

Previsão do fluxo de turistas para o Brasil utilizando Singular Spectrum Analysis (SSA)

Junior Gabriel Nardino Fumagali¹
Rafael Morais de Souza²

Resumo

Este estudo buscou prever o fluxo de turistas estrangeiros no Brasil, com intuito de contribuir com informações relevantes aos formuladores de políticas públicas. Isso possibilita dinamizar o setor hoteleiro e outros agentes que estão envolvidos com o setor de turismo, promovendo assim um desenvolvimento regional. Foram utilizados dois métodos de previsão: Singular Spectrum Analysis (SSA) e Auto-regressivo integrado com médias móveis (ARIMA), considerando a série mensal do fluxo de turistas estrangeiros entre 2003 a 2019 para o Brasil e dez estados brasileiros, proveniente do Anuário Estatístico de Turismo (Ministério do Turismo). Foram considerados os seguintes horizontes de previsão: 1, 3, 6 e 12 meses. Verificou-se que o fluxo de turistas estrangeiros cresceu no período e apresenta uma sazonalidade bem definida de 12 meses, tendo o pico na alta temporada no final do ano. Considerando os modelos de previsão utilizados, viu-se que o ARIMA foi o melhor método de previsão frente ao SSA. Sendo que o ARIMA foi superior na previsão que SSA em 30 das 44 estimações feitas, ou 68,18%. Apesar do SSA ter tido resultados piores frente ao ARIMA no total, considerando as previsões em relação a 12 meses o SSA venceu o ARIMA.

Palavras-Chave: Singular Spectrum Analysis; Previsão; Turismo.

Abstract

This study sought to predict the flow of foreign tourists in Brazil, with the aim of contributing relevant information to public policy makers. This makes it possible to boost the hotel sector and other agents that are involved with the tourism sector, thus promoting regional development. Two forecasting methods were used: Singular Spectrum Analysis (SSA) and Auto-regressive integrated with moving averages (ARIMA), considering the monthly series of the flow of foreign tourists between 2003 and 2019 to Brazil and ten Brazilian states, from the Statistical Yearbook of Tourism (Ministry of Tourism). The following forecast horizons were considered: 1, 3, 6 and 12 months. It was verified that the flow of foreign tourists grew in the period and presents a well-defined seasonality of 12 months, with a peak in the high season at the end of the year. Considering the forecast models used, it was seen that ARIMA was the best forecasting method compared to SSA. ARIMA was superior in forecasting than SSA in 30 of the 44 estimates made, or 68.18%. Although SSA had worse results compared to ARIMA in total, considering the 12-month forecasts, SSA beat ARIMA.

Keywords: Singular Spectrum Analysis; Forecasting; Tourism.

JEL: M22; Z32

Área 10. Cultura, lazer, turismo e desenvolvimento regional.

¹ Doutorando pelo programa de pos-graduação em Economia Aplicada da Universidade Federal de Juiz de Fora (PPGE/UFJF)

² Professor do Departamento de Economia (DE/UFJF) e do PPGE/UFJF
Agradecimento ao apoio financeiro da CAPES

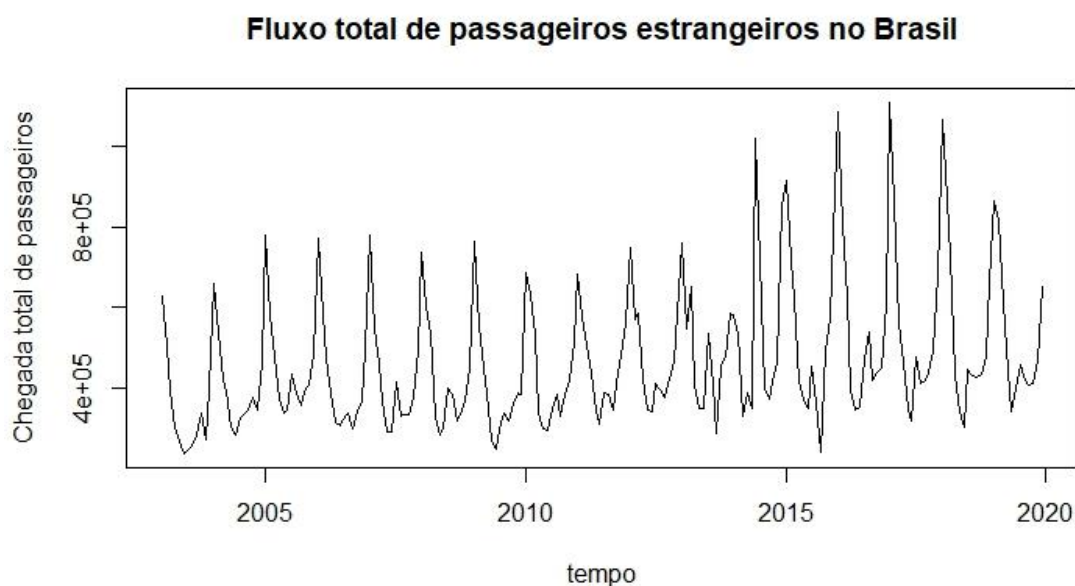
1 Introdução

O turismo é uma fonte de renda relevante para diversos países ao redor do mundo, sobretudo o turismo internacional. Segundo a Organização Mundial do Turismo (OMT, 2019) os dez países que mais receberam turistas estrangeiros em 2019 foram os seguintes: França (89 milhões); Espanha (83 milhões); Estados Unidos (80 milhões), China (63 milhões), Itália (62 milhões), Turquia (46 milhões), México (41 milhões), Alemanha (39 milhões), Tailândia (38 milhões) e Reino Unido (36 milhões). Assim, a receita do turismo impacta positivamente na economia dessas nações, a França em 2019, teve uma receita de 70.776 milhões de dólares com o turismo internacional.

Em relação ao Brasil, o país recebe em média 6 milhões de turistas estrangeiros anualmente, obtendo uma receita de 6.127 milhões de dólares, em média. Uma forma de evidenciar como o turismo internacional é importante para um país é observando que o México que é o sétimo país mais visitado do mundo, tem uma receita de 25.847 milhões de dólares. Esta receita é mais do que 4,21 vezes maior a receita de turismo internacional do Brasil (OMT, 2019). Isso mostra que o fomento do turismo é um fator importante a ser considerado pelo formulador de políticas públicas, principalmente para países em desenvolvimento como Brasil e México.

O fluxo de turistas estrangeiros no Brasil vem crescendo desde os anos 2000 até 2020. Na figura 1, abaixo é possível observar que houve um aumento do fluxo de estrangeiros (período de janeiro de 2003 até dezembro de 2019), sobretudo, no período de férias de final de ano.

Figura 1- Fluxo total de estrangeiros no Brasil.



Fonte: Resultados da pesquisa.

A principal característica deste gráfico é que o fluxo de estrangeiros no Brasil apresenta uma sazonalidade bem definida de 12 meses, ou seja, a cada final de ano o país recebe mais estrangeiros. A principal origem desses estrangeiros, provém de países da América do Sul, sendo a Argentina o principal país de origem desses turistas de acordo com o Anuário Estatístico do Turismo de 2019 (Ministério do Turismo, 2019).

