



## ABORDAGEM DE MÉDIAS MÓVEIS SIMÉTRICAS PARA GERAÇÃO DE SÉRIES SINTÉTICAS DE VAZÃO

*Daniel Henrique Marco Detzel<sup>1\*</sup> & Miriam Rita Moro Mine<sup>2</sup>*

**Resumo** – Mostra-se no presente trabalho uma aplicação da abordagem de médias móveis simétricas (SMA) para a geração de cenários sintéticos de vazão em escala anual. Trata-se de um algoritmo elaborado para a modelagem de séries que apresentam o chamado comportamento de Hurst-Kolmogorov, associado à interação entre fenômenos que se manifestam em múltiplas escalas temporais. Diferentemente de estudos tradicionais, nos quais as consequências dessa interação são tratadas como manifestação de um comportamento não estacionário (e.g. tendências), a técnica SMA considera essas flutuações como característica natural das séries. O estudo apresenta o equacionamento do algoritmo e uma aplicação em seis séries de vazões anuais provenientes da Bacia do Rio Iguaçu, no Paraná. Os resultados evidenciam a potencialidade dessa abordagem, ao preservar perfeitamente as estatísticas básicas das séries e gerar eventos de baixa probabilidade de ocorrência referentes a épocas de secas e cheias relevantes para estudos hidrológicos.

**Palavras-Chave** – Séries sintéticas, comportamento de Hurst-Kolmogorov, modelos estocásticos.

## SYMMETRIC MOVING AVERAGE APPROACH FOR SYNTHETIC STREAMFLOW TIME SERIES GENERATION

**Abstract** – A symmetric moving average (SMA) approach application for synthetic annual streamflow scenarios generation is presented. It is an algorithm proposed to model time series with the so-called Hurst-Kolmogorov behavior, associated to multiple time scales phenomenon interaction. Apart from traditional approaches where the consequences of this interaction are considered as nonstationarities (e.g. trends), SMA considers it as a natural phenomenon expected behavior. The algorithm presentation is followed by an application in six annual streamflow time series, collected from Iguaçu River Basin, in Paraná State. Results show perfect agreement between synthetic and historical time series statistics. Moreover, SMA algorithm provided the generation of low probability events, associated to droughts and floods relevant to hydrological studies.

**Keywords** – Synthetic time series, Hurst-Kolmogorov behavior, stochastic models.

## INTRODUÇÃO

No contexto da modelagem de séries hidrológicas de vazão, algumas questões recentes têm cumprido um papel importante na escolha das técnicas estatísticas e matemáticas utilizadas por pesquisadores. A maior delas faz referência ao comportamento não estacionário das séries, ou seja, a

<sup>1</sup> Doutorando do Programa de Pós Graduação em Engenharia de Recursos Hídricos e Ambiental da UFPR, daniel@lactec.org.br

<sup>2</sup> Professora Titular do Departamento de Hidráulica e Saneamento da UFPR, mrmine.dhs@ufpr.br

\* Autor Correspondente



variação de seus momentos estatísticos no tempo. Essa característica vem sendo explicada pelas ações antrópicas nas bacias hidrográficas e também pela questão da variabilidade/mudança climática (Milly *et al.*, 2008).

Para contornar esses problemas e permitir o uso de tradicionais modelos estocásticos, algumas técnicas matemáticas passaram a ser incorporadas nas formulações, como por exemplo, a representação da alteração do regime hidrológico através de tendências lineares (portanto determinísticas), detectadas com auxílio de técnicas estatísticas (e.g. Fleming e Weber, 2012). No entanto, alguns autores questionam esse artifício principalmente por não ser possível obter uma função determinística que explique adequadamente as tendências observadas. Ademais, análises sobre séries de longa duração permitiram observar que as flutuações existentes seriam mais bem modeladas através de uma abordagem integralmente estocástica (Koutsoyiannis, 2006).

