

SUAVIZAÇÃO MSSA NA PREVISÃO DE PRECIPITAÇÃO PLUVIOMÉTRICA NAS CINCO REGIÕES GEOGRÁFICAS DO BRASIL

João Marcos Amorim dos Santos

Departamento de Estatística – Universidade Federal Fluminense
Bloco G, 3º. Andar, Campus Gragoatá, São Domingos, Niterói – RJ, 24.210-201.
joaomarcosest@hotmail.com

Moisés Lima de Menezes

Departamento de Estatística – Universidade Federal Fluminense
Bloco G, 3º. Andar, Campus Gragoatá, São Domingos, Niterói – RJ, 24.210-201.
moises_lima@id.uff.br

Keila Mara Cassiano

Departamento de Estatística – Universidade Federal Fluminense
Bloco G, 3º. Andar, Campus Gragoatá, São Domingos, Niterói – RJ, 24.210-201.
keilamath@hotmail.com

RESUMO

SSA é uma técnica de suavização de séries temporais. MSSA é uma extensão de SSA para múltiplas séries temporais simultaneamente. Este artigo apresenta uma abordagem deste método de suavização, na qual os modelos de amortecimento exponencial de Holt-Winters e de Box & Jenkins são aplicados. Na ocasião, séries temporais de precipitação pluviométrica de cidades das cinco regiões geográficas do Brasil são modeladas antes e após a suavização SSA e MSSA. Os resultados obtidos mostram que tanto a modelagem de Holt-Winters quanto o modelo de Box & Jenkins são beneficiados pela suavização prévia da série temporal por SSA/MSSA e que a suavização MSSA supera SSA além do que os modelos de Box & Jenkins apresentam maior aderência aos dados originais.

Palavra-chave: MSSA; Precipitação Pluviométrica; Box & Jenkins; Holt-Winters.

ABSTRACT

SSA is a time series smoothing technique. MSSA is an extension of SSA to multiple time series simultaneously. This paper presents an approach to this smoothing method in which the exponential damping models of Holt-Winters and Box & Jenkins are applied. At the time, rainfall time series of cities of the five geographic regions of Brazil are modeled before and after the SSA and MSSA smoothing. The results show that both the Holt-Winters model and the Box & Jenkins model are benefited by the earlier smoothing of the time series by SSA / MSSA and that the MSSA smoothing outperforms SSA in addition to the Box & Jenkins models have higher adherence to the original date.

Keywords: MSSA; Rainfall; Box & Jenkins; Holt-Winters.

Como Citar:

SANTOS, João Marcos Amorim dos; MENEZES, Moisés Lima de; CASSIANO, Keila Mara. Suavização MSSA na previsão de precipitação pluviométrica nas cinco regiões

geográficas do Brasil. In: SIMPÓSIO DE PESQUISA OPERACIONAL E LOGÍSTICA DA MARINHA, 19., 2019, Rio de Janeiro, RJ. **Anais** [...]. Rio de Janeiro: Centro de Análises de Sistemas Navais, 2019.

1. INTRODUÇÃO

Precipitação pluviométrica é um fenômeno meteorológico que possui aleatoriedade tanto em sua ocorrência quanto em sua intensidade. O conhecimento prévio sobre precipitações pluviométricas é de suma importância no planejamento de geração de energia hidrelétrica e de abastecimento d'água das grandes cidades. Prevenções como racionamento de água ou controle de nível de uso da água são exemplos de como se pode fazer uso de uma modelagem e previsão adequada desta precipitação (BEIJO *et al.*, 2005).

Singular Spectrum Analysis (SSA) é um método para análise e previsão de séries temporais e incorpora elementos de análise de séries temporais clássica, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamento de sinais (GOLYANDINA *et al.*, 2001). SSA pode ser aplicada em vários campos como matemática, física, economia, meteorologia, oceanografia e ciências sociais (HASSANI, 2007).

Com base na decomposição em valores singulares (*singular value decomposition - SVD*), o método SSA decompõe uma série de tempo em uma soma de componentes, de tal forma que em cada componente se concentre uma parte da energia contida na série temporal. Além disso, um pequeno conjunto de componentes concentra a maior parte da estrutura de autocorrelação contida na série temporal, enquanto os restantes podem ser considerados como componentes de ruído. Portanto, o conjunto de componentes pode ser classificado em dois grupos mutuamente exclusivos e exaustivos: sinal e ruído. O primeiro grupo contém os componentes que possuem algumas informações sobre a estrutura da série, enquanto o último abrange as componentes ruidosas. Assim, uma versão suavizada da série temporal pode ser obtida pela soma das componentes de sinal. Tradicionalmente, uma série de tempo pode ser expressa como a soma de componentes: tendência, harmônica e ruído. A remoção do componente de ruído pode contribuir para uma melhor identificação do processo estocástico subjacente da série temporal e, conseqüentemente, melhorar a precisão das previsões. Jayawardena & Gurung (2000) mostram que a remoção de ruídos em série hidrológica melhora a precisão das previsões. Uma boa revisão do assunto é apresentada em (ELSHORBAGY *et al.*, 2002).

Multi-channel Singular Spectrum Analysis (MSSA) é a extensão natural de SSA para uma série temporal multivariada. Os procedimentos SSA e MSSA oferecem um método de pré-tratamento de dados. Portanto, ambos os processos podem ser utilizados a fim de melhorar o ajuste de modelos de previsão e ajudar a melhorar a precisão das previsões (HASSANI & MAHMOUDVAND, 2013). Mas quando o interesse reside na previsão de um conjunto de séries temporais múltiplas qual procedimento é o mais adequado? Temos de aplicar o procedimento SSA para cada série temporal individualmente ou aplicamos o procedimento MSSA ao conjunto de séries temporais? Para responder as perguntas acima, neste artigo são apresentados os resultados de um experimento computacional onde os procedimentos de SSA e MSSA foram aplicados a séries temporais de precipitação pluviométrica de cidades das cinco regiões do Brasil.

Em seguida, os modelos de amortecimento exponencial de Holt-Winters e de Box & Jenkins foram aplicados para cada série temporal suavizada obtida por SSA e MSSA. Os resultados mostram que SSA contribui para os ajustes tanto de Holt-Winters quanto de Box & Jenkins no modelo de cada série, porém a contribuição de MSSA é ainda maior. As estatísticas de aderência utilizadas para verificar qual modelo tem maior ganho preditivo são

o MAPE (*Mean Absolute Percentage Error*), RMSE (*Root Mean Square Error*), BIC (*Bayesian Information Criterion*) e o coeficiente de determinação R^2 e os programas Forecast Pro for Windows (FPW), E-views, CaterpillarSSA e Microsoft Excel foram utilizados para auxiliar na parte operacional.

O presente artigo está organizado em seis seções. A seguir, na seção 2, há uma breve descrição de SSA. Na seção 3 tem-se uma descrição resumida de MSSA. As modelagens e as estatísticas de aderência são apresentadas na seção 4. Na seção 5 está o estudo de caso com os resultados obtidos. Enfim, na seção 6, são apresentadas as principais conclusões do trabalho.

2. SINGULAR SPECTRUM ANALYSIS

A versão básica do método SSA pode ser dividida em duas etapas: decomposição e reconstrução.

2.1. DECOMPOSIÇÃO

A etapa de decomposição pode ser subdividida em incorporação e SVD.