Ainda sobre o ano de 2019, o Brasil recebeu 6,353 milhões de estrangeiros, sendo que o estado de São Paulo recebeu 2,358 milhões desses estrangeiros, ou seja, quase 40% desse total. Os cinco estados brasileiros que mais receberam estrangeiros foram São

Paulo, Rio de Janeiro, Paraná, Rio Grande do Sul e Santa Catarina. O turismo no Brasil também é um gerador de empregos formais, tendo criado 2,104 milhões, sendo o maior em quatro anos. Para o estado o turismo gerou uma arrecadação de 20,775 bilhões de reais, o maior em cinco anos (Ministério do Turismo, 2019).

O turismo pode influenciar no desenvolvimento das regiões brasileiras. Segundo Camargo et al. (2011) o fomento do turismo ecológico no pantanal sul-matogrossense na cidade de Bonito-MS e adjacentes promoveu um desenvolvimento local. Dado que atraiu turistas do país e do exterior a visitar os balneários e hotéis da região, promovendo um crescimento econômico. Aliado a isso, há também o turismo rural que busca mostrar um pouco do ambiente rural, promovendo uma culinária local tradicionais ajudando produtores locais. Logo, há um desenvolvimento regional atrelado ao turismo (Zuin e Zuin, 2007).

Dessa maneira, tendo em vista que o Brasil possui uma sazonalidade bem definida, é possível utilizar modelos de séries temporais a fim de prever a chegada de turistas ao país. Preparando assim o setor hoteleiro e outras infraestruturas turísticas para melhor receber os turistas. Com bons modelos preditivos o formulador de políticas públicas, seja o Ministério do Turismo ou secretárias estaduais, podem ajudar o empresariado local antever o fluxo turístico.

Uma vantagem de se ter modelos de previsão robustos, pois lhe confere uma estimativa de quanto será o fluxo de estrangeiros dentro do país. Isso reduz a dependência de se realizar grandes eventos esportivos internacionais (Exemplo: Copa do Mundo de Futebol 2014 e Olimpíadas do Rio de Janeiro 2016) para impulsionar o turismo internacional. Uma vez que tais eventos, normalmente requerem um estudo de fluxo de estrangeiros para a realização de investimentos turísticos seja estrutural ou de marketing. Atualmente, no Brasil há poucos estudos de previsão da demanda de turistas estrangeiros, Porto et al. (2019) utilizam modelos univariados (ARIMA e Holt-Winters) e redes neurais para prever a demanda por turismo no estado de Mato Grosso do Sul, entre janeiro de 2007 e dezembro de 2017.

No final dos anos 1980 nasceu uma nova abordagem para realizar previsão de series temporais proveniente dos estudos de Broomhead e King (1986) e Broomhead et al. (1987). Tal método é conhecido como Singular Spectral Analysis (SSA) que é uma técnica robusta de previsão. Esse novo método é não paramétrico, permitindo reconstruir a série de fluxo de turistas estrangeiros com base na tendência e na sazonalidade das series. Um diferencial do método é a redução do ruído na série. Esse novo método será comparado com modelos tradicionais de previsão de series temporais como o modelo SARIMA.

Sendo assim em virtude dos números apresentados sobre o turismo no Brasil. É possível prever a chegada de turistas estrangeiros no país no período mensal de 2003 até 2019, excluindo-se o período de 2020 e 2021 em virtude da pandemia da Covid-19. Logo, questiona-se: Qual é o melhor método de previsão entre SSA ou ARIMA para o fluxo de turistas estrangeiros no Brasil nos horizontes de previsão de 1,3,6 e 12 meses? Tal previsão será importante para que os formuladores de políticas públicas prever o fluxo de turistas.

Hassani et al. (2015) utiliza o optimal SSA afim de prever o fluxo turístico mensal internacional para os Estados Unidos. Os resultados encontrados mostram que o SSA teve o menor nível de erro frente a outros modelos de previsão. Em outro estudo Hassani et al. (2009) utilizou o SSA para prever a produção industrial de alguns países europeus. Os resultados encontrados mostraram que o 83,33% de todos os horizontes de previsão estimados. Por fim, Ete, Suhartono e Atok utilizaram o SSA afim de prever o fluxo da entrada de turistas para quatro aeroportos na Indonésia. Os autores consideraram apenas

o horizonte de 12 meses à frente, os resultados encontrados mostraram que o SSA teve um nível de erro menor que o ARIMA.

Assim o objetivo deste trabalho é prever o fluxo turístico internacional no Brasil, afim de apresentar uma abordagem não-paramétrica (SSA) e comparar com o ARIMA. Os resultados a respeito do melhor método podem servir de base para ajudar a desenvolver maneiras de o setor hoteleiro receber melhor os turistas estrangeiros. Além de ajudar a promover o desenvolvimento regional com uma melhor previsão sobre o turismo internacional.

Assim, este artigo está estruturado da seguinte forma. Na seção 2 foi feita uma revisão de literatura sobre o tema de previsão da demanda por turismo. A seção 3 apresenta uma breve explanação dos modelos de previsão utilizados neste artigo. A seção 4 mostra uma discussão sobre os dados de demanda do turismo brasileiro e as chegadas internacionais de turistas ao país. A seção 5 apresenta os resultados encontrados a partir do método SSA e outros métodos de comparação. Por fim, a seção 6 tem-se a conclusão do artigo e suas contribuições para o desenvolvimento do turismo brasileiro.

2 Revisão de literatura

A previsão pela demanda no turismo, de modo geral, é um assunto crescente de estudos na literatura internacional ao longo das últimas duas décadas (Hassani et al. 2017). Onde foram empregadas diversas técnicas paramétricas e não-paramétricas dentro do contexto de séries temporais com intuito de prever a demanda por turismo.

Para a literatura internacional tem-se preocupações distintas entre os autores, contendo assim objetivos diferentes. Alguns autores se preocupam mais em prever a demanda turística e como o setor pode antever isso e se preparar através de investimentos para desenvolver o turismo. Já outros se preocupam majoritariamente com apenas a previsão futura de turistas em seus países, sem haver uma ligação direta de como realizar investimentos para o melhor desenvolvimento do setor. Havendo um foco exclusivamente em comparar diversos métodos diferentes de previsão e encontrar aquele que possui o melhor ajuste.

Assim, alguns autores como Coshall e Charlesworth (2011) e Chu (2008) mostraram como uma previsão da demanda turística com bom ajuste pode contribuir para a formação de investimentos para o setor de turismo. Ambos autores concluíram que sem uma estimativa confiável do fluxo de turistas futuro, em caso de alguma crise ou choque externo será comprometido os investimentos no setor. Portanto, quanto melhor for a estimativa futura do turismo, melhor será a capacidade de realizar investimento para alavancar a demanda turística dos países.

Por outro lado, alguns autores como Álvarez-Díaz e Rósselló-Nadal (2010) focam em encontrar um melhor método que seja obtenha a melhor estimativa de previsão, tendo assim o melhor nível de acurácia. Neste caso, os autores compararam os métodos de previsão ARIMA e Redes Neurais autorregressivas para encontrar a melhor previsão para a demanda turística Britânica nas ilhas Baleares. Outros autores como Assaf, Barros e Gil-Alana (2011) usaram métodos como o processo autorregressivo AR (1) e modelo SARIMA com intuito de capturar a persistência e sazonalidade, respectivamente, para a chegada de turistas na Austrália. Além dos métodos citados anteriormente, autores como Jackman e Greenidge (2010) utilizaram o modelo estrutural de series de tempo para prever a chegada de turistas em Barbados.