Esse enfoque se traduz em uma linha de pensamento ainda pouco explorada na literatura hidrológica e, por esse motivo, é o tema central do presente artigo. Ao contrário das técnicas tradicionais, considera-se que as flutuações observadas nas séries são parte de seu comportamento natural esperado, fruto da interação de eventos que ocorrem em diversas escalas temporais. Dessa maneira, utilizam-se os conceitos de processos estocásticos agregados em uma classe de modelos que abordam múltiplas escalas temporais, cujo objetivo é, justamente, representar a citada interação entre eventos. A base teórica dessa modelagem está no chamado comportamento de Hurst-Kolmogorov (HK – Koutsoyiannis, 2013), que atesta que a magnitude da interação pode ser mensurada através do expoente de Hurst.

Até então a aplicação dos modelos de múltiplas escalas temporais esteve limitada a séries hidrológicas particularmente longas, como a sequência de níveis mínimos anuais do rio Nilo medida na ilha de Rhoda em Cairo, no Egito, que totaliza 663 anos (Koutsoyiannis, 2002). Evidentemente, séries hidrológicas com tal comprimento são extremamente raras (senão únicas), fazendo com que a aplicação desses modelos em séries de comprimento condizente com as trabalhadas na hidrologia operacional seja uma investigação de grande potencial. Assim, propõe-se aplicar essa abordagem em séries anuais da Bacia do Rio Iguaçu, no Paraná, com auxílio do algoritmo de médias móveis simétricas, ferramenta específica para a modelagem de séries nas quais é importante a preservação do comportamento HK.

## **ALGORITMO DE MÉDIAS MÓVEIS SIMÉTRICAS**

O algoritmo de médias móveis simétricas (SMA, do inglês *Symmetric Moving Average*) é uma das três abordagens propostas por Koutsoyiannis (2003) para a geração de variáveis aleatórias que obedecem os princípios dos modelos de múltiplas escalas temporais. As outras duas técnicas mostram níveis diferentes de precisão, variantes de acordo com a distribuição de probabilidades marginal dos dados, assimetria e peso computacional. Entretanto o SMA é o único aplicável a séries com qualquer distribuição marginal (gaussiana ou não) e que garante a reprodução da assimetria das séries, característica importante quando se trabalha com registros hidrológicos de vazão.

Como o próprio nome sugere, a base do algoritmo SMA está na definição de um processo (infinito) puro de médias móveis, tal como mostrado em Box *et al.*, (2008, p. 9-10). Koutsoyiannis (2002) propôs limitar esse processo a um número finito  $q$  de elementos, compreendendo-o a um somatório entre  $-q \leq t \leq q$  e de forma que o elemento  $t = 0$  fique centrado na série, resultando na equação (1):



$$y_t = \sum_{j=-q}^q \theta_{|j|} a_{t+j} = \theta_q a_{t-q} + \dots + \theta_1 a_{t-1} + \theta_0 a_t + \theta_1 a_{t+1} + \dots + \theta_q a_{t+q} \quad (1)$$

onde  $y_t$  é a variável aleatória que representa o fenômeno considerado (no caso, a série de vazão),  $a_t$  é a série de resíduos do modelo,  $\theta_j$  são coeficientes (ou pesos) a serem estimados e  $q$  é a variável que delimita o processo, como comentado anteriormente. Ressalta-se que no caso do processo SMA, o número  $q$  de elementos é significativamente maior do que as ordens dos modelos estocásticos lineares (e.g. MA[ $q$ ], ver Box *et al.*, 2008) tradicionalmente utilizados. A escolha pelo valor de  $q$  é feita de acordo com o número de termos da função de autocorrelação (FAC) da série a ser reproduzido pelo modelo. Alternativamente, Koutsoyiannis (2002) propõe um critério matemático para sua seleção, mostrado pela equação (2):

$$q \geq \max \left[ m, \left( \frac{H^2 - 0,25}{2\beta} \right)^{1/(1,5-H)} \right] \quad (2)$$

onde  $m$  é, no mínimo, o número de termos significativos da FAC amostral,  $H$  é o expoente de Hurst e  $\beta$  é a tolerância aceita para o erro da estimativa de  $\theta_0$  (tolerâncias de  $\beta \leq 0,001$  são consideradas suficientemente pequenas). Em específico, coeficiente  $\theta_0$  tem direta relação com a autocovariância do processo, como será mostrado na sequência.

