau = \frac{\sum w(\tau)}{n} \quad \bar{w}_{\tau+1} = \frac{\sum w(\tau+1)}{n} \quad (15)$$

onde n é o número de pares contíguos obtidos. O valor do coeficiente de correlação ρ é obtido de Reigier e Hamdam (1971), com $\hat{\rho}'$ e β .

5. A geração pode ser iniciada. A série gerada será aqui representada por maiúsculas $Q(t)$, com $t = 1, \dots, \zeta$ onde ζ é o número de vazões diárias geradas. O primeiro valor gerado, $Q(0)$, é sorteado da distribuição empírica $q(\tau)$, $\tau = 1, \dots, h$. Um primeiro valor inicial $z(0)$ é sorteado da distribuição normal padrão $\phi(\cdot)$.

6. Faça $t = 1$.

7. Sorteie $e(t)$ da distribuição normal padrão $\phi(\cdot)$.

8. Calcule onde $z(t) = \rho z(t-1) + \sqrt{1-\rho^2} \varepsilon(t)$, onde $\varepsilon(t)$ é um ruído aleatório com média nula e variância unitária. Aqui o processo $z(t)$ sofre a censura da fórmula (12): faça $w(t) = \max(\beta, z(t))$.

9. Obtenha $u(t)$: faça $\Psi(t) = \phi^{-1}[w(t)]$. O valor de $u(t)$ é obtido através de $u(t) = F_U^{-1}[\Psi(t)]$, onde F_U é a distribuição empírica acumulada da variável $u(t)$ (ver passo 3).

10. Se $u(t) > 0$, faça $Q(t) = u(t) + \lambda Q(t-1)$. Se $u(t) = 0$, obtenha $k(t)$ da função empírica de distribuição acumulada de $k(\tau)$, ou seja, $F_K[k(\tau)]$, através da função $k(\tau) = F_K^{-1}[\Xi]$, onde Ξ é um número aleatório uniforme entre 0 e 1 (ver passo 2). Então faça $Q(t) = k(t) Q(t-1)$.

11. Faça $t = t + 1$ e retorne ao passo 7.

3 - HIDROLOGIA DA ESTAÇÃO ESCOLHIDA

A estação fluviométrica de Rio dos Patos (código: 64620000), na Bacia do Rio Ivaí, Estado do Paraná, está situado a 25°12' latitude sul, 50°56' longitude leste, e tem uma área de drenagem correspondente de 1090 km². O período estudado vai de 1o de janeiro de 1931 a 31 de dezembro de 2003. A vazão média deste período é de 22,2 m³/s, e a vazão máxima observada foi de 573,0 m³/s, em 31/5/1992. A série de vazões no local foi obtida no site *Hidroweb* (<http://hidroweb.ana.gov.br/>), da ANA (Agência Nacional de Águas). Por simplicidade computacional, tanto na análise histórica quanto na geração, não foram considerados dias 29 de fevereiro: os anos serão considerados como sempre tendo 365 dias, sejam eles bissextos ou não.

A figura 1 apresenta um fluviograma típico, de 6 meses de vazões diárias, em Rio dos Patos.