Seja $Y_T = [y_1, \dots, y_T]$ uma série temporal de comprimento T . Por incorporação, entende-se como sendo um procedimento no qual uma série temporal Y_T é levada a uma matriz $X = [X_1, \dots, X_K]_{L \times K}$, onde $X_k = [y_k, \dots, y_{k+L-1}]^T$, para todo $k = 1 \dots K$, onde $K = T - L + 1$. A matriz X é conhecida como matriz trajetória (HASSANI, 2007) e o parâmetro L , que assume algum valor inteiro no intervalo $2 \leq L \leq T$, é o comprimento da janela da matriz trajetória (GOLYANDINA et al., 2001). A matriz trajetória X pode ser expandida, via SVD, em (1).

$$X = E_1 + E_2 + E_3 + \dots + E_L, \tag{1}$$

onde $E_i = \lambda_i^{1/2} U_i V_i^T$ e os conjuntos $\{\lambda_i^{1/2}\}_{i=1}^L$ e $\{U_i\}_{i=1}^L$ são, respectivamente, denominados *espectro singular* e *vetores singulares* da matriz trajetória X . A coleção (λ_i, U_i, V_i) é conhecida como *autotripla* na SVD da matriz trajetória X . A contribuição de cada componente em (1) pode ser mensurada pela razão de valores singulares, dada por $(\lambda_i)^{1/2} / \sum_{i=1}^L (\lambda_i)^{1/2}$. Considere que d seja o posto (isto é, o número de autovalores não nulos) da matriz trajetória X . Segue que a identidade descrita em (1) pode ser reescrita tal como em (2).

$$X = \sum_{i=1}^d E_i, \text{ onde } d \leq L. \tag{2}$$

2.2. RECONSTRUÇÃO

A etapa de reconstrução pode ser subdividida em agrupamento e média diagonal. A etapa de agrupamento consiste no procedimento de agrupar algumas sequências de matrizes elementares resultantes da decomposição SVD em grupos disjuntos e, após isso, somá-las, gerando novas matrizes elementares.

Considerando a sequência $\{E_i\}_{i=1}^d$ de matrizes elementares na SVD em (2) e

agrupando-as em $m \leq d$ grupos disjuntos e assumindo que, após o agrupamento, o conjunto de índices gerado é dado por $\{I_1, \dots, I_m\}$, onde, para todo i , $I_i = \{I_{i1}, \dots, I_{ip_i}\}$ e p_i é a cardinalidade do grupo I_i , então a matriz elementar X_{I_i} gerada a partir do grupo $\{X_{I_{ij}}\}_{j=1}^{p_i}$ é dada por $X_{I_i} = \sum_{j=1}^{p_i} X_{I_{ij}}$, de modo que a identidade em (2) pode ser reescrita como em (3).

$$X = \sum_{i=1}^m X_{I_i} \quad (3)$$

É importante salientar que o procedimento de agrupamento pode ser realizado também sob a sequência $\bigcup_{i=1}^m \{X_{I_{ij}}\}_{j=1}^{p_i}$ de igual forma ao realizado na sequência $\{E_l\}_{l=1}^d$ utilizando, inclusive, um método ou critério de agrupamento diferente do aplicado sob a sequência $\{E_l\}_{l=1}^d$ em (2). Um dos objetivos da análise de agrupamentos é separar as componentes do sinal das componentes do ruído. A contribuição da componente X_{I_i} pode ser mensurada pela razão de valores singulares dada por $\sum_{j=1}^{p_i} (\lambda_{I_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$.

Dada a matriz trajetória X e assumindo que $L^* = \min(L, K)$ e que $K^* = \max(L, K)$, considere que $x_{lk}^{(i)}$ seja um elemento na linha l e coluna k na matriz X_{I_i} . O elemento $y_t^{(i)}$ da componente SSA $[y_t^{(i)}]_{1 \times T}$ é calculado por meio da média diagonal, que é definida em (4), a partir da matriz elementar X_{I_i} .

$$y_t^{(i)} = \begin{cases} \frac{\sum_{l=1}^t x_{l,t-l+1}^{(i)}}{t}, & \text{se } 1 \leq t < L^* \\ \frac{\sum_{l=1}^{L^*} x_{l,t-l+1}^{(i)}}{L^*}, & \text{se } L^* \leq t < K^* \\ \frac{\sum_{l=t-K^*+1}^{T-K^*+1} x_{l,t-l+1}^{(i)}}{T-K^*+1}, & \text{se } K^* \leq t \leq T \end{cases} \quad (4)$$

Cada componente $[y_t^{(i)}]_{1 \times T}$ concentra parte da energia da série temporal original

$[y_t]_{1 \times T}$ que pode ser mensurada pela razão de valores singulares $\sum_{j=1}^{p_i} (\lambda_{I_{ij}})^{1/2} / \sum_{l=1}^d (\lambda_l)^{1/2}$.

De acordo com Golyandina *et al.* (2001), as componentes SSA $[y_t^{(i)}]_{1 \times T}$ podem ser classificadas em três categorias: *tendência*, *componentes harmônicas* e *ruído*.

2.3. CORRELAÇÃO PONDERADA

Por correlação ponderada (*w*-correlation) entende-se como sendo a função que quantifica a dependência linear entre duas componentes SSA $Y_t^{(i)}$ e $Y_t^{(j)}$. A correlação ponderada é dada por

$$\rho_{ij}^{(w)} = \frac{(Y_t^{(i)}, Y_t^{(j)})_w}{\|Y_t^{(i)}\|_w \|Y_t^{(j)}\|_w},$$

Onde $\| \cdot \|$ é a norma euclidiana, $(\cdot)_w$ é o produto interno tal que:

$$\|Y_t^{(i)}\|_w = \sqrt{(Y_t^{(i)}, Y_t^{(i)})_w}, \quad (Y_t^{(i)}, Y_t^{(j)})_w = \sum_{k=1}^T w_k y_k^{(i)} y_k^{(j)} \quad e \quad w_k = \min\{k, L, T - k\}.$$

Por meio da separabilidade, pode-se verificar estatisticamente se as duas componentes SSA estão bem separadas, em termos de dependência linear. Se o valor absoluto $\rho_{ij}^{(w)}$ é pequeno), então as componentes SSA correspondentes são classificadas como *w*-ortogonais (ou quase *w*-ortogonais). Caso contrário, são ditas mal separadas (HASSANI, 2007).

A análise das coordenadas da série temporal na base definida pelos vetores singulares resultantes da SVD permite identificar as componentes de tendência e da sazonalidade da série. O problema geral aqui consiste em identificar e separar as componentes oscilatórias das componentes que fazem parte da tendência. De acordo com Golyandina *et al.* (2001) a análise gráfica de tais coordenadas aos pares permite identificar por meio visual as componentes harmônicas da série.

As coordenadas da série temporal em duas componentes ortogonais podem ser dispostas em um diagrama de dispersão. Considere um harmônico puro com frequência igual a ω , fase igual a δ , amplitude igual a ξ e período $\rho = 1/\omega$ definido como um divisor do comprimento da janela L e K . Se o parâmetro ρ assume um valor inteiro, então ρ é classificado como período do harmônico. Por exemplo, as funções seno e o cosseno com frequências, amplitudes e fases iguais resultam em um diagrama de dispersão que exibe um padrão circular. Por sua vez, se $\rho = 1/\omega$ é um inteiro, então o diagrama de dispersão exibe um polígono regular com ρ vértices. Para uma frequência $\omega = m/n < 0.5$ com m e n inteiros e primos, os pontos são vértices de um polígono regular de n vértices (GOLYANDINA *et al.*, 2001). Dessa forma, a identificação dos componentes que são gerados por um harmônico é reduzida à análise pictórica do padrão determinado nos diferentes pares de componentes.

3. MULTI-CHANNEL SINGULAR SPECTRUM ANALYSIS

MSSA é uma extensão de SSA para trabalhar com análise e previsão de séries temporais multidimensionais. O procedimento MSSA segue a mesma estrutura do procedimento SSA com a diferença de fazê-lo usando um conjunto de séries temporais em face de uma única série.

Considere o sistema de s séries temporais de comprimento T , em (5)

$$Y^{(k)} = \left(y_t^{(k)} \right)_{t=1}^T \tag{5}$$

onde $k = 1, \dots, s$. O caso particular do procedimento MSSA para $s = 1$, equivale ao procedimento usando SSA (GOLYANDINA & STEPANOV, 2005).