A estimativa dos coeficientes  $\theta_q$  é feita através das equações (3), propostas por Koutsoyiannis (2002):

$$\begin{cases} \theta_0 = \frac{\sqrt{(2-2H)\gamma_0}}{1,5-H} \\ \theta_j = \frac{\theta_0}{2} [(j+1)^{H+0,5} + (j-1)^{H+0,5} - 2j^{H+0,5}], \quad j > 0 \end{cases} \quad (3)$$

onde  $\gamma_0$  é a autocovariância do processo (que em aplicações práticas pode ser substituída pela variância amostral da série) e  $H$  é novamente o expoente de Hurst. Em meio aos diversos métodos de estimação desse expoente, sugere-se o emprego de um dos três procedimentos propostos em Tyralis e Koutsoyiannis (2011), pois eles permitem estimar  $H$  e a variância do processo simultaneamente. Para o presente artigo, optou-se pelo estimador de mínimos quadrados baseados no desvio padrão, técnica iterativa detalhada em Koutsoyiannis (2003).

Como dito no início da seção, o algoritmo SMA pode trabalhar com séries assimétricas, o que é desejável quando as variáveis em questão são vazões em rios. Para isso, é necessário aplicar as equações (4), que estabelecem as relações entre os três primeiros momentos da série modelada com os resíduos utilizados no modelo SMA (Koutsoyiannis, 2002):



$$\left\{ \begin{array}{l} E[a_t] = \hat{\mu}_a = \hat{\mu}_y / \left( \theta_0 + 2 \sum_{j=1}^q \theta_j \right) \\ VAR[a_t] = \hat{\sigma}_a^2 = 1 \\ \hat{\xi}_a = \hat{\xi}_y \gamma_0^{3/2} / \left( \theta_0^3 + 2 \sum_{j=1}^q \theta_j^3 \right) \end{array} \right. \quad (4)$$

onde  $\hat{\mu}$ ,  $\hat{\sigma}^2$  e  $\hat{\xi}$  são os respectivos estimadores amostrais da média, variância e assimetria das séries de vazão  $y$  e dos resíduos  $a$ ,  $E[\ ]$  e  $VAR[\ ]$  são os operadores esperança e variância e  $\gamma_0$  é a autocovariância do processo que, novamente, pode ser substituída pela variância amostral da série.

Tendo-se os três primeiros momentos estatísticos para a série de resíduos, a assimetria pode ser modelada aplicando-se uma distribuição de probabilidades conhecida assimétrica. No presente trabalho, optou-se pela distribuição log-normal a três parâmetros, utilizada para a geração dos números pseudoaleatórios submetidos à equação principal do modelo.

## ESTUDO DE CASO

### Área de estudo

O local escolhido para aplicação do método estudados se localiza na Bacia Hidrográfica do Rio Iguaçu (área de 70.799 km<sup>2</sup>), uma das sub-bacias que contribuem para a Bacia Hidrográfica do Rio Paraná. Esta bacia se encontra em uma região de clima subtropical úmido, com inverno fresco a frio na porção oeste e com inverno frio na porção leste, segundo classificação oferecida por Mendonça e Danni-Oliveira (2007, p. 151).

Em termos hidrológicos, a área de estudo selecionada é conhecida por não ter sazonalidade bem definida e pelas baixas autocorrelações entre anos consecutivos. Além disso, a região sofre com fenômenos atmosférico-oceânicos tais como El Niño, que causa reflexos como aumento da temperatura e do volume de precipitações repercutindo diretamente no regime de vazões dos rios da bacia. Por esses e outros fatores, a área oferece grandes desafios para modelagem.