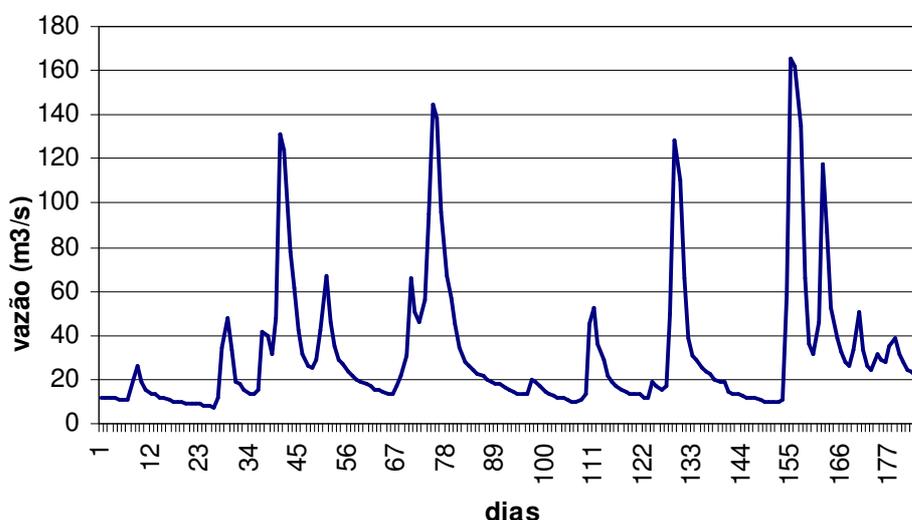


Figura 1 - Vazões diárias em Rio dos Patos, de 1/7/1976 a 31/12/1976

4 - TESTES COMPARATIVOS

Conforme Müller (1997), os dois modelos não paramétricos analisados não preservam estatísticas sazonais. Por isto, os testes farão apenas comparações de estatísticas de máximas (aplicação freqüente para geração de séries de vazões diárias) e de vazões diárias de longo período.

Uma qualidade dos modelos não-paramétricos, citada anteriormente, é a preservação da forma dos hidrogramas e, por isto, também serão testadas grandezas que dão uma idéia destas formas. Também neste sentido é feito o teste visual: escolhe-se ao acaso seis meses de vazões diárias

geradas pelo modelo e analisa-se de maneira expedita se ele "se parece" com a série histórica apresentada na figura 1.

Dada a alta correlação serial normalmente encontrada em séries de vazões diárias, o autocorrelograma diário também é analisado.

3.1 – Índice de avaliação global

O índice de avaliação global de O'Connell (apud Barbosa, 1980 e Kelman et al, 1983) consiste em calcular para a série histórica e para diversas séries sintéticas com o mesmo comprimento do histórico, estatísticas específicas ou índices. Disto resulta uma distribuição de probabilidades para os índices e a coerência entre as séries sintéticas pelo modelo e a série histórica pode ser medida pela posição do índice histórico nesta distribuição. Um índice I qualquer é definido como uma função de uma série de vazões, $I=f[Q(1),Q(2),\dots,Q(h)]$. Por exemplo, $I=\max[Q(1),\dots,Q(h)]$. Aplicando-se a função à série de vazões históricas obtém-se o índice histórico $i(\text{hist})$. Com n séries geradas com o mesmo comprimento do histórico obtém-se n índices gerados $i(j)$, $j=1,\dots,n$. O desempenho do modelo é então medido pela proporção de índices gerados maiores do que o histórico, $P[i(j)>i(\text{hist})]$. Se esta proporção for muito pequena ou muito grande, isto é uma indicação de que a observação histórica é atípica para o modelo. Em geral, deseja-se testar m índices, $I(i)$, $i=1,\dots,m$, arranjados no vetor

$$\mathbf{r} = \begin{bmatrix} I(1) \\ I(2) \\ \dots \\ I(m) \end{bmatrix} \quad (16)$$

$$I(i)=f_i [Q(1),\dots,Q(h)] \quad (17)$$

O vetor obtido com o histórico é denotado \mathbf{r}_{hist} e existem n vetores \mathbf{r}_j , $j=1,\dots,n$, cada um obtido com uma série gerada. Define-se portanto uma distribuição de probabilidades multivariada. Para cada vetor \mathbf{r}_{hist} e para \mathbf{r}_j , define-se a distância ao centro da distribuição pelo escalar positivo:

$$d_j = (\mathbf{r}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r}_j - \boldsymbol{\mu}) \quad (18)$$

$$d_{\text{hist}} = (\mathbf{r}_{\text{hist}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{r}_{\text{hist}} - \boldsymbol{\mu}) \quad (19)$$

onde

$$\boldsymbol{\mu} = \sum_{j=1}^n \mathbf{r}_j \quad (20)$$

$$\Sigma = \frac{1}{n} \sum_{j=1}^n (\mathbf{r}_j - \boldsymbol{\mu})(\mathbf{r}_j - \boldsymbol{\mu})^T \quad (21)$$