Enquanto que Cang (2011) utiliza o modelo SARIMA, amortização exponencial e Redes neurais vetoriais para prever a chegada de turistas no Reino Unido. Os modelos

ARIMA e SARIMA são muito empregados para prever demanda futuras de turistas e seus fluxos, Wan, Wang e Woo (2013) fazem uso do modelo SARIMA para prever a chegada de turistas em Hong Kong. O modelo ARIMA por ter características autorregressivos e de médias móveis de forma integrada é um modelo muito utilizado como baseline para prever o fluxo de turistas futuros.

Alguns autores como Gounopoulos et al. (2012) utilizam o método ARIMA em suas previsões futuras de demanda de turismo, porém fazem uso de um modelo VAR. Os autores incluem choques macroeconômicos no modelo VAR e tentam mensurar como a chegada de turistas na Grécia pode ser afetada por tais choques macroeconômicos. Como uma medida de robustez, eles compararam o modelo VAR com o modelo ARIMA.

Por último, a aplicação de modelos de volatilidade como os modelos ARCH são usados para a previsão de turistas. O uso desses modelos é pelo fato dos modelos ARCH modelarem a volatilidade, sendo a mesma uma maneira muito eficiente de capturar a sazonalidade existente nas séries de demanda turística e produzindo bons resultados de previsão. Kim e Wong (2006) fizeram uso dos modelos EGARCH e TGARCH com objetivo de prever a demanda turística da Coreia do Sul, considerando respostas assimétricas com base em novos choques.

Entretanto, nos últimos anos um novo método vem sendo adotado pelos pesquisadores devido suas características únicas de combinar a tendência e sazonalidade de uma série e assim gerar as melhores previsões com os menores erros. O método conhecido como Singular Spectrum Analysis (SSA) vem sendo utilizado por autores como Beneki, Eeckels e Leon (2012) e Hassani et al. (2015). No primeiro artigo os autores usam SSA para prever o fluxo de renda proveniente de turistas para o Reino Unido, no segundo artigo os autores usam a SSA para prever a demanda futura de turistas para o Estados Unidos. O SSA tem poucas aplicações para a área do turismo, sendo mais abrangente usado em áreas como meteorologia e telecomunicações.

Enquanto que para a literatura nacional, os trabalhos em sua maioria abordam o tema de maneira mais qualitativa, e sem se preocupar com o desenvolvimento do turismo em larga escala e para muitos anos à frente. Muitos autores fizeram estudos de demanda turística para regiões e lugares específicos do país, sendo assim um recorte muito distinto. Esses estudos contribuem pouco para a formulação de políticas públicas em larga escala para o desenvolvimento do país. Sendo que um dos objetivos desse artigo é cobrir essa lacuna da literatura nacional, tão bem explorada em outros países pela literatura internacional.

Um dos autores já citados é Porto et al. (2019) com uso de modelos ARIMA e amortização exponencial (Hot-Winters) para prever a demanda turística no Mato Grosso do Sul. Casanova et al. (2012) também faz uso do modelo ARIMA para prever a demanda turística na cidade de Foz do Iguaçu entre 1999 e 2009. Em geral, há pouquíssimos estudos sobre a demanda turística no Brasil e previsões a respeito.

A fim de contribuir com a literatura nacional, foi proposto neste artigo uma nova abordagem sobre o tema. Utilizou-se os estudos de Hassani et al. (2015) como baseline, assim foi utilizado os métodos de Modelo SARIMA e SSA com objetivo de prever a demanda futura do fluxo de turistas estrangeiros para o Brasil. Tal estudo visa contribuir com a formação de políticas públicas para o setor de turismo, visto que ao prever a chegada de turistas tem-se a possibilidade de desenhar políticas públicas mais eficientes para o desenvolvimento do turismo no país.

3 Métodos de Previsão

3.1 Auto-Regressivo Integrado com Médias Móveis (ARIMA)

O modelo ARIMA foi proposto inicialmente por Box e Jenkins (1970). Essa metodologia consiste em ajustar os modelos auto-regressivos integrados a modelos de médias móveis, ARIMA (p, d, q), a um conjunto de dados. É correto dizer também que o modelo ARIMA é uma evolução do modelo ARMA, que um modelo ARIMA, porém sem integração. Todavia para esse estudo será usado uma evolução do modelo ARIMA, sendo o modelo SARIMA, que é o modelo ARIMA com a presença de sazonalidade (p, d, q)_s, sendo S= 12.

Em relação ao d, refere-se ao número de vezes que é necessário fazer-se-a diferença na série com intuito de tornar ela estacionaria. É possível usar três testes para verificar a estacionariedade da série: o teste de Dick-Fuller (ADF), o teste de Phillips-Perron (PP) e o teste KPSS, todos esses testes verificam a raiz unitária da série e por fim a sua estacionariedade. Além disso para a definição dos valores de p e q, será utilizado o critério de informação Bayesiano (BIC). Foi considerado também, em quais valores de p e q são livres de autocorrelação através do teste de Ljung-Box.

3.2 Singular Spectrum Analysis (SSA)

O SSA (sigla em inglês para *Singular Spectrum Analysis*) é uma técnica não paramétrica para análise e previsão de séries temporais. O SSA decompõe a série temporal original na soma de um pequeno número de componentes independentes e interpretáveis, eles podem ser elencados como: uma tendência de variação lenta, componentes oscilatórios e ruído³. Neste trabalho será apresentada, e trabalhada, a versão univariada do modelo para realizar as previsões das séries em foco.

O SSA tem uma ampla gama de aplicações: desde meteorologia e física até economia e matemática financeira. O SSA foi aplicado pela primeira vez para extrair tendências e componentes harmônicos em séries temporais meteorológicas e geofísicas (Vautard, Yiou, & Ghil, 1992). Nos últimos anos, o SSA continuou em desenvolvimento e sua aplicação se estendeu a diversos problemas práticos, como o estudo de Golyandina e Korobeynikov (2012) que estuda as concentrações de CO₂ do Observatório de Mauna Loa, no Havaí, e contém 468 observações mensais de 1959 a 1997.

O SSA é especialmente útil para analisar e prever séries com componentes sazonais complexos e não estacionários, na qual, ao contrário dos modelos Box&Jenkins da classe ARMA ou ARIMA, aqui não é necessária a escolha de um grau apropriado de diferenciação da série, ou até mesmo a normalidade dos resíduos – como exigem também outras metodologias (HASSANI, 2007). Outro aspecto importante do SSA, que pode se destacar em aplicações econômicas, é que sua aplicação se mostra adequada até mesmo para séries com, relativamente, poucas observações (HASSANI, 2007; HASSANI & ZHIGLJAVSKY, 2009; VAUTARD *et al.*, 1992).

3.2.1 Breve descrição do SSA Básico

De acordo com Golyandina, Nekrutkin e Zhigljavsky (2001), a aplicação do SSA pode ser dividida, basicamente, em dois grandes passos: Decomposição e Reconstrução. Cada um desses dois grandes passos pode ser subdividido em duas categorias cada, sendo essas: incorporação; decomposição do valor singular; agrupamento; aplicação da média

³Os fundamentos teóricos e práticos da técnica SSA podem ser encontrados em Golyandina, Nekrutkin e Zhigljavsky (2001), e uma introdução ao assunto é dada por Elsner e Tsonis (1996).

diagonal⁴. A aplicação desses passos se dá na sequência em que foram apresentados, portanto, o SSA realiza primeiro a decomposição da série e depois a sua reconstrução.