Para o trabalho, selecionaram-se seis postos fluviométricos cujas principais características são descritas na Tabela 1. Partiu-se da premissa de escolher séries com histórico de boa duração (mínimo de 50 anos) e com dados consistidos.

Tabela 1 – Postos fluviométricos utilizados (dados coletados da Agência Nacional de Águas – ANA, via portal Hidroweb – <http://hidroweb.ana.gov.br/>)

Nome	Cód. ANA	Duração	Área (km <sup>2</sup> )	MLT* (m <sup>3</sup> /s)
São Bento (SB)	65155000	1931-2005	2012	34,8
União da Vitória (UV)	65310000	1930-2007	24211	468,4
Porto Amazonas (PA)	65035000	1931-2005	3662	67,9
Jangada do Sul (JS)	65370000	1946-2005	1055	25,9
Rio Negro (RN)	65100000	1930-2007	3379	67,7
Fazenda Maracanã (FM)	65415000	1946-2005	323	9,0

\*MLT: Média de longo termo



## Premissas do estudo

Para cada localidade foi gerado um conjunto de 1000 séries sintéticas por modelo, cada uma com comprimento de 77 anos, equivalente ao das séries de maior número de elementos (UV e RN). As métricas de comparação foram constituídas primeiramente em estatísticas de curto termo tais como médias, desvios padrão, coeficientes de assimetria, vazões mínimas e máximas e autocorrelações. A seguir, determinaram-se estatísticas de longo termo, relacionadas às sequências de vazões consecutivas acima ou abaixo da MLT das respectivas séries históricas, tanto em termos de números de sequências quanto em termos de vazões acumuladas em cada uma. Calcularam-se também os déficits médios e máximos necessários para suprir a regularização de um reservatório hipotético, assumindo que 80% da afluência média da série seja requerida para tal objetivo. Trabalha-se com um sistema sem consideração de falhas, mas, ainda assim, este índice é uma boa ferramenta de análise de desempenho de modelos.

## RESULTADOS E DISCUSSÕES

### Análises preliminares

Previamente à apresentação dos resultados da aplicação do algoritmo SMA às séries estudadas, mostram-se as estimativas do expoente  $\hat{H}$  e da quantidade  $q$  de pesos que limitam o processo. A Tabela 2 reúne esses resultados, lembrando-se que para a quantidade  $q$  foi utilizada a equação (2) com uma tolerância  $\beta = 0,001$ .

Tabela 2 – Resultados dos expoentes  $\hat{H}$  e quantidade de elementos  $q$  utilizados

Estação	$\hat{H}$	$q$
SB	0,67	243
UV	0,64	174
PA	0,67	246
JS	0,59	74
RN	0,73	655
FM	0,67	246

No caso do expoente  $\hat{H}$ , nota-se que JS e RN são as séries com menor e maior estimativas, respectivamente. No entanto, de uma maneira geral, o que se percebe é que os expoentes têm valores apreciáveis em todas as estações, evidenciando o comportamento HK definido por Koutsoyiannis (2013). Entende-se, portanto, que as flutuações em múltiplas escalas temporais estão presentes nas séries estudadas, justificando a utilização do algoritmo SMA.

Para a quantidade  $q$ , nota-se que os resultados são reflexo direto da magnitude do expoente de Hurst em cada série, resultado esperado pela própria interpretação do critério matemático da equação (2). Assim, as estações com os maiores  $\hat{H}$  foram as que apresentaram as maiores estimativas para  $q$ . Na intenção de uniformizar o procedimento de geração, adotou-se como padrão o valor arredondado do maior  $q$  calculado (RN), resultando em  $q = 700$  para todas as séries.

### Análise das séries sintéticas geradas

A Tabela 3 exibe as estatísticas de curto termo obtidas com a aplicação do algoritmo SMA às séries estudadas. Para as séries sintéticas, os resultados se referem à média dos cenários gerados. Nota-se que todas as estatísticas foram reproduzidas fielmente pelo modelo, com uma pequena subestimação da assimetria nas estações de PA e JS. A exatidão obtida com a reprodução do expoente de Hurst atesta que a implementação do algoritmo foi feita com sucesso.