Nas fórmulas acima, os símbolos em negrito significam notação matricial. A variável d_{hist} é então comparada com as distâncias d_j através de duas medidas de desempenho: a proporção de distâncias geradas maiores que a distância histórica, $P[d_j > d_{\text{hist}}]$, e a distância padronizada (nota - Ψ) definida por:

$$\psi = \frac{d_{\text{hist}} - m}{s} \quad (22)$$

onde m é a média dos valores de d_j , $j = 1, \dots, n$ e s é definido como

$$s = \sqrt{\frac{1}{n} \sum_{j=1}^n (d_j - m)^2} \quad (23)$$

Quanto mais $P[d_j > d_{\text{hist}}]$ se aproximar de 0,5 e quanto menor o valor absoluto de Ψ , melhor o desempenho do modelo.

Os índices escolhidos para a aplicação neste trabalho, $m=10$ no total, à exceção do de número 10, volume de espera para enchentes, foram os mesmos escolhidos por Kelman et al. (1983) para a análise do modelo Diana. Os índices estão relacionados no quadro 1.

Quadro 1 - Índices testados na avaliação global desenvolvida por O'Connell

Número	Índice
1	Média da vazão diária
2	Desvio-padrão da vazão diária
3	Assimetria da vazão diária
4	Curtose da vazão diária
5	Média da vazão máxima anual
6	Desvio-padrão da vazão máxima anual
7	Assimetria da vazão máxima anual
8	Curtose da vazão máxima anual
9	Vazão máxima de todo o período
10	Correlação lag 1 da vazão média diária

3.2 – Porcentagens que se relacionam com as formas dos hidrogramas

Os índices que se relacionam com a forma das hidrógrafas são:

1. *porcentagem acima*: porcentagem de dias em que a vazão foi maior que a média mensal;
2. *ascensões*: porcentagem de dias em que a hidrógrafa subiu;
3. *saltos*: porcentagem de dias em que ocorreram saltos, isto é, onde $Q(i-1) < Q(i) > Q(i+1)$, onde $Q(i)$ é a vazão para um dia i .

4 – APLICAÇÃO DOS MODELOS EM RIO DOS PATOS

4.1 – Modelo de Yakowitz (1997)

Do mesmo modo que em Müller (1997), as médias geradas de longo período variaram bastante com a utilização de m (o número de vetores de estados representativos), $B(\min)$ (que é o menor valor de B nos m conjuntos de sucessores) e N (a ordem do modelo). Para obter o melhor modelo na geração de vazões diárias em Rio dos Patos, geraram-se séries para cada caso de combinação dos seguintes parâmetros: $N = 1, 2, 3$; $m = 28, 75, 100, 160$; $B(\min) = 70, 95, 120, 140$, resultando $3 \times 4 \times 4 = 48$ séries, cada uma com dez períodos iguais ao tamanho do histórico. Foi considerado que o melhor resultado seria o melhor colocado nos critérios média e índice de O'Connel.

O melhor resultado foi obtido com séries geradas utilizando os seguintes parâmetros: $N = 1$; $m = 160$; $B(\min) = 95$. Esta combinação ficou em terceiro lugar tanto em termos de média de longo período quanto no quesito índice de O'Connel.

Para as simulações realizadas em Rio dos Patos, a função discreta $G_v(q)$ em (4) foi transformada numa função contínua $F_v(q)$ através de interpolação linear.

4.2 – Modelo Diana

Para a aplicação do modelo no Rio dos Patos, Müller (1997) calculou uma nova geração de séries para cada um dos 20 quantis de $F_\Delta[(\tau)]$ (o 5^o, 10^o, ... até o 100^o), correspondendo, a cada uma, um novo valor para λ (a taxa de recessão característica). O escolhido foi aquele que gerou séries com média mais próxima à histórica.

Com 10 gerações de séries do tamanho da do histórico para cada λ , o valor definitivo foi o obtido no 70^o quantil de $F_\Delta[(\tau)]$.