Ainda seguindo Golyandina, Nekrutkin e Zhigljavsky (2001), para entender o primeiro passo da incorporação, primeiro considere uma série temporal real diferente de zero $Y_N = (y_0, \dots, y_{N-1})$ de comprimento suficiente N . O processo de incorporação produz K vetores defasados X_i , sendo: $X_i = (y_{i-1}, \dots, y_{i+L-2})^N$, $1 \leq i \leq K$; $K = N - L + 1$; e L um número inteiro de tamanho $L \leq N/2$ – este chamado de janela de observação. Com isso, é definida a chamada “Matriz de Trajetória” da série ($\mathbf{X} = [X_1: \dots: X_K] = (X_{i,j})_{i,j=1}^{L,K}$), em que as colunas dessa matriz são os vetores defasados que se encontram em um espaço L -dimensional:

$$\begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_{K-1} \\ y_1 & y_2 & y_3 & \cdots & y_K \\ y_2 & y_3 & y_4 & \cdots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{L-1} & y_L & y_{L+1} & \cdots & y_{N-1} \end{bmatrix} \quad (1)$$

Observe que \mathbf{X} é uma Matriz de Hankel, em que, por definição, estas são as matrizes tais que suas (i, j) -ésimas entradas dependem apenas da soma $i + j$. Considere-se, então, \mathbf{X} como dados multivariados com características L e $K = N - L + 1$ observações.

No que tange ao segundo passo, assim como o próprio nome se refere, é realizada a decomposição de valor singular (DVS) da matriz trajetória. Defina uma matriz S tal que $S = \mathbf{X}\mathbf{X}^T$, a decomposição de valor singular fornece os autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L$, e seus respectivos autovetores ortonormais $U_1 \geq U_2 \geq \dots \geq U_L$, logo, a DVS da matriz trajetória pode ser escrita como:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d \quad (2)$$

em que $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$ com $i = 1, \dots, d$; $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$. O subscrito d é o número de autovalores diferentes de zero de $\mathbf{X}\mathbf{X}^T$ e V_i são os componentes principais. Os autores Golyandina, Nekrutkin e Zhigljavsky (2001) chamam o grupo $(\sqrt{\lambda_i}, U_i, V_i)$ de “autotriplas” resultantes do processo DVS.

Passando para a discussão do processo de reconstrução, no terceiro passo o objetivo é agrupar as autotriplas obtidas, e escolhidas⁵, no passo anterior a partir dos m subconjunto disjuntos I_1, \dots, I_m (com $I = \{i_1, \dots, i_p\}$). Tais subconjuntos são obtidos a partir da DVS do anterior, em que, ao expandir a eq. (2), obtém-se:

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_2} \quad (3)$$

Por fim, tem-se no último passo a aplicação da média diagonal que converte cada grupo obtido na decomposição em uma nova série temporal reconstruída. Para entender como o processo funciona, Golyandina, Nekrutkin e Zhigljavsky (2001) apresentam uma matriz F de tamanho $L \times K$ constituída pelos elementos f_{ij} – com $1 \leq i \leq L$ e $1 \leq j \leq K$, com isso, a média diagonal passa a matriz F para as séries g_0, \dots, g_{N-1} utilizando a fórmula:

⁴ Em inglês Golyandina, Nekrutkin e Zhigljavsky (2001) nomeiam esses passos como: *Embedding; Singular value Decomposition; Grouping; Diagonal averaging*.

⁵ O motivo de escolha das autotriplas se deve ao fato de que nem todas fornecidas pelo modelo representarão o sinal da série, algumas podem representar processos ruidosos – processos os quais não devem entrar na reconstrução da série. Para mais detalhamento, ver Hassani (2007).

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} f_{m,k-m+2}^* & \text{para } 0 \leq k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} f_{m,k-m+2}^* & \text{para } L^* - 1 \leq k \leq K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} f_{m,k-m+2}^* & \text{para } K^* \leq k < N. \end{cases} \quad (4)$$

O resultado de (4) permite a reconstrução da série, em que se obtém $\tilde{Y}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$. Logo a série inicial, conforme os autores, é uma soma das m séries decompostas no processo.

3.2.2. Escolha do tamanho da janela (L) e das autotriplas.

O método SSA requer então a seleção de dois parâmetros, o comprimento da janela (L) e o número de matrizes elementares (r), na qual sua escolha depende da estrutura dos dados e da análise que será realizada. Em resumo, o comprimento da janela L é o único parâmetro que deve ser selecionado no estágio de decomposição. Maiores valores de L (em que pode-se assumir $L \leq T/2$) conduzem a uma decomposição com maior quantidade de vetores defasados, sendo que, ao selecionar L , deve-se tentar alcançar “separabilidade” o suficiente dos componentes na etapa da decomposição (GOLYANDINA *et al.*, 2001).

Entretanto, para este trabalho foi utilizada a seguinte abordagem proposta pelos autores como Hassani (2007) e Hassani, Heravi e Zhigljavsky (2009) para determinar o tamanho do L ideal para cada série. Sendo que pode haver algum tipo de periodicidade anual, semestral, trimestral e afins dentro da série. Logo, caso as series tenham periodicidade anual (que é o caso nesse trabalho) deve-se prestar atenção nas frequências $k/12$ ($k = 1$ até 12). Por fim, é importante lembrar que para determinar a frequência presente em cada série é necessário a utilização do períodograma, onde o mesmo informará qual o tipo de frequência presente na série.

A seguinte formulação (chamada de correlação ponderada ou correlação w) é uma medida natural da dependência entre duas séries temporais $Y_T^{(1)}$ e $Y_T^{(2)}$, em que:

$$\rho_{12}^{(w)} = (Y_T^{(1)}, Y_T^{(2)})_w / \left\| Y_T^{(1)} \right\|_w \left\| Y_T^{(2)} \right\|_w \quad (5)$$

Na qual $(Y_T^{(i)}, Y_T^{(j)}) = \sum_{k=1}^T w_k y_k^{(i)} y_k^{(j)}$; $w_k = \min\{k, L, T - k\}$; $\left\| Y_T^{(i)} \right\|_w = \sqrt{(Y_T^{(i)}, Y_T^{(i)})_w}$ com $(i, j = 1, 2)$.

Se dois componentes reconstruídos tiverem uma w -correlação igual a zero, então estes estão bem separados. Grandes valores de w -correlações entre os componentes reconstruídos indicam que os componentes devem ser considerados como um grupo e, possivelmente, correspondem ao mesmo componente na decomposição SSA.

A primeira matriz elementar V_1 com a norma $\sqrt{\lambda_1}$ tem a maior contribuição para a norma de X em (2), e a última matriz elementar V_d com a norma $\sqrt{\lambda_d}$ tem a menor contribuição para a norma de X . O gráfico do logaritmo dos autovalores (exemplo: Figura 1) $\lambda_1 \geq \dots \geq \lambda_d$ fornece uma visão geral dos valores dos autovalores e é essencial para decidir onde truncar a soma de (2) para construir uma boa aproximação da matriz original. Uma sequência lentamente decrescente de autovalores é tipicamente relacionada ao

componente de ruído da série. Valores semelhantes dos autovalores permitem identificar as autotriplas que correspondem à mesma componente harmônica da série.