Tabela 3 – Estatísticas de curto termo das séries sintéticas geradas

Posto	Posto	Média (m <sup>3</sup> /s)	D. Padrão (m <sup>3</sup> /s)	Assimetria	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	Hurst
SB	Histórico	34,8	12,6	1,15	14,7	80,4	0,67
	SMA	34,8	12,2	1,14	15,6	77,8	0,66
UV	Histórico	468,4	184,4	1,08	183,6	1127,8	0,64
	SMA	468,4	180,6	1,03	172,7	1088,9	0,64
PA	Histórico	67,9	25,4	1,02	27,3	151,9	0,67
	SMA	67,6	24,5	0,92	26,7	147,5	0,66
JS	Histórico	25,9	9,5	0,97	8,4	54,5	0,59
	SMA	25,8	9,2	0,80	10,0	53,9	0,58
RN	Histórico	67,7	25,5	0,90	25,5	154,4	0,73
	SMA	67,5	24,3	0,84	24,8	146,6	0,73
FM	Histórico	9,0	3,5	1,35	3,6	22,1	0,67
	SMA	9,0	3,4	1,37	4,3	21,2	0,66

A análise das autocorrelações foi feita utilizando os gráficos das FAC para todas as estações, porém sua exibição está limitada somente a RN (Figura 1). Os motivos foram porque a estrutura de persistência de todas as séries é muito semelhante e RN é a que possui o maior valor da autocorrelação de primeira ordem. Da mesma forma, as FAC obtidas com o algoritmo SMA foram parecidas em todas as séries, respeitando a primeira defasagem e com posterior decaimento para dentro do limite de significância estatística. Visto que a série histórica também possui rápido decaimento, considera-se o desempenho do modelo satisfatório.

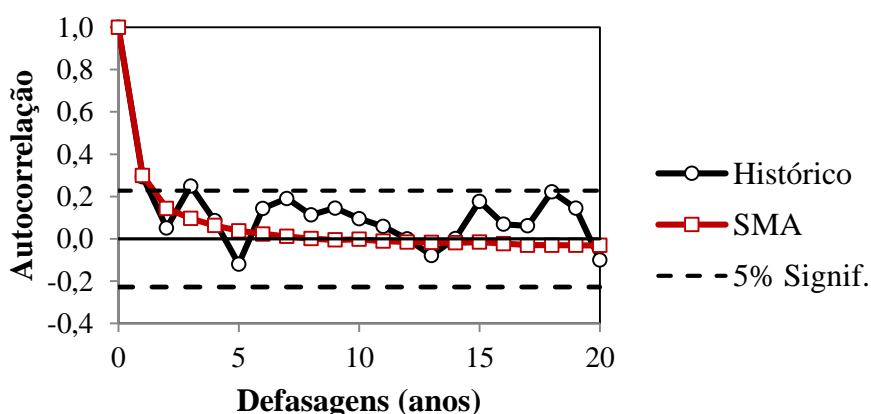


Figura 1 – Comparação da FAC histórica e modelada para RN



A exibição dos resultados prossegue com as estatísticas de longo termo, mostradas na Tabela 4. Alerta-se que a interpretação desses resultados distingue-se da feita para as estatísticas de curto termo, quando o objetivo era a reprodução das propriedades observadas na série histórica. Em particular, as estatísticas de longo termo evidenciam as características da abordagem SMA que, claramente, geram eventos severos de vazões baixas e altas. Isso pode ser inferido ao notar que praticamente todas as métricas calculadas resultaram em valores maiores que a série histórica.