Como neste artigo a estação utilizada foi a mesma de Müller (1997) - mas com a série de vazões históricas atualizadas - considerou-se que utilizar o mesmo quantil de $F_\Delta[(\tau)]$ aqui não iria causar discrepâncias nos resultados, e este foi o procedimento adotado.

4.3 – Geração de séries

Escolhidos os parâmetros de geração foram geradas 100 séries de vazões diárias do tamanho das do histórico (1/1/1931-31/12/2003) para comparação entre os dois modelos.

5 – COMPARAÇÃO DE RESULTADOS

5.1 – Índice de O'Connel

O quadro 2 apresenta os resultados dos parâmetros estatísticos relacionados anteriormente no quadro 1, tanto para a série histórica de vazões, quanto para os 100 períodos do tamanho do histórico gerados pelos modelos de Yakowitz e Diana.

Pelos resultados do quadro 2, vê-se que nas estatísticas de vazões diárias de longo período (itens 1 a 4 e 10) os dois modelos saíram-se bem, chegando a resultados muito próximos da série histórica - com exceção da curtose diária.

Quanto às estatísticas de máximas, o modelo de Yakowitz apresentou resultados sensivelmente mais próximos dos do histórico que o modelo Diana, o que sugere que o primeiro modelo é o melhor na simulação de vazões máximas - possivelmente por causa da introdução da fórmula de Barbosa (1981), a segunda fórmula em (4).

Quadro 2 - Resultados dos parâmetros estatísticos utilizados no teste de O'Connel

N.	Grandeza	Histórico	Yakowitz	Diana
1	Média da vazão diária	22,23	21,90	21,94
2	Desvio-padrão da vazão diária	31,97	30,12	31,42
3	Assimetria da vazão diária	4,51	4,08	4,26
4	Curtose da vazão diária	34,25	29,23	28,79
5	Média da vazão máxima anual	197,73	196,20	215,85
6	Desvio-padrão da vazão máxima anual	92,33	90,40	76,62
7	Assimetria da vazão máxima anual	1,59	1,30	0,58
8	Curtose da vazão máxima anual	6,50	5,87	2,89
9	Vazão máxima de todo o período	573,00	531,36	421,72
10	Correlação lag 1 da vazão média diária	0,901	0,875	0,909

O quadro 3, que apresenta os resultados do índice de O'Connel para a série histórica e para os dois modelos gerados, corrobora o resultado anteriormente. Por causa do seu melhor ajuste para vazões máximas, a nota do modelo de Yakowitz é mais próxima de zero (ou seja, resultou num melhor ajuste) do que a do modelo Diana, e a correspondente distância $d > d_{\text{hist}}$ (fórmula (18)) ficou mais próxima de 50%. A melhor geração de máximas em Yakowitz também fica evidente nos

resultados das probabilidades $\text{sint} > \text{hist}$ - que mostram a porcentagem de vezes em que o valor da estatística sintética resultou maior do que a sua correspondente histórica nas 100 gerações do tamanho do histórico - para os índices de número 6 a 10, que estão mais próximas de 50% do que as correspondentes para o modelo Diana.

Ainda sobre o quadro 3, é um resultado significativo o do índice 10, correlação lag 1: todas as 100 séries geradas por Yakowitz resultaram em um valor menor do que o do histórico, e apenas uma das 100 geradas pelo modelo Diana resultou numa autocorrelação lag 1 menor do que a do histórico. Este resultado será ainda citado neste trabalho em outro teste, o visual.