3.2.3. Previsão SSA

O método de previsão SSA pode ser aplicado a séries temporais que satisfazem, aproximadamente, a fórmula recorrente linear. A classe de séries temporais regidas por fórmulas recorrentes lineares é bastante ampla; inclui harmônicos, polinômios e séries temporais exponenciais.

Para este trabalho, foi usado o algoritmo de previsão Vetor SSA. Assim, faremos uma breve descrição sobre o algoritmo de previsão, baseando-se no trabalho de Golyandina *et al.* (2001)⁶. A escolha pelo Vetor SSA frente ao outro algoritmo de previsão o recorrente SSA, deve-se ao fato de muitas das séries utilizadas neste trabalho terem raiz unitária, logo segundo a própria Golyandina et al. (2001) é melhor usar o Vetor SSA frente ao recorrente SSA. É importante lembrar que o Vetor SSA utiliza a base do recorrente SSA.

Defina a série original $Y_T = (y_1, \dots, y_T)$ e a série reconstruída $\tilde{Y}_T = (\tilde{Y}_1, \dots, \tilde{Y}_T)$. Para um autovetor $U \in \mathbb{R}^L$ denotamos o vetor dos primeiros $L - 1$ componentes do vetor U como $U^\nabla \in \mathbb{R}^{L-1}$. Tomando $v^2 = \pi_1^2 + \dots + \pi_r^2 < 12$, onde o π_i é o último componente do autovetor U_i ($i = 1, \dots, r$), pode-se provar que o última componente y_L de qualquer vetor $Y = (y_1, \dots, y_L)^T$ é uma combinação linear dos primeiros componentes (y_1, \dots, y_{L-1}) . Ou seja, $y_L = a_1 y_{L-1} + \dots + a_{L-1} y_1$, em que o vetor de coeficientes $A = (a_1, \dots, a_{L-1})$ pode ser expresso como :

$$A = \sum_{i=1}^r \pi_i U_i^\nabla / (1 - v^2) \quad (6)$$

Portanto, as previsões são obtidas conforme:

$$\hat{y}_i = \begin{cases} \tilde{y}_i & \text{para } i = 1, \dots, T \\ \sum_{j=1}^{L-1} a_j \hat{y}_{i-j} & \text{para } i = T + 1, \dots, T + h \end{cases} \quad (7)$$

Assim, a matriz que corresponde ao Vetor SSA é a seguinte:

$$\Pi = V^\nabla (V^\nabla)^T + (1 - v^2) A A^T \quad (8)$$

onde: $V^\nabla = [U_1^\nabla, \dots, U_r^\nabla]$. Também é necessário considerar o operador linear:

$$\theta^{(v)}: \mathcal{R}_r \mapsto \mathbb{R}^L \quad (9)$$

onde

$$\theta^{(v)} U = \begin{pmatrix} \Pi U^\nabla \\ A^T U^\nabla \end{pmatrix} \quad (10)$$

Assim, podemos definir o vetor Z_i da seguinte forma:

$$Z_i = \begin{cases} \tilde{X}_i, & \text{onde } i = 1, \dots, K \\ \theta^{(v)} Z_{i-1}, & \text{onde } i = K + 1, \dots, K + h + L - 1 \end{cases} \quad (11)$$

onde, \tilde{X}_i 's são as colunas reconstruídas da matriz trajetória após o agrupamento e a eliminação do ruído. Por fim, para construir a matriz $Z = [Z_1, \dots, Z_{K+h+L-1}]$, além de que ao fazer média diagonal, é possível obter as novas series $y_1, \dots, y_{K+h+L-1}$, onde y_{N+1}, \dots, y_{N+h} sendo na forma de termos em h da previsão do Vetor SSA.

⁶ Outros trabalhos, como Golyandina, Nekrutkin e Zhigljavsky (2001) e Hassani (2007), também fornecem *insights* a respeito do algoritmo de previsão do SSA.

3.3 Testes de acurácia

Com a finalidade de testar a acurácia dos testes de previsão para os modelos propostos (SSA e ARIMA) foram retiradas de todas as series os últimos 24 meses (ano de 2018 e 2019) para ser o *out-of-sample*. A escolha por 24 meses, deve-se que na literatura, Hassani *et al.* (2009), Hassani *et al.* (2015) aponta-se de usar pelo menos os últimos dois anos para a acurácia das séries. Logo, com base nos 24 meses de *out-of-sample*, foi possível ter uma previsão para diferentes horizontes, sendo eles $h=1$; $h=3$; $h=6$ e $h=12$.

Como medida de acerto para as previsões foi utilizado o RMSE (Root mean squared error) que é bastante usado na literatura de previsão (vide, Zhang, Patuwo e Hu, 1998). Com intuito de dar mais espaço, apenas será colocado a razão do RMSE entre o SSA e ARIMA, como pode ser visto abaixo:

$$RRMSE = \frac{(\sum_{i=1}^N (\hat{y}_{T+h,i} - y_{T+h,i})^2)^{\frac{1}{2}}}{(\sum_{i=1}^N (\tilde{y}_{T+h,i} - y_{T+h,i})^2)^{\frac{1}{2}}} \quad (12)$$

onde: $\hat{y}_{T+h,i}$ refere-se ao número (h) de passos à frente para a previsão obtidos por meio da previsão do SSA, enquanto que $\tilde{y}_{T+h,i}$ refere-se ao número (h) de passos à frente na previsão obtidos pela previsão via ARIMA. Logo, se o $RRMSE < 1$, significa que a performance de previsão do SSA é superior à previsão via ARIMA.

4 Dados Utilizados

As séries utilizadas para este estudo foram retiradas do Anuário Estatístico do Turismo, disponível no site do Ministério do Turismo. As séries são mensais tendo início em janeiro de 2003 até dezembro de 2019. Foram utilizadas as series que mostram o número de turistas estrangeiros no Brasil, como também para 10 estados brasileiros: São Paulo, Rio de Janeiro, Paraná, Rio Grande do Sul, Santa Catarina, Bahia, Ceará, Pernambuco, Mato Grosso do Sul e Pará. Em virtude da pandemia da Covid-19, optou-se por excluir os anos de 2020 e 2021 deste trabalho, pelo fato de não ter fluxo turístico no período.

4.1 Modelos utilizados para a análise do poder de previsão do SSA

Os modelos escolhidos como base de comparação frente ao poder preditivo do SSA foram modelos da classe ARIMA⁷, na qual dois motivos levaram a sua escolha. O primeiro motivo foi que muitos trabalhos na literatura que envolvem comparação de métodos preditivos frente ao SSA, usam o ARIMA como base. Em virtude dos modelos ARIMA serem bem quistos pelos pesquisadores. Logo, trabalhos como Hassani *et al.* (2009); Hassani *et al.* (2015); Ete e Suhartono (2020); Hassani (2007); Abreu *et al.* (2018); Menezes *et al.* (2013) e outros. O segundo motivo é a ampla, e recorrente, utilização da metodologia Box&Jenkins na literatura econômica, já possuindo conceito bem pesquisados e estabelecidos – o que fornece uma sólida base de comparação. Dado isso na tabela 1, mostra os parâmetros para os modelos ARIMA para o turismo brasileiro.

⁷ Tendo em vista a vasta aplicação da metodologia Box&Jenkins na literatura de séries temporais, optou-se por omitir a explicação da metodologia dessa classe de modelos no artigo. Para uma introdução, e explicação da metodologia citada, ver Morettin e Tolo (2018).