Escolhendo um único comprimento de janela de defasagem L para todas as s séries, onde $1 < L < T$, na fase de incorporação MSSA são obtidos $K = T - L + 1$ vetores defasados $X_j^{(k)} = \left(y_j^{(k)}, \dots, y_{j+L-1}^{(k)} \right)^T$, $j = 1, \dots, K$ para cada série $Y^{(k)}$, $k = 1, \dots, s$. Assim, para cada série $Y^{(k)}$, é possível obter uma matriz trajetória como em (6)

$$X^{(k)} = \begin{bmatrix} y_1^{(k)} & y_2^{(k)} & \dots & y_K^{(k)} \\ y_2^{(k)} & y_3^{(k)} & \dots & y_{K+1}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ y_L^{(k)} & y_{L+1}^{(k)} & \dots & y_T^{(k)} \end{bmatrix} \tag{6}$$

A matriz trajetória da série multidimensional $(Y^{(1)}, Y^{(2)}, \dots, Y^{(s)})$ é, então, uma matriz de dimensão $Ls \times K$ e tem a seguinte forma:

$$X = \left[X_1^{(1)} : \dots : X_K^{(1)} : \dots : X_1^{(s)} : \dots : X_K^{(s)} \right]^T = \left[X^{(1)} : \dots : X^{(s)} \right]^T \tag{7}$$

O espaço trajetória é definido por um espaço linear *spanado* pelos vetores defasados (colunas da matriz trajetória X).

A partir de $S = XX^T$, onde os autovalores de S em ordem de magnitude são $\lambda_1 \geq \dots \geq \lambda_{Ls} \geq 0$ e U_1, \dots, U_{Ls} são os respectivos autovetores associados, $d = \max \{j: \lambda_j > 0\}$ é o posto da matriz S e $V_j = \frac{X^T U_j}{\sqrt{\lambda_j}}$, $j = 1, \dots, d$. Denotando $E_j = \sqrt{\lambda_j} U_j V_j^T$, então o SVD da matriz trajetória X pode ser escrito por: $X = E_1 + \dots + E_d$.

De forma análoga, a fase de agrupamento particiona o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos disjuntos I_1, \dots, I_m de modo que a matriz trajetória seja reescrita na forma conhecida como decomposição agrupada:

$$X = X_{I_1} + \dots + X_{I_m} \tag{8}$$

Por fim a média diagonal é aplicada a cada uma das séries decomposta em (8) e então o grupo de sinais reconstruídos dado por $\tilde{Y}^{(k)} = \left(\tilde{Y}_t^{(k)} \right)_{t=1}^T$, $k = 1, \dots, s$ é obtido.

4. MÉTODOS PREDITIVOS

Nesta seção são apresentados os métodos preditivos a serem testados na previsão de séries filtradas e não filtradas via SSA e MSSA. Os métodos testados neste artigo são os modelos de Amortecimento Exponencial de Holt-Winters e os modelos ARIMA. Além disso, são apresentadas as estatísticas de aderência e a metodologia proposta.

4.1. MODELOS DE HOLT-WINTERS

De acordo com Morettin & Toloi (2001), os métodos de amortecimento exponencial baseiam-se na premissa de que os dados são diferentemente ponderados. Comumente, as observações recentes contêm informações mais relevantes que as antigas, de forma que a ponderação dos dados (série temporal) decresce exponencialmente à medida que a observação se torna mais antiga. Um caso particular dos métodos de amortecimento exponencial é o método de Holt-Winters multiplicativo, o qual realiza a modelagem de forma dinâmica (isto é, com parâmetros variantes no tempo) das componentes: nível ($a_{1,t}$), tendência ($a_{2,t}$) e sazonalidade ($\rho_{m(t)}$), conforme a equação (9).

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ y_t = [a_{1,t} + (a_{2,t} \times t)] \times \rho_{m(t)} + \varepsilon_t, \end{cases} \quad (9)$$

onde ε_t é um erro estocástico, y_t é o valor observado em t e $\rho_{m(t)}$ é o fator sazonal em t relativo ao mês m . De acordo com Hamilton (1994), a família $\{\rho_{m(\cdot)}\}_{m \in \Gamma}$ de fatores sazonais, onde Γ é o conjunto com todos os meses no ano, deve obedecer à restrição $\sum_{m \in \Gamma} \rho_{m(\cdot)} = S$, sendo S o tamanho do ciclo sazonal (quando o ciclo sazonal é anual e a periodicidade dos dados é mensal, então $S = 12$). No processo de estimação dos parâmetros da equação (9), são utilizados três hiperparâmetros (quantidades invariantes no tempo), denotados por α , β e γ os quais estão associados, respectivamente, às estimativas de nível, tendência e sazonalidade (quantidades variantes no tempo). De acordo com Morettin & Toloi (2001), os valores ótimos dos hiperparâmetros α , β e γ se encontram na região convexa $[0, 1]^3$ e tem como objetivo a minimização da estatística *MSE* (*mean square error*).

4.2. MODELOS DE BOX & JENKINS

De acordo com Hamilton (1994), um processo estocástico estacionário de segunda ordem é definido como uma família $\{Y_t\}_{t=1}^T$ de variáveis aleatórias cujos momentos (média, variância e covariância) são invariantes no tempo. Isto é: $Y_t \sim Dist(\mu; \sigma^2)$ para todo t . Considere que a sequência $\{y_t\}_{t=1}^T$ seja uma realização de $\{Y_t\}_{t=1}^T$ (isto é, uma série temporal estacionária de segunda ordem), onde y_t consiste na notação da observação em t . Box &

Jenkins (1970) propõem a equação (10) para a modelagem de sua dinâmica temporal $\{y_t\}_{t=1}^T$.

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (10)$$

onde ε_t é um erro estocástico e y_t é o valor observado em t . O modelo (10) pode ser representado em termos de dois polinômios obtidos através do operador defasagem B (Hamilton, 1994), definido por $B^d Y_t = Y_{t-d}$ e possui formulação geral definida em (11)

$$(1 - \phi_1 B^1 - \dots - \phi_p B^p) y_t = (1 - \theta_1 B^1 - \dots - \theta_q B^q) \varepsilon_t, \quad (11)$$

onde $\phi_k, \forall k \geq 1$ e $\theta_j, \forall j \geq 1$ denotam os parâmetros do modelo *ARMA* (autorregressivo e de médias móveis) ε_t e y_t são como definidos em (10). Em (11) as defasagens de y_t representam a parte autoregressiva de ordem p (*AR(p)*), enquanto as defasagens dos erros representam a parte média móvel de ordem q (*MA(q)*). O exame de estacionariedade de segunda ordem de $\{y_t\}_{t=1}^T$ pode ser realizado por meio da análise da função de autocorrelação simples (ρ_k), definida em (12).

$$\rho_k = (\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})) / \sum_{t=1}^T (y_t - \bar{y})^2, \quad (12)$$

onde \bar{y} é a média da série temporal e k é a defasagem (*lag*) da autocorrelação. A determinação das ordens p e q em (10) é realizada pela análise do perfil das funções de autocorrelação (*ACF*) ρ_k e autocorrelação parcial (*PACF*) ϕ_{kk} cujos padrões teóricos podem ser verificados em Souza & Camargo (2004). Uma vez identificada a ordem do modelo *ARMA*, os coeficientes das partes *AR(p)* e *MA(q)* são estimados e as respectivas significâncias estatística são avaliadas (HAMILTON, 1994). Em especial, caso uma série temporal $\{y_t\}_{t=1}^T$ não estacionária na média, mas estacionária na variância e covariância, deve-se diferenciá-la d vezes, a fim de se gerar uma nova série temporal que apresente estacionariedade na média (HAMILTON, 1994). Para tais classes de séries temporais, pode-se utilizar o modelo *ARIMA(p, d, q)*, cuja formulação geral é descrita em (13).

$$(1 - \phi_1 B^1 - \dots - \phi_p B^p)(1 - B)^d y_t = (1 - \theta_1 B^1 - \dots - \theta_q B^q) \varepsilon_t. \quad (13)$$

Quando a série possui sazonalidade, o modelo de Box & Jenkins utilizado é o modelo *SARIMA(p, d, q) × (P, D, Q)_s*, descrito em (14).

$$\phi(B)\Phi(B^S)(1 - B)^d(1 - B^S)^D y_t = \theta(B)\Theta(B^S)\varepsilon_t, \quad (14)$$

onde $\phi(B)$ é o polinômio autorregressivo, $\Phi(B^S)$ é o polinômio autorregressivo

sazonal, $\theta(B)$ é o polinômio de média móvel, $\theta(B^S)$ é o polinômio de média móvel sazonal, d é o número de diferenças necessárias para tornar a série estacionária e D é o número de diferenças sazonais necessárias para tornar a série sazonalmente estacionária (MORETTIN & TOLOI, 2001).