Quadro 3 - Resultados do índice de O'Connel

N.	Grandeza		Yakowitz	Diana
1	Média da vazão diária	$\text{prob}(\text{sint} > \text{hist}) =$	31,0%	33,0%
2	Desvio-padrão da vazão diária	$\text{prob}(\text{sint} > \text{hist}) =$	12,0%	32,0%
3	Assimetria da vazão diária	$\text{prob}(\text{sint} > \text{hist}) =$	9,0%	11,0%
4	Curtose da vazão diária	$\text{prob}(\text{sint} > \text{hist}) =$	18,0%	5,0%
5	Média da vazão máxima anual	$\text{prob}(\text{sint} > \text{hist}) =$	44,0%	99,0%
6	Desvio-padrão da vazão máxima anual	$\text{prob}(\text{sint} > \text{hist}) =$	39,0%	0,0%
7	Assimetria da vazão máxima anual	$\text{prob}(\text{sint} > \text{hist}) =$	23,0%	0,0%
8	Curtose da vazão máxima anual	$\text{prob}(\text{sint} > \text{hist}) =$	26,0%	0,0%
9	Vazão máxima de todo o período	$\text{prob}(\text{sint} > \text{hist}) =$	27,0%	0,0%
10	Correlação lag 1 da vazão média diária	$\text{prob}(\text{sint} > \text{hist}) =$	0,0%	99,0%
		$\text{prob}(d > d_{\text{hist}}) =$	(fórmula (18)) = 83,0%	100,0%
		Nota =	(fórmula (22)) = -0,38	-10,47

O quadro 4 apresenta os resultados de porcentagem acima, saltos e ascensões para o período histórico e para os dois modelos.

Quadro 4 - Estatísticas dos hidrogramas

	Histórico	Yakowitz	Diana
Saltos	9,20%	13,20%	16,90%
Acima	27,70%	28,40%	26,10%
Ascensões	23,70%	22,30%	40,10%

Os dois modelos saíram-se bem nas estatísticas dos hidrogramas, como era de se esperar. Ocorreu, porém uma discrepância entre os resultados do histórico e os do modelo Diana na estatística *ascensões*: a série gerada por este modelo resultou em muito mais ascensões do que a série histórica (40,1% contra 23,7%).

O auto-correlograma de vazões diárias para a série histórica e para os dois modelos de geração de vazões é apresentado na figura 2.

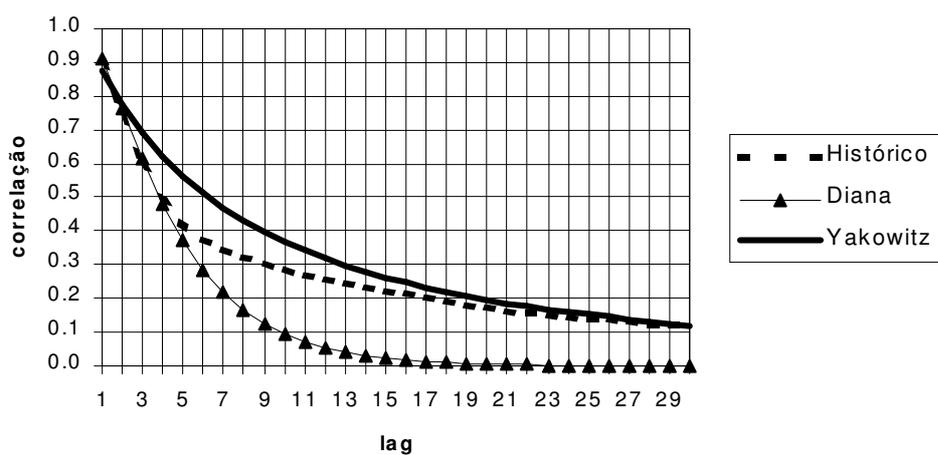


Figura 2 - Autocorrelograma de vazões diárias

A figura 2 mostra que o modelo Diana reproduz melhor as autocorrelações de lag menor do que o modelo Yakowitz, mas este se comporta de maneira mais adequada nos lags maiores.

A figura 3 apresenta seis meses de vazões diárias geradas pelo modelo Diana em Rio dos Patos, enquanto que a figura 4 apresenta a mesma quantidade de vazões gerada pelo modelo de Yakowitz.