Já para o SSA, conforme discutido anteriormente, foi necessário definir o tamanho da janela (L) e as autotriplas utilizadas no modelo para reconstruir as séries. A decisão para a escolha do tamanho do parâmetro L, foi com base na periodicidade que a série deve possuir, com base nos trabalhos de Hassani, como visto anteriormente. Além de que se buscou também verificar qual L apresentaria os melhores resultados *out-of-sample* frente ao modelo ARIMA. Dado isso, na tabela 2 observa-se a seleção dos parâmetros L. É importante lembrar que a escolha dos L também representa o primeiro estágio do SSA, que é a decomposição. Logo, ao inserir os L's desejados é possível fazer o primeiro estágio do SSA.

TABELA 1- Parâmetros para a previsão do modelo ARIMA, para o Brasil e os estados brasileiros.

Séries	Modelos ARIMA
Brasil	(2,1,2)(0,1,1) ¹²
São Paulo	(1,0,0)(1,0,0) ¹²
Rio de Janeiro	(1,0,0)(1,1,1) ¹²
Paraná	(1,1,2)(1,1,4) ¹²
Rio Grande do Sul	(2,1,4)(1,1,2) ¹²
Santa Catarina	(0,1,1)(0,1,1) ¹²
Bahia	(0,1,1)(0,1,1) ¹²
Ceará	(0,1,2)(0,1,1) ¹²
Pernambuco	(0,1,3)(0,1,4) ¹²
Mato Grosso do Sul	(1,1,2)(1,1,3) ¹²
Pará	(1,0,1)(1,0,0) ¹²

Fonte: Resultados da pesquisa.

Tabela 2 – Tamanho da Janela (L) e o número de autotriplas.

Séries	Janela de Comprimento(L)	Número de Autotriplas
Brasil	84	50
São Paulo	36	36
Rio de Janeiro	84	50
Paraná	24	24
Rio Grande do Sul	84	50
Santa Catarina	72	50
Bahia	84	50
Ceará	36	36
Pernambuco	24	24
Mato Grosso do Sul	72	50
Pará	90	50

Fonte: Resultados da pesquisa.

5 Resultados da Previsão

O primeiro resultado foi utilizando sempre apenas tendência e sazonalidade, conforme a ideia central do SSA. A tabela 3, mostra a separação do conjunto de autotriplas. A separação das autotriplas, conhecidas como (r) na literatura de SSA é o segundo estágio do SSA, que é a reconstrução das séries. Sendo quando se agrupam as series em relação as autotriplas que possuem tendência e/ou sazonalidade, logo é possível reconstruir a série livre de ruído.

Tabela 3- Representação das autotriplas obtidas pelo SSA.

Séries	Autotriplas		
	Tendência	Sazonalidade	Ruído
Brasil	1	2-12	13-50
São Paulo	1	2-17	18-36
Rio de Janeiro	1	.2-22	23-50
Paraná	1	2-19	20-24
Rio Grande do Sul	1	2-11	12-50
Santa Catarina	1	2-12	13-50
Bahia	1,6	2-5,7-8	9-50
Ceara	1	2-10	11-36
Pernambuco	1	2-14	15-24
Mato Grosso do Sul	1	2-15	16-50
Pará	1-2	3-19	20-50

Fonte: Resultados da pesquisa.

Assim, foi possível obter os resultados onde comparou-se a performance tanto no *in-sample* quanto no *out-of-sample* entre SSA e ARIMA. No apêndice A, será mostrado em mais detalhes sobre os parâmetros escolhidos (tamanho do L e autotriplas) para o SSA. Na tabela 4, observa-se os resultados para as series do turismo brasileiro.

Tabela 4- Estatísticas descritivas dos erros *in-sample* e *out-of-sample* para as series de fluxo de turistas estrangeiros no Brasil.

Séries	In-Sample: RMSE		In-Sample: RRMSE	h	Out-of-Sample: RRMSE
	SSA	ARIMA	SSA		SSA
			ARIMA		ARIMA
Brasil	63508,49	76289,62	0,8324	1	1,6761 ^a
				3	1,6085 ^{ab}
				6	1,6795 ^{ab}
				12	1,6272 ^{ab}
São Paulo	14672,09	29667,81	0,4945	1	0,9525
				3	0,8601 ^b
				6	0,8071 ^{ab}
				12	0,6764 ^{ab}
Rio de Janeiro	10479,74	19875,61	0,5272	1	2,4254 ^a
				3	1,7467 ^{ab}
				6	1,6859 ^{ab}
				12	1,5497 ^{ab}

Paraná	3775,829	9520,374	0,3966	1	1,1898
				3	1,1739
				6	0,9915
				12	0,9079
Rio Grande do Sul	30095,76	30400,18	0,9899	1	2,3129 ^a
				3	2,4347 ^{ab}
				6	2,4037 ^{ab}
				12	2,2617 ^{ab}
Santa Catarina	3645,944	4879,842	0,7471	1	1,4288
				3	1,3848
				6	1,2017
				12	0,7182 ^{ab}
Bahia	2635,864	3193,959	0,8252	1	1,057
				3	1,0157
				6	0,9577
				12	0,9463
Ceará	1634,181	1756,44	0,9303	1	1,6595 ^a
				3	1,5984 ^{ab}
				6	1,6594 ^{ab}
				12	1,3105 ^{ab}
Pernambuco	765,9233	1866,037	0,4104	1	0,9721
				3	1,0700
				6	1,0531
				12	1,1697 ^{ab}
Mato Grosso do Sul	1513,093	2167,113	0,6982	1	1,3125
				3	1,2827
				6	1,1916
				12	1,1608 ^{ab}
Pará	448,8929	707,7081	0,6342	1	0,8871
				3	0,8482
				6	0,8268 ^b
				12	0,8647 ^b
Pontuação			11	1	3
				3	2
				6	4
				12	5

Fonte: Resultados da pesquisa. a indica a significância do teste de Diebold-Mariano para 10% ou menos. b indica a significância do *Encompassing test* para 10% ou menos.

Na tabela 4, mostra que o SSA foi superior para todas as 11 séries do turismo brasileiro nos erros *in-sample*. Todavia, para os erros *out-of-sample* nota-se pela última linha (pontuação) que em apenas três séries o SSA foi superior ao ARIMA para h=1, enquanto que para h=3 apenas duas séries o SSA foi superior ao ARIMA, já para h=6 o SSA foi superior em quatro séries, e finalmente para h=12 o SSA teve seu melhor

resultado com cinco séries, tendo uma performance superior ao ARIMA, ou seja, quase 50% das séries.

Tais resultados mostram que na previsão out-of-sample, o SSA não foi superior ao ARIMA, tendo apenas 31,81% dos horizontes de previsão no geral superior ao ARIMA. Isso contrasta com a maioria da literatura sobre SSA onde geralmente o SSA é superior aos demais métodos de previsão (HASSANI et al. 2009). Até mesmo em relação a literatura sobre turismo onde é aplicado o SSA, os resultados do SSA são superiores (ETE e SUHARTONO, 2020; HASSANI et al. 2015). Apesar disso, há trabalhos relacionados ao turismo onde o SSA realmente não foi superior a outros métodos de previsão, visto em Hassani et al. (2017) onde a previsão para as chegadas de turistas em países europeus mostrou que não houve um método que sobressaísse frente a outro método. Logo para alguns países (exemplo, a Espanha) o ARIMA e o TBATS foram superiores a outros métodos de previsão, especialmente a respeito do SSA.