4.3. ESTATÍSTICAS DE ADERÊNCIA

Os resultados obtidos ao longo dos experimentos computacionais realizados são comparados em termos das estatísticas de aderência: MAPE (*Mean Absolute Percentage Error*), BIC (*Bayesian Information Criterion*) e RMSE (*Root Mean Square Error*) que devem ser minimizadas e o coeficiente de determinação R^2 que deve ser maximizado:

$$MAPE = \sum_{t=1}^T \left[\frac{|y_t - \hat{y}_t|}{y_t} \right] \times \frac{1}{T}, \quad BIC = \ln(T)p - 2 \ln(M),$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}}, \quad R^2 = 1 - \left(\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \right),$$

onde p é o número de parâmetros do modelo, M é máximo valor da função de máxima verossimilhança, y_t é o valor observado e \hat{y}_t é a previsão, ambos para o instante t , onde $1 \leq t \leq T$.

4.4. METODOLOGIA PROPOSTA

O objetivo deste trabalho é avaliar o ganho preditivo nas modelagens de séries temporais quando se aplica a abordagem SSA e MSSA. Para isso, 5 séries temporais de precipitação pluviométrica, uma em cada região do Brasil, foi modelada via modelos de amortecimento exponencial de Holt-Winters e modelos ARIMA de Box & Jenkins sem e com a filtragem SSA e MSSA para análises e comparação via estatísticas de aderência. A Figura 1 apresenta um fluxograma desta proposta.

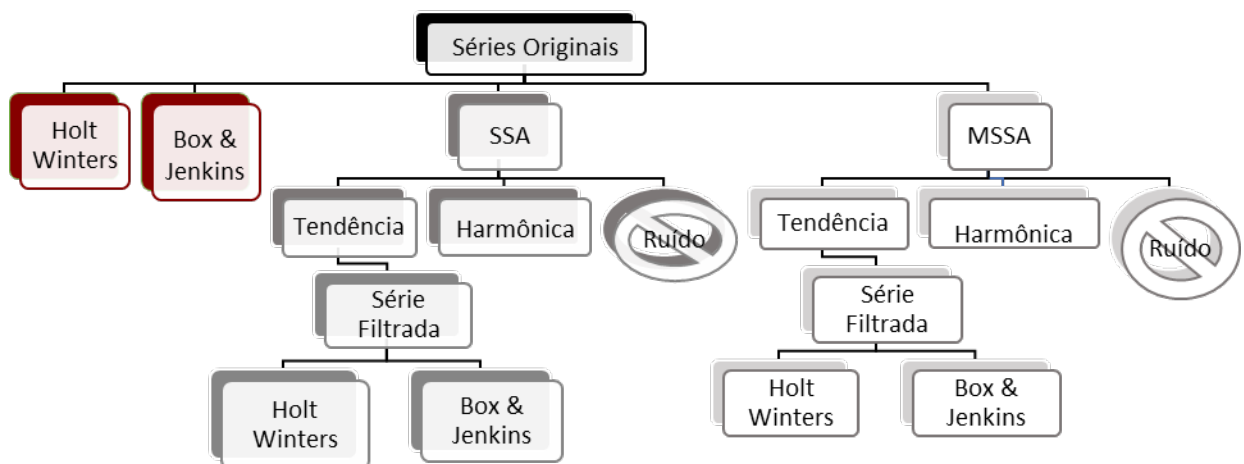


Figura 1. Fluxograma da Metodologia.

5. RESULTADOS E DISCUSSÃO

As cinco séries temporais de médias mensais de precipitação pluviométricas (medidas em *mm*) foram obtidas a partir da Agência Nacional de Águas (ANA) e do Instituto Nacional de Meteorologia (INMET) e datam de janeiro de 1970 a dezembro de 2014. Foram escolhidas cinco cidades espalhas pelo Brasil, sendo uma em cada região. São elas: Rio Branco (AC), Recife (PE), Brasília (DF), Rio de Janeiro (RJ) e Porto Alegre (RS). Cada série tem comprimento $T = 528$. As Figuras 2 – 6 apresentam as cinco séries originais.

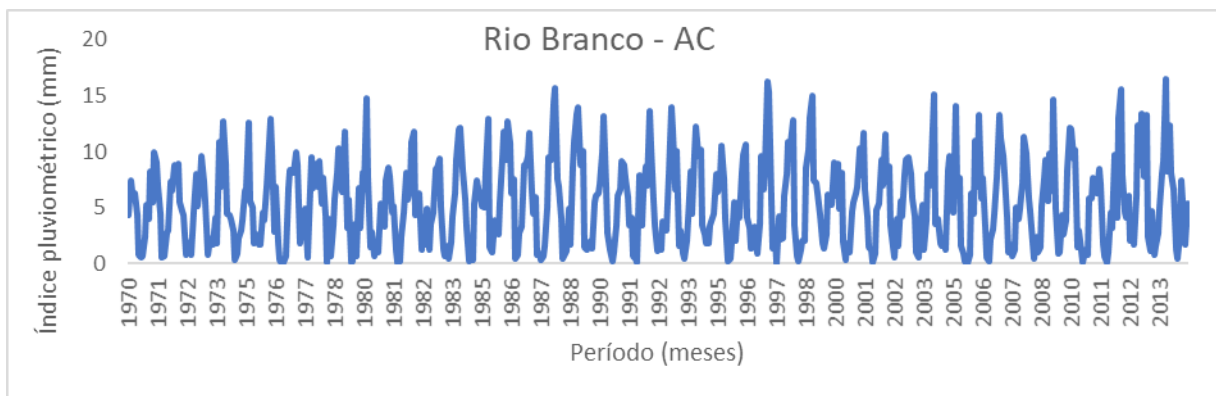


Figura 2. Médias mensais de precipitação pluviométrica (em *mm*) em Rio Branco (AC). Jan/1970 – dez/2014.

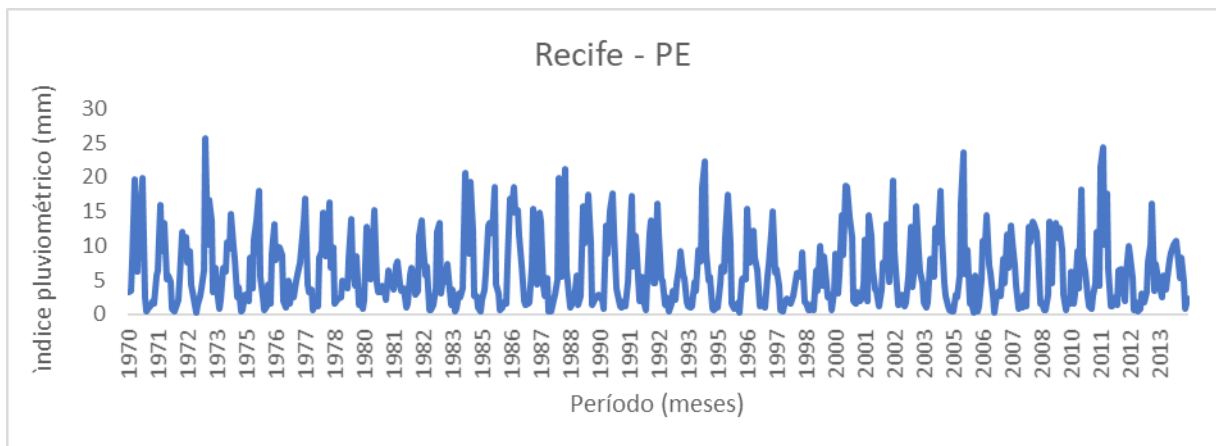


Figura 3. Médias mensais de precipitação pluviométrica (em *mm*) em Recife (PE). Jan/1970 – dez/2014.

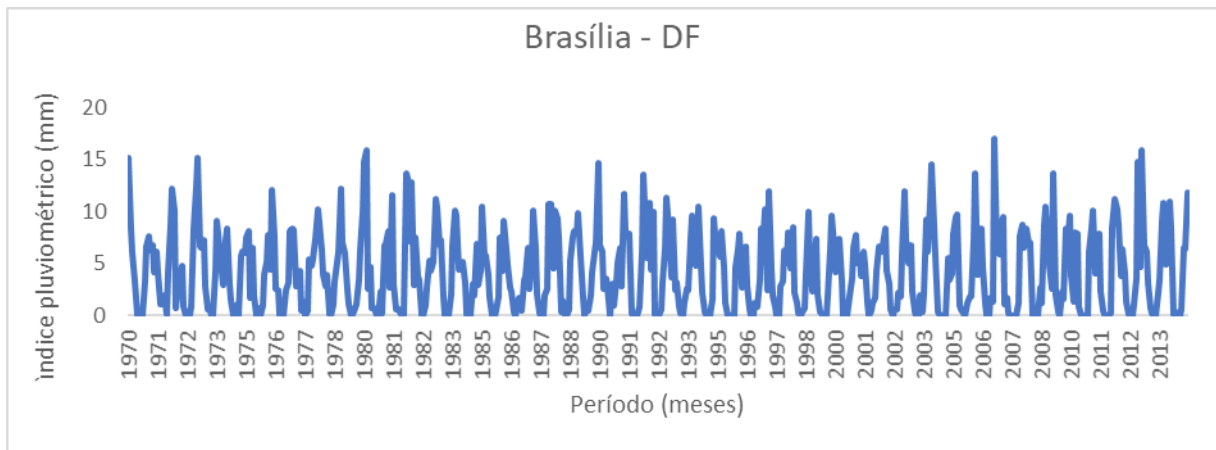


Figura4. Médias mensais de precipitação pluviométrica (em mm) em Brasília (DF). Jan/1970 – dez/2014.

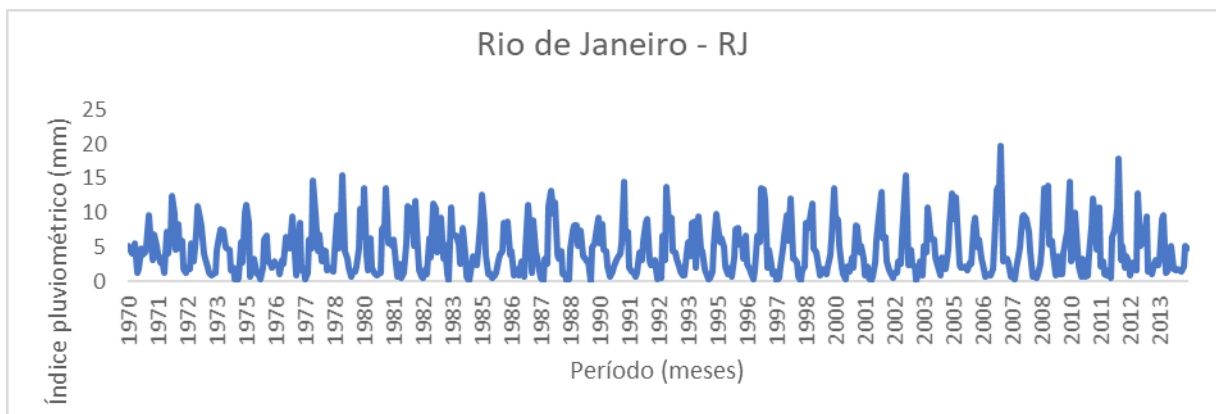


Figura 5. Médias mensais de precipitação pluviométrica (em mm) no Rio de Janeiro (RJ). Jan/1970 – dez/2014.

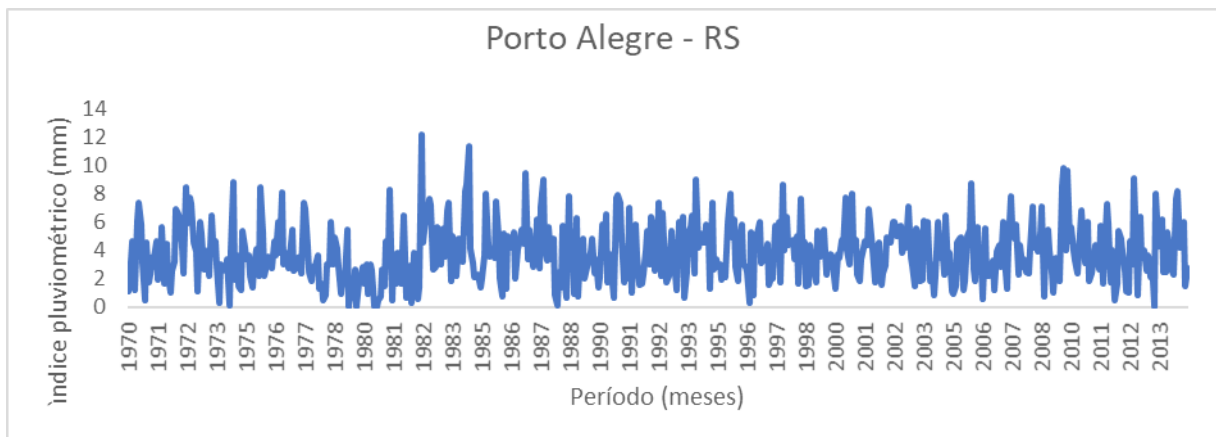


Figura 6. Médias mensais de precipitação pluviométrica (em mm) em Porto Alegre (RS). Jan/1970 – dez/2014.

A Tabela 1 apresenta as precipitações pluviométricas médias em cada mês, bem como o desvio padrão.

As estatísticas apresentadas na Tabela 1 mostram a diferença de comportamento das precipitações ao longo do ano, como por exemplo, em Recife, cujas maiores médias estão concentradas nos meses de maio, junho e julho enquanto em Brasília, estes meses estão entre os que têm as menores médias.

A Figura 7 apresenta um resumo da Tabela 1, mostrando o comportamento das precipitações pluviométricas a cada mês. Pode-se perceber que o comportamento em Recife vai na contramão das demais cidades.

Tabela 1. Médias e desvios-padrão de precipitação pluviométrica (em mm) por mês em cada cidade pesquisada.

Mês	Estatística	Rio Branco	Recife	Brasília	Rio de Janeiro	Porto Alegre
Jan	Média	9,27	3,30	7,27	8,53	3,45
	Desvio Padrão	3,08	2,17	3,56	3,97	1,69
Fev	Média	10,22	5,00	6,65	5,73	3,86
	Desvio Padrão	2,73	3,48	3,18	3,20	1,73
Mar	Média	8,33	6,93	6,53	5,28	3,26
	Desvio Padrão	2,99	3,96	2,66	2,54	1,61
Abr	Média	6,10	9,93	4,44	3,69	3,39
	Desvio Padrão	2,72	5,72	2,47	1,75	1,84
Mai	Média	3,08	10,29	0,99	2,15	3,45
	Desvio Padrão	1,76	4,57	0,86	1,01	1,89
Jun	Média	1,26	12,75	0,32	1,34	4,72
	Desvio Padrão	1,20	4,42	0,67	0,99	2,65
Jul	Média	1,16	11,37	0,24	1,39	4,42
	Desvio Padrão	1,11	4,62	0,57	1,01	2,33
Ago	Média	1,69	6,68	0,57	1,45	4,28
	Desvio Padrão	1,16	3,49	0,78	1,33	2,19
Set	Média	3,15	3,70	1,72	2,95	4,36
	Desvio Padrão	1,86	2,50	1,37	2,09	2,42
Out	Média	4,91	2,06	5,27	4,54	3,90
	Desvio Padrão	2,04	1,49	3,16	1,80	2,14
Nov	Média	6,65	1,32	7,90	7,89	3,53
	Desvio Padrão	2,39	0,81	2,76	3,12	2,14
Dez	Média	8,13	1,93	8,05	9,57	3,35
	Desvio Padrão	2,23	1,48	2,80	2,79	1,48

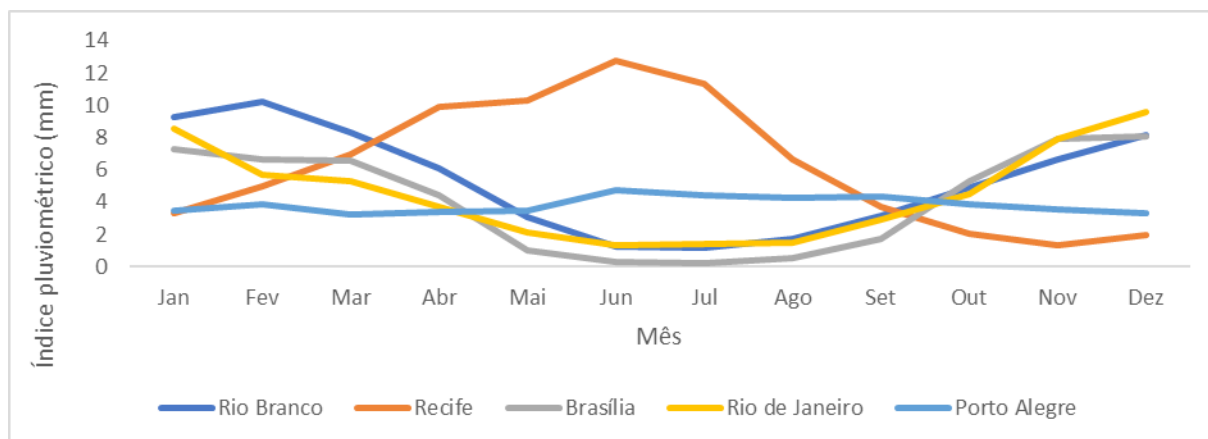


Figura 7. Médias mensais de precipitação pluviométrica (em mm) nas 5 cidades. Jan/1970 – dez/2014.

5.1. FILTRAGENS SSA E MSSA

De acordo com a proposta deste trabalho, as 5 séries serão ajustadas via modelos de amortecimento exponencial de Holt-Winters e modelos da classe ARIMA. Estas modelagens ocorrem sem e com as filtragens SSA e MSSA. Nos dois processos, o comprimento de janela utilizado foi $L = 264$, ou seja, $L = T/2$, como sugere Golyandina *et. al*, 2001. No processo de filtragem SSA, as 5 séries são filtradas separadamente. As séries são decompostas em

Tendência, Harmônica e Ruído e a componente ruidosa de cada série é removida, reconstruindo as séries menos ruidosas. A Tabela 2 apresenta as correlações ponderadas na decomposição das 5 séries na abordagem SSA.

Tabela 2. Correlação ponderada entre as componentes de cada série na filtragem SSA.

Série	Componente	Tendência	Harmônica	Ruído
Rio Branco	Tendência	1	0,000	0,001
	Harmônica	0,000	1	0,018
	Ruído	0,001	0,018	1
Recife	Tendência	1	0,000	0,005
	Harmônica	0,000	1	0,010
	Ruído	0,005	0,010	1
Brasília	Tendência	1	0,001	0,004
	Harmônica	0,001	1	0,023
	Ruído	0,004	0,023	1
Rio de Janeiro	Tendência	1	0,000	0,001
	Harmônica	0,000	1	0,013
	Ruído	0,001	0,013	1
Porto Alegre	Tendência	1	0,004	0,001
	Harmônica	0,004	1	0,053
	Ruído	0,001	0,053	1

Após verificar que as componentes estão bem separadas, a componente ruidosa de cada série é removida e as demais são somadas obtendo as cinco séries filtradas via SSA menos ruidosas.

No processo MSSA, as análises das séries são feitas simultaneamente. Na ocasião, as três componentes obtidas contêm informação das cinco séries. A Tabela 3 apresenta os resultados da correlação ponderada nesta abordagem.

Tabela 3. Correlação ponderada entre as componentes na filtragem MSSA.

Componente	Tendência	Harmônica	Ruído
Tendência	1	0,00	0,004
Harmônica	0,00	1	0,017
Ruído	0,004	0,017	1

De forma análoga ao processo SSA, após verificar a boa separação, a componente ruidosa é removida e as séries filtradas via MSSA são reconstruídas. As Figuras 8 – 12 apresentam as sobreposições das séries originais pelas séries filtradas via MSSA.

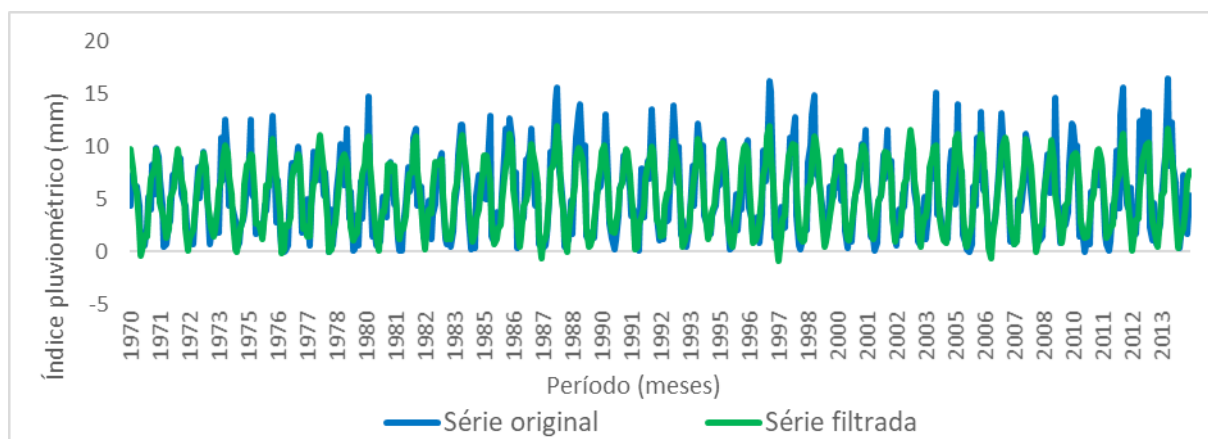


Figura 8. Sobreposição série original × série filtrada via MSSA. Rio Branco (AC). Jan/1970 – dez/2014.

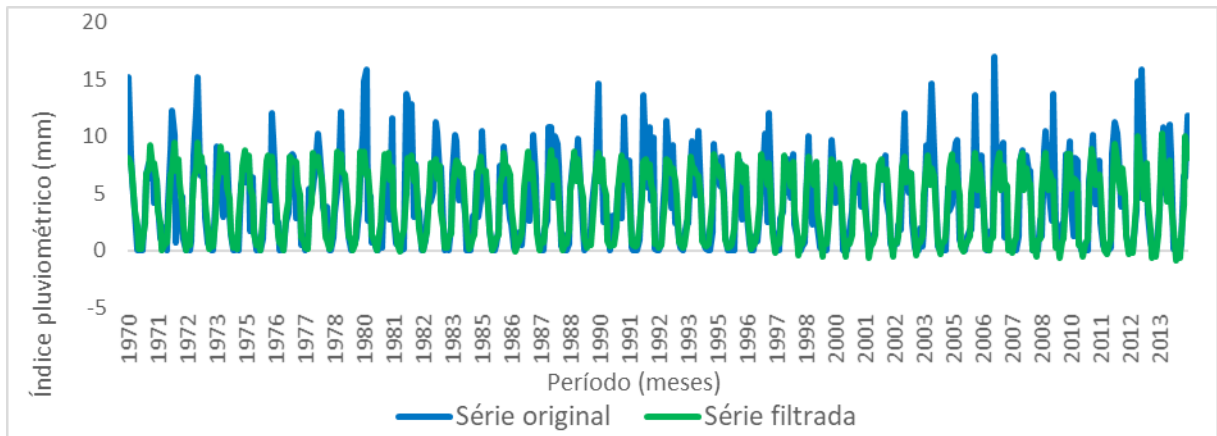


Figura 9. Sobreposição série original × série filtrada via MSSA. Brasília (DF). Jan/1970 – dez/2014.

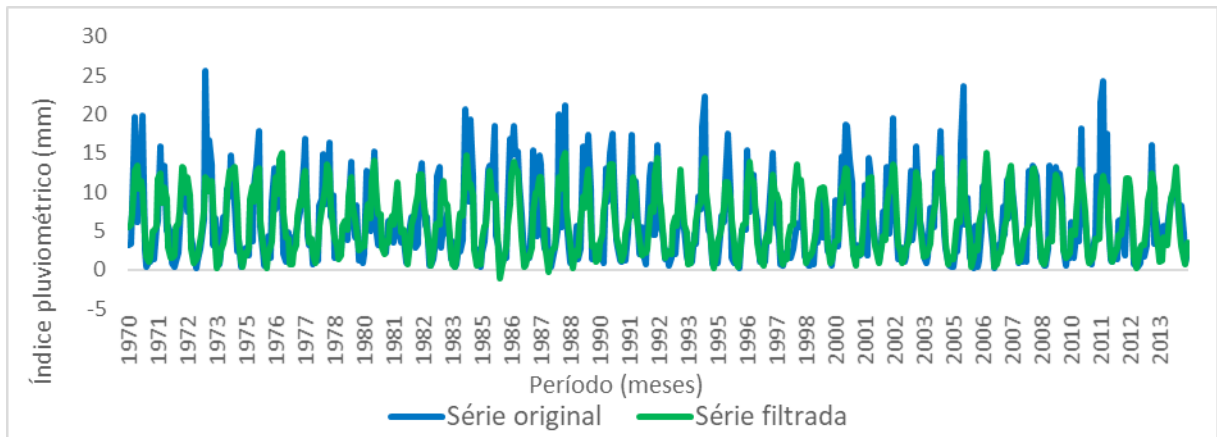


Figura 10. Sobreposição série original × série filtrada via MSSA. Recife (PE). Jan/1970 – dez/2014.

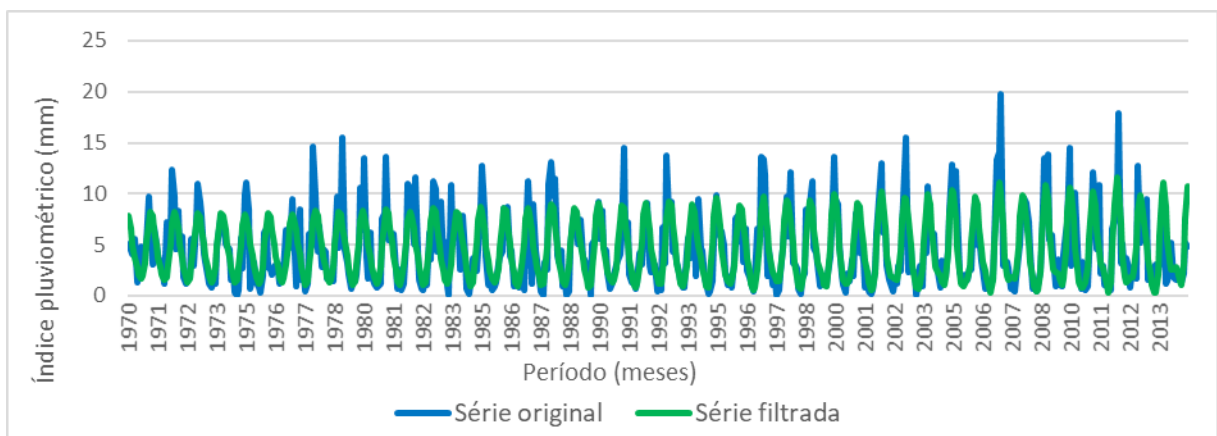


Figura 11. Sobreposição série original × série filtrada via MSSA. Rio de Janeiro (RJ). Jan/1970 – dez/2014.

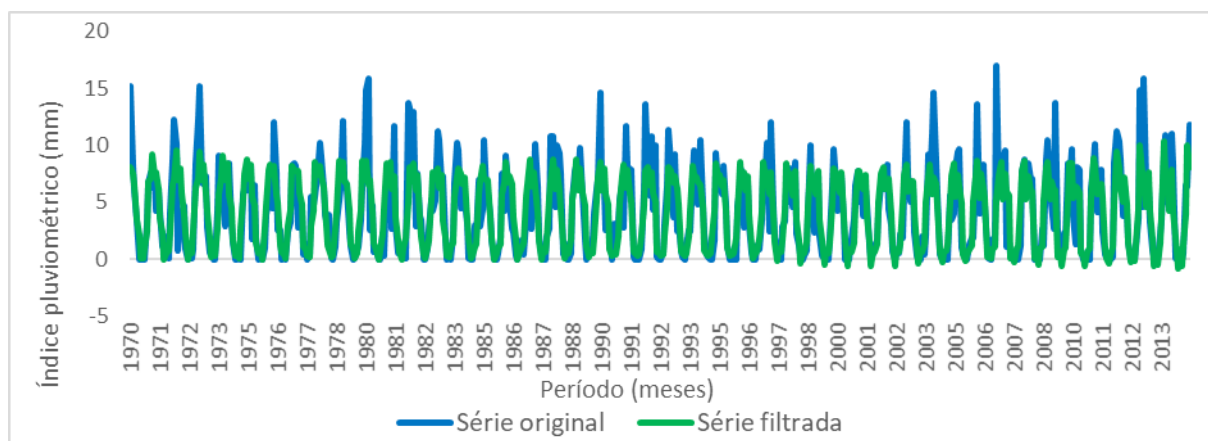


Figura 12. Sobreposição série original \times série filtrada via MSSA. Porto Alegre (RS). Jan/1970 – dez/2014.

Após as filtrações SSA e MSSA, as séries originais e filtradas são modeladas via modelos de Box & Jenkins e Holt-Winters. Diversos modelos foram testados na classe de Amortecimento Exponencial, desde modelos de tendência constante até modelos de tendência linear com sazonalidade. Os modelos de Box & Jenkins foram testados de acordo com as análises de correlogramas e de testes de normalidade e de estacionariedade. Em todos os casos foram feitas análises de resíduos para averiguar a adequação do modelo e para todos os casos estudados foram escolhidos os modelos que minimizavam as estatísticas de aderência e o BIC. A Tabela 4 apresenta os modelos campeões em cada situação.

Tabela 4. Modelos escolhidos com base nas estatísticas de aderência em cada caso.

Série	Modelo	Holt-Winters	Box & Jenkins
Rio Branco	Original	HW1	$SARMIA(1,0,4) \times (1,1,1)_{12}$
	SSA	HW1	$SARMIA(1,1,3) \times (1,1,2)_{12}$
	MSSA	HW1	$SARMIA(0,0,0) \times (1,0,1)_{12}$
Recife	Original	HW2	$SARMIA(1,0,0) \times (1,0,1)_{12}$
	SSA	HW3	$SARMIA(2,0,3) \times (1,1,3)_{12}$
	MSSA	HW1	$SARMIA(0,1,3) \times (1,1,2)_{12}$
Brasília	Original	HW4	$SARMIA(0,0,0) \times (2,0,1)_{12}$
	SSA	HW3	$SARMIA(1,0,3) \times (1,1,3)_{12}$
	MSSA	HW1	$SARMIA(1,1,4) \times (1,1,2)_{12}$
Rio de Janeiro	Original	HW1	$SARMIA(0,0,0) \times (2,0,1)_{12}$
	SSA	HW1	$SARMIA(2,0,2) \times (1,1,4)_{12}$
	MSSA	HW1	$SARMIA(0,1,3) \times (1,1,2)_{12}$
Porto Alegre	Original	HW1	$SARMIA(1,0,0) \times (1,0,0)_{12}$
	SSA	HW4	$SARMIA(2,0,3) \times (2,0,2)_{12}$
	MSSA	HW2	$SARMIA(1,1,1) \times (0,1,4)_{12}$

onde HW1 é o modelo de Holt-Winters com tendência constante e sazonalidade aditiva, HW2 é o modelo de Holt-Winters com tendência constante e sazonalidade multiplicativa, HW3 é o modelo de Holt-Winters com tendência linear e sazonalidade aditiva e HW4 é o modelo de Holt-Winters com tendência constante e sem sazonalidade.

As comparações das estatísticas de aderência para cada modelo de Amortecimento Exponencial de Holt-Winters em cada situação estão apresentadas na tabela 5 e as comparações destas estatísticas em relação aos modelos de Box & Jenkins estão na tabela 6.

Tabela 5. Estatísticas de Aderência para os modelos de Holt-Winters.

Série	Abordagem	<i>MAPE</i>	<i>RMSE</i>	<i>BIC</i>	R^2
Rio Branco	Original	1,1260	2,2210	2,2470	0,6569
	SSA	0,6744	0,7918	0,8011	0,9412
	MSSA	0,0515	0,1919	0,1942	0,9964
Recife	Original	0,7856	3,5680	3,6100	0,5265
	SSA	0,3257	1,1400	1,1600	0,9216
	MSSA	0,1759	0,7467	0,7554	0,9647
Brasília	Original	3,8670	3,5770	3,5980	0,1123
	SSA	0,2982	0,2378	0,2420	0,9940
	MSSA	0,2876	0,1910	0,1932	0,9960
Rio de janeiro	Original	0,9572	2,3460	2,3730	0,5786
	SSA	0,1334	0,4196	0,4245	0,9785
	MSSA	0,0591	0,2173	0,2199	0,9943
Porto Alegre	Original	0,8139	2,0440	2,0680	0,0392
	SSA	0,0952	0,4183	0,4232	0,5017
	MSSA	0,0536	0,2771	0,2804	0,7888

Tabela 6. Estatísticas de Aderência para os modelos de Box & Jenkins.

Série	Abordagem	<i>MAPE</i>	<i>RMSE</i>	<i>BIC</i>	R^2
Rio Branco	Original	1,0660	2,2230	2,2490	0,6564
	SSA	0,3290	0,4704	0,4899	0,9790
	MSSA	0,0046	0,0198	0,0207	1,0000
Recife	Original	0,6273	3,6130	2,6790	0,6032
	SSA	0,0946	0,3226	0,3400	0,9976
	MSSA	0,0283	0,1119	0,1572	0,9983
Brasília	Original	1,8510	2,3570	2,3980	0,6177
	SSA	0,1210	0,0635	0,0666	0,9996
	MSSA	0,0271	0,0289	0,0303	0,9999
Rio de janeiro	Original	0,9508	2,3450	2,3870	0,5718
	SSA	0,0324	0,1136	0,1430	0,9973
	MSSA	0,0105	0,0418	0,0463	0,9997
Porto Alegre	Original	0,8376	2,0818	2,0570	0,0187
	SSA	0,0068	0,0318	0,0335	0,9971
	MSSA	0,0054	0,0289	0,0310	0,9976

Como pode ser observado nas Tabelas 5 e 6, a filtragem SSA melhora a qualidade do ajuste tanto na modelagem de Holt-Winters quanto na modelagem de Box & Jenkins em todas as cidades analisadas uma vez que minimizam as estatísticas *MAPE*, *RMSE* e *BIC* e maximizam o R^2 . Pode-se perceber também que a filtragem MSSA supera o SSA uma vez que assumem estatísticas menores que as registradas na abordagem SSA e R^2 maiores que as filtragens quando feitas separadamente. Também é possível perceber, ao se fazer uma comparação transversal entre as tabelas 5 e 6, que os modelos de Box & Jenkins minimizam as estatísticas de aderência e maximizam o coeficiente de determinação em todas as abordagens e todas as cidades avaliadas, sendo esta classe de modelos a mais adequada para estas séries.

6. CONCLUSÕES

Este artigo propôs o uso da filtragem MSSA na modelagem de Holt-Winters e de Box & Jenkins de precipitação pluviométrica em cinco capitais espalhadas pelo Brasil, sendo

uma em cada região. Para isso, as cinco séries foram filtradas via SSA e MSSA de modo criterioso e as séries originais e filtradas foram modeladas nas duas classes de modelos analisadas com base nas premissas que permeiam as análises de séries temporais considerando diversos modelos para cada cidade. A escolha dos modelos adequados para cada caso se deu a partir das análises das estatísticas de aderência. Os resultados obtidos mostram que a filtragem SSA melhora a qualidade do ajuste tanto para as modelagens de Holt-Winters quanto nas modelagens de Box & Jenkins e que a filtragem MSSA supera a filtragem SSA nas mesmas modelagens. Outro resultado obtido indica que os modelos de Box & Jenkins são mais adequados para estas séries estudadas que os modelos de Holt-Winters. Sendo assim, são recomendáveis modelos de Box & Jenkins após a filtragem MSSA. Para trabalhos futuros, são sugeridas outras classes de modelos como Modelos de Inteligência Artificial ou Modelos Fuzzy associados a SSA e MSSA bem com outras metodologias de filtrações como, por exemplo, uso de Wavelets.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] BEIJO, L. A., Muniz, J. A., Neto, P. C. (2005). Tempo de retorno das precipitações máximas em Lavras (MG) pela distribuição de valores extremos do tipo I. *Ciência e Agrotecnologia (UFPA)*, v. 29, pp. 657-667.
- [2] GOLYANDINA, N., Nekrutkin, V., Zhihgljavsky, A. (2001). *Analysis of time series structure: SSA and related techniques*. Chapman&Hall/CRC. New York, USA.
- [3] Hassani, H. (2007). Singular Spectrum Analysis: Methodology and Comparison. *Journal of Data Science*. 5, pp. 239-257.
- [4] JAYAWARDENA, A.W., Gurung, A.B. (2000). Noise reduction and prediction of hydrometeorological time series: dynamical systems approach vs. stochastic approach, *Journal of Hydrology*, 228, pp. 242-264.
- [5] ELSHORBAGY, A., Simonovic, S.P., Panu, U.S. (2002). Noise reduction in chaotic hydrologic time series: facts and doubts, *Journal of Hydrology*, 256, pp. 147-265.
- [6] HASSANI, H., Mahmoudvand, R. (2013). Multivariate singular spectrum analysis: a general view and new vector forecasting approach. *International Journal of Energy and Statistics*. 1 (1), pp. 55 – 83.
- [7] GOLYANDINA, N., Stepanov. D. (2005). SSA-based approaches to analysis and forecast of multidimensional time series. Proceedings of the Fifth Workshop on Simulation. Department of Mathematics, St. Petersburg State University, Russia. pp. 293-298.
- [8] MORETTIN, P. A. e Toloi, C. M. C. (2001). *Análise de Séries Temporais*. Edgard Blücher, São Paulo.
- [9] HAMILTON, J. (1994). *Time Series Analysis*. Princeton University Press.
- [10] BOX, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
- [11] SOUZA, R. C. e Camargo, M. E. (2004). *Análise de Séries Temporais: os Modelos ARIMA*, 2ª.ed., Gráfica e Editora Regional, Rio de Janeiro.