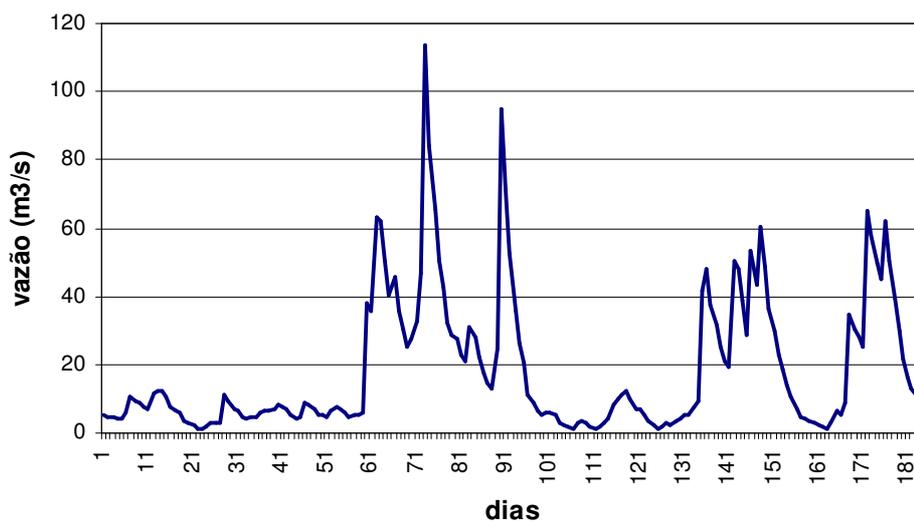


Figura 3 - Seis meses de vazões diárias geradas pelo modelo Diana em Rio dos Patos

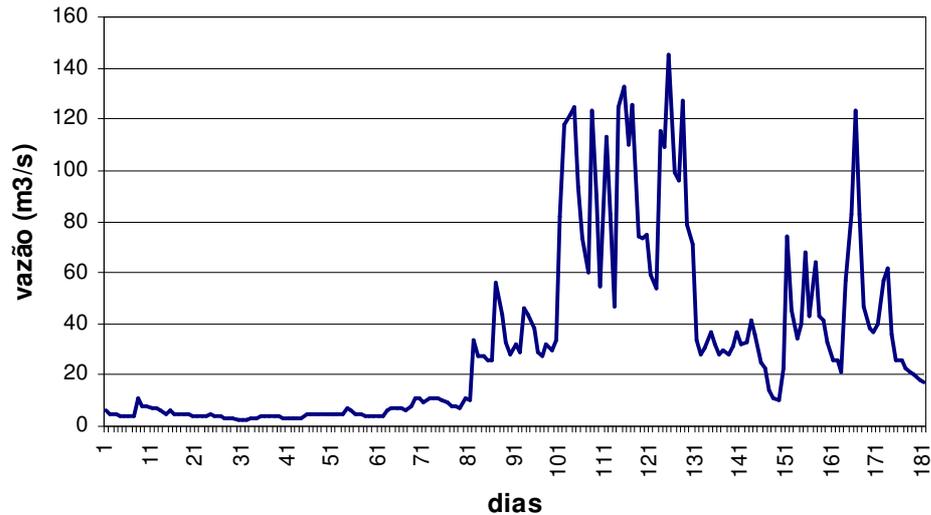


Figura 4 - Seis meses de vazões diárias em Rio dos Patos gerados pelo modelo de Yakowitz