5.1 Estatísticas de Diebold-Mariano

Como medida de robustez, com intuito de confirmar as previsões estatisticamente, foi utilizado o teste de Diebold-Mariano com intuito de comparar os erros de previsão. Podendo assim, atestar que estatisticamente uma previsão é melhor que a outra. A letra *a* significa que os resultados obtiveram um nível de significância igual a 10% ou menos (DIEBOLD e MARIANO, 1995).

Em virtude do modelo ARIMA ter sido melhor em oito vezes das onze estimações para $h=1$, nota-se que o ARIMA tem uma performance melhor que o SSA, já no âmbito estatístico vemos que o ARIMA foi melhor quatro vezes frente ao SSA. Em relação ao $h=3$, o ARIMA foi estatisticamente melhor que o SSA, quatro vezes também. Já para o $h=6$, o ARIMA foi melhor em quatro estimações. Por fim, em $h=12$, o ARIMA foi estatisticamente melhor que o SSA em seis estimações.

Além do teste de Diebold-Mariano, foi considerado também o encompassing test proposto por Harvey, Leybourne e Newbold (1998) também como uma medida de robustez. Considerando esse teste, notamos que o ARIMA foi melhor que o SSA em 14 das 44 estimações totais, considerando todos os horizontes de previsão. Algo como 31,8% das estimações totais, mais que o dobro das estimações onde o SSA ganhou que foram 5 de 44, ou 11,36%.

6 Conclusões

Para este trabalho, houve uma comparação entre dois modelos de previsão, SSA e ARIMA, com objetivo de prever mensalmente o fluxo de turistas estrangeiros que chegam no Brasil e para dez estados brasileiros. O resultado contrariou a literatura econômica que apresentava que o SSA seria superior a outros métodos de previsão, dentre eles o ARIMA. O que foi observado é que o SSA não foi melhor que o ARIMA em nenhum dos horizontes de previsão ($h=1$; $h=3$; $h=6$ e $h=12$), de acordo com o critério do RMSE. Apesar disso, para o horizonte de previsão 12 meses à frente o SSA teve o melhor desempenho, quase superando a performance do modelo ARIMA. Portanto, o SSA não foi um modelo de previsão confiável como se esperava para prever o fluxo de turistas estrangeiros vindos ao Brasil.

Apesar do método SSA não ter sido o mais eficaz como se esperava, segundo os resultados vistos em Hassani et al. (2009); fomentar o turismo é importante para a promoção do desenvolvimento regional. Dado que a atração de novos turistas provenientes de outras partes do mundo irá fomentar desenvolvimento local. Podendo

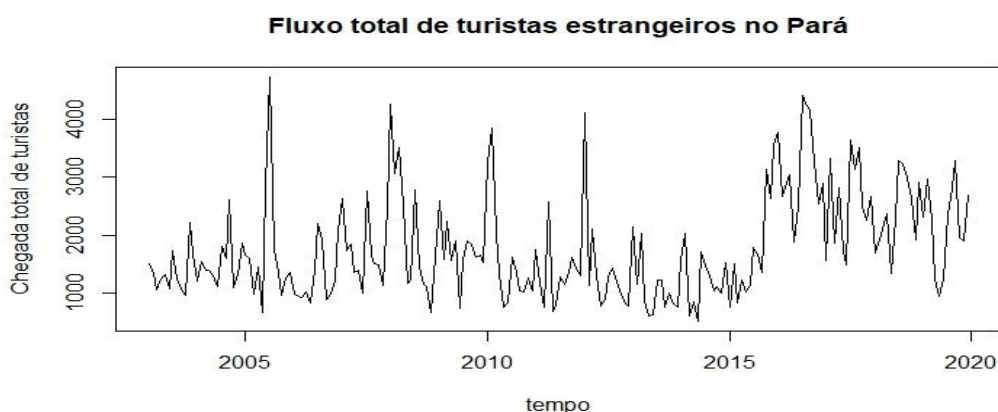
assim aumentar o setor hoteleiro, através da criação de novos postos de trabalho. A chegada de turistas ao Brasil também será importante para impulsionar o comércio regional por via de produtos locais, gerando uma riqueza regional como também mostrando a cultura regional brasileira.

Ademais, é necessário para trabalhos futuros incluir outros modelos preditivos para que se tenha mais comparativos de previsão. Considerar também, o período pós pandemia (2020 e 2021) e ver se os resultados preditivos para os métodos, especialmente o SSA tem alguma alteração significativa.

APÊNDICE A. Aplicação do SSA para o fluxo de turistas estrangeiros no estado do Pará, Brasil.

Foi usado a série do fluxo de turistas estrangeiros no estado do Pará, com intuito de ilustrar um exemplo de aplicação do SSA, selecionando seus parâmetros e a reconstrução da série original. Na figura 2, vemos apresentamos o gráfico da série, onde percebe-se que não apresenta uma tendência, e a alguns sinais de uma possível sazonalidade.

Figura 2- Fluxo de turistas estrangeiros no estado do Pará

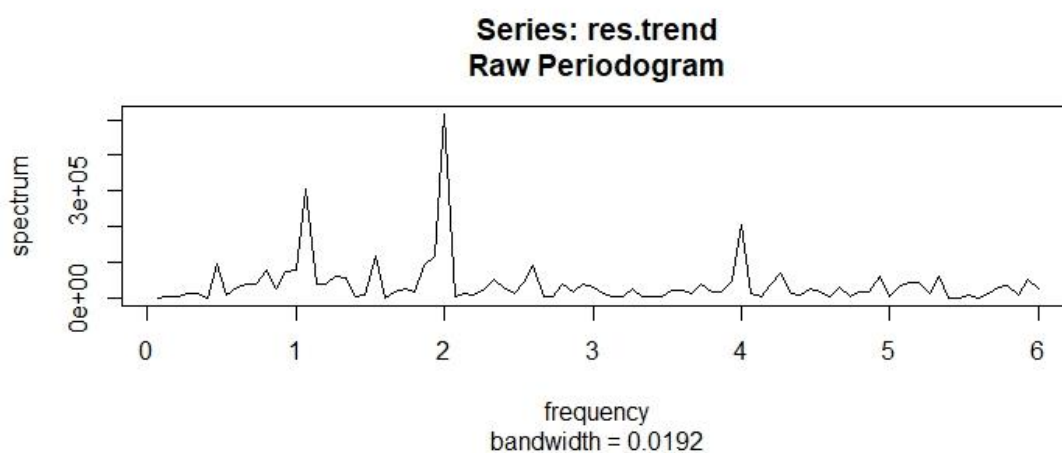


Fonte: Resultados da pesquisa.

A.1. Seleção da janela de comprimento (L)

A janela de comprimento, ou *window length*, é o único parâmetro no estágio da decomposição. Ao analisar a série é possível perceber que talvez haja um componente de periodicidade na série. Logo, para obter a melhor separabilidade da série é necessário que o L seja proporcional a essa periodicidade. Por exemplo, assumimos que haja uma periodicidade anual na série é interessante prestar atenção na frequência da série, $k/12$ ($k=1,2, \dots,12$). No caso do estado do Pará, de acordo com o periodograma (figura 3), notamos que a frequência é igual a $2/12$, logo repete-se a cada seis meses, sendo o período mais forte de repetição. Assim um L que seria razoável é $L=90$, que é um múltiplo de 6.

Figura 3- Periodograma da série para o estado do Pará.

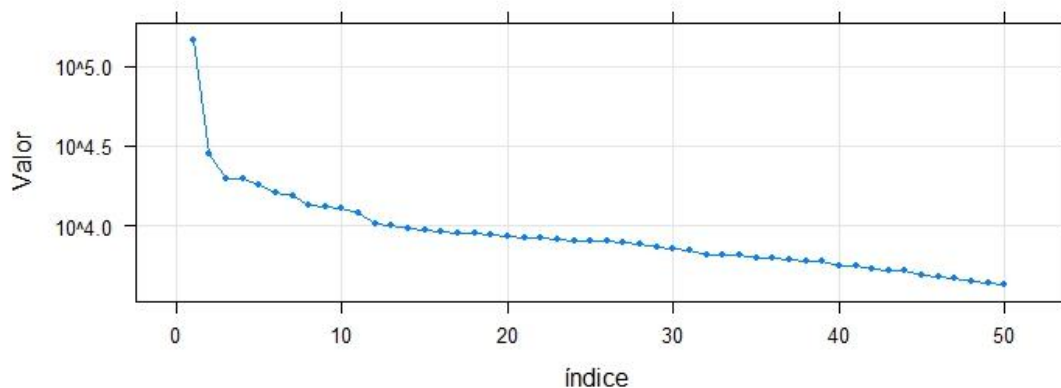


Fonte: Resultados da pesquisa.

B.2. Seleção das autotriplas utilizadas.

Primeiramente, será apresentado o gráfico do log dos autovalores que indica possíveis tendências e/ou sazonalidades presentes na série. Assim, como escolhemos um $L=90$, o gráfico do log dos autovalores irá apresentar até a autotripla número 50.

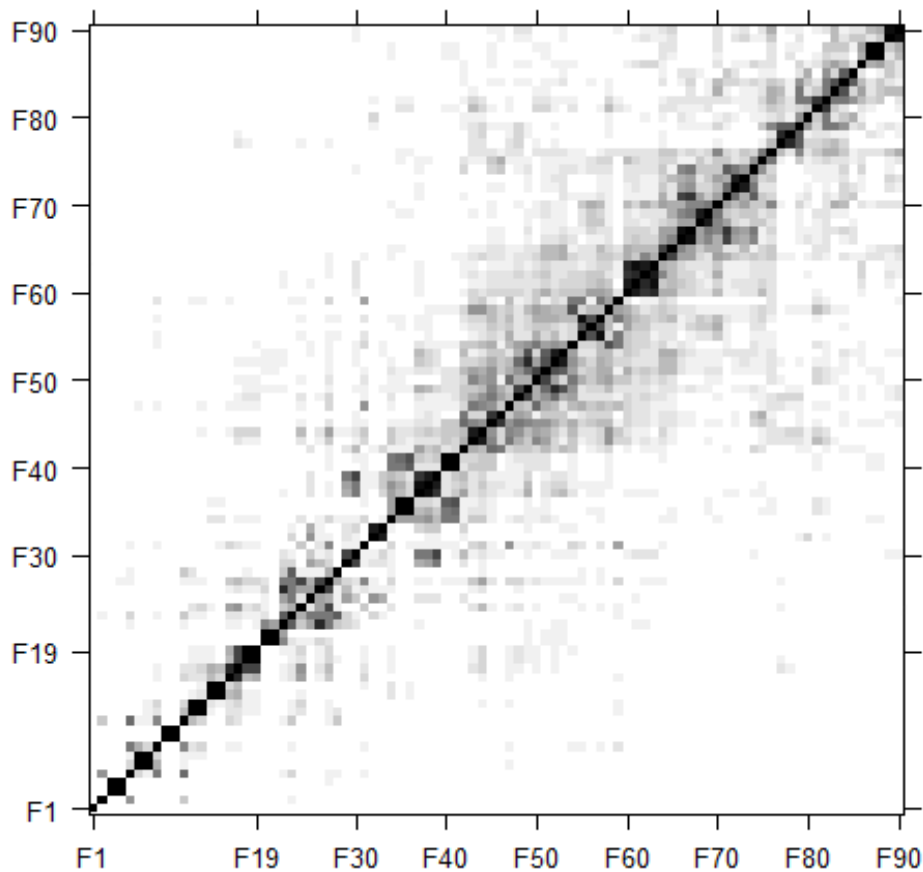
Figura 4- Log dos autovalores para o estado do Pará.



Fonte: Resultados da pesquisa.

Nota-se que até a autotripla 50, há algum poder de explicação e que as duas primeiras autotriplas indicam uma forte presença de tendência. Enquanto a partir da autotripla 3 em diante, vemos que há uma sazonalidade. Porém como o gráfico mostra que até a autotripla 50 há um poder de explicação, visto que o valor ainda não é abaixo $10^{2.0}$. É necessário utilizar o gráfico da matriz de correlação w afim de verificar o que é de fato tendência, sazonalidade e ruído. A figura 5, apresenta o gráfico da matriz de correlação w .

Figura 5- Matriz de correlação w para o estado do Pará.



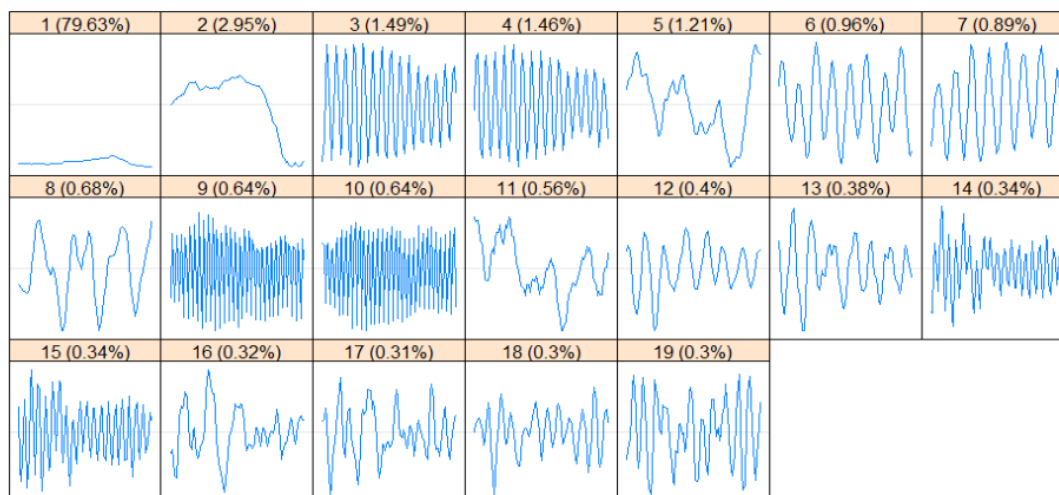
Fonte: Resultados da pesquisa.

Notamos que até a autotripla 19, há algum grau de explicação e após isso a série apresenta correlações entre autotriplas de forma espalhadas o que configura ruído. Assim pelo gráfico, notamos que as autotriplas 1 e 2, estão bem separadas (quanto mais preto estiver os quadradinhos F1-F1, por exemplo, e estiver mais branco a correlação F1-F2, melhor será a separabilidade). Já para a autotripla 3 até a 19, vemos uma presença forte de sazonalidade em algumas dessas autotriplas. Logo é correto supor que 1 e 2 são tendência e da 3 até 19 temos sazonalidade.

Além disso, será apresentado