Tabela 4 – Estatísticas de longo termo das séries sintéticas geradas

Posto	Série	Nº médio de seq.	Nº máx. de seq.	Vazões acum. médias (m <sup>3</sup> /s)	Vazões acum. máx. (m <sup>3</sup> /s)	Déficits acum. médios (m <sup>3</sup> /s)	Déficits acum. máx. (m <sup>3</sup> /s)
SB	Histórico	3	7	86,2	187,1	4,3	29,9
	SMA	4	8	102,7	218,2	5,1	33,5
UV	Histórico	3	7	1046,8	2531,8	64,7	370,6
	SMA	4	8	1294,2	2739,4	93,0	563,2
PA	Histórico	4	7	183,1	352,2	9,1	58,0
	SMA	4	8	190,2	392,3	11,8	72,5
JS	Histórico	4	6	72,1	122,1	2,7	12,3
	SMA	3	7	66,9	127,6	3,4	21,7
RN	Histórico	3	7	170,4	382,5	10,2	72,5
	SMA	4	9	201,5	433,2	15,7	92,9
FM	Histórico	4	7	26,4	44,6	1,7	7,1
	SMA	4	8	27,4	54,8	1,4	8,3

Considerando as premissas dos modelos de múltiplas escalas temporais discutidas na introdução do trabalho, percebe-se que os resultados obtidos foram bastante coerentes. A geração de cenários extremos e com baixa probabilidade de ocorrência pode ser tratada como um resultado importante, pois a reprodução desses eventos por modelos estocásticos tradicionais é notadamente complexa.

## CONCLUSÕES

A proposta de estruturação e o ajuste do algoritmo SMA nas séries da Bacia do Rio Iguaçu foram feitas com sucesso. As estatísticas de curto termo foram muito bem representadas em todas as situações, enquanto que as métricas de longo termo permitiram atestar que o modelo é capaz de gerar eventos extremos, o que pode ser de grande importância em estudos hidrológicos diversos. Dessa maneira, atesta-se que a utilização dessa abordagem para a modelagem de séries de vazões é promissora, visto que, diferentemente de modelos mais tradicionais, não há a necessidade de se verificar premissas como distribuição de probabilidades marginal dos dados ou ainda a presença de tendências.

Como trabalhos futuros, estão previstos a generalização do algoritmo para o caso multivariado, objetivando a geração de séries sintéticas que respeitem a estrutura de correlações espaciais entre as estações. Na sequência, pretende-se trabalhar em uma forma de se incorporar a sazonalidade na formulação, permitindo a modelagem de séries em escala mensal.



## **REFERÊNCIAS**

- BOX, G.E.P.; JENKINS, G.M.; REINSEL, G.C. (2008). *Time Series Analysis Forecasting and Control*. 4th ed. New Jersey: John Wiley & Sons, 746 p.
- FLEMING, S.W., WEBER, F.A. (2012). Detection of long-term change in hydroelectric reservoir inflows: Bridging theory and practise. *Journal of Hydrology*, 470-471, pp. 36–54.
- KOUTSOYIANNIS, D. (2002). The Hurst phenomenon and fractional Gaussian noise made easy. *Hydrological Sciences Journal*, 47(4), pp. 573–596.
- KOUTSOYIANNIS, D. (2003). Climate change, the Hurst phenomenon, and hydrological statistics. *Hydrological Sciences Journal*, 48(1), pp. 3–24.
- KOUTSOYIANNIS, D. (2006). Nonstationarity versus Scaling in Hydrology. *Journal of Hydrology*, 324(1-4), pp. 239–254.
- KOUTSOYIANNIS, D. (2013). Hydrology and change. *Hydrological Sciences Journal*, 58(6), pp. 1177–1197.
- MENDONÇA, F.; DANNI-OLIVEIRA, I.M. (2007). *Climatologia: noções básicas e climas do Brasil*. São Paulo: Oficina de Textos, 208 p.
- MILLY, P.C.D., BETANCOURT, J., FALKENMARK, M., HIRSCH, R.M., ZBIGNIEW, W., LETTENMAIER, D.P., STOUFFER, R.J. (2008). Stationarity Is Dead: Whither Water Management? *Science*, 319, pp. 573–574.
- TYRALIS, H., KOUTSOYIANNIS, D. (2011). Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. *Stochastic Environmental Research and Risk Assessment*, 25(1), pp. 21–33.