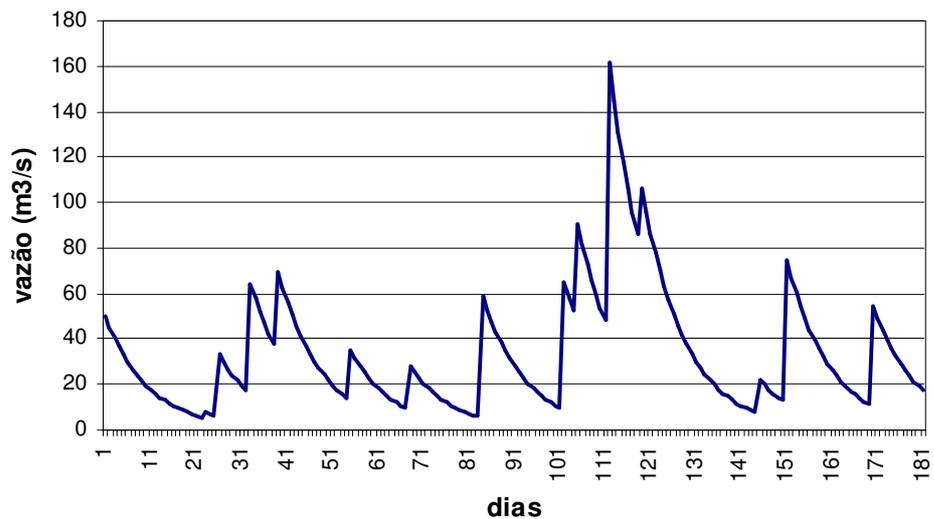


Figura 5 - Seis meses de vazões diárias em Rio dos Patos gerados pelo método shot noise de primeira ordem

As figuras 3 e 4 mostram que os hidrogramas gerados pelos dois modelos não-paramétricos analisados neste estudo têm uma aparência bastante próxima daqueles do período histórico. Apenas a título ilustrativo, a figura 5 apresenta seis meses de vazão diária gerados em Rio dos Patos pelo método paramétrico shot noise (Weiss, 1977) conforme metodologia descrita em Müller (1997). Os hidrogramas do shot noise têm uma aparência de serra dentada (Weiss, 1977), diferentemente do que ocorre nos fluviogramas histórico.

Quanto à comparação visual entre os dos dois modelos não-paramétricos objeto do presente estudo, nota-se que as linhas dos hidrogramas gerados pelo modelo Yakowitz - apesar do bom

ajuste mostrado nas porcentagens da hidrógrafa (Quadro 4) - são sutilmente mais irregulares do que aquelas do histórico, o que provavelmente acabou acarretando uma autocorrelação lag 1 menores no modelo de Yakowitz do que a do período histórico, conforme comentado anteriormente na análise do quadro 3.

5 – CONCLUSÕES

Os modelos não-paramétricos estudados no presente trabalho, de Yakowitz (1979) e Diana (Kelman et al, 1983) apresentam hidrogramas com boa similitude com as do histórico, embora não preservem estatísticas sazonais.

Nas estatísticas de vazões diárias de longo período os resultados foram similares para os dois períodos, enquanto que na análise de máximos o modelo de Yakowitz apresentou melhores resultados, provavelmente por causa do bom comportamento da fórmula introduzida por Barbosa (1981).

As autocorrelações diárias de lag até 4 mostraram um resultado melhor para o modelo Diana do que para Yakowitz, enquanto que esta situação se inverte nos lags maiores.

Em termos puramente visuais, os hidrogramas do modelo Diana parecem um pouco mais consistentes com os do histórico, resultado causado provavelmente pelo melhor ajuste das autocorrelações mais baixas.

BIBLIOGRAFIA

BARBOSA, P. R. D. J. *Modelagem estocástica de vazões diárias*. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 1981. Dissertação (Mestrado em Engenharia Civil). Universidade Federal do Rio de Janeiro.

MULLER, F. *Geração univariada de vazões diárias por métodos diretos*. Curitiba: Universidade Federal do Paraná, 1997. Dissertação (Mestrado em Engenharia Hidráulica). Universidade Federal do Paraná.

MULLER, F. "*Dois modelos de desagregação de vazões mensais em diárias*". Revista Brasileira de Recursos Hídricos, v.3, n.2, 1998. p.23-46.

KAVISKI, E. *Métodos de regionalização de eventos e parâmetros hidrológicos*. Curitiba: Universidade Federal do Paraná, 1992. Dissertação (Mestrado em Engenharia Hidráulica). Universidade Federal do Paraná.

KELMAN, J. DAMÁZIO, J.M. e COSTA, J.P. "*Geração de séries sintéticas de vazões diárias - modelo Diana*". Revista Brasileira de Eng., v.1, n.2, 1983. p.5-22.

WEISS, G. *"Shot noise models for the generation of synthetic streamflow data"*. Water Resources Res., v.13, n.1, Feb. 1977. pp. 101-108.

YAKOWITZ, S. J. *"A nonparametric Markov model for daily river flow"*. Water Resources Res., v.15, n.5, Oct. 1979. pp. 1035-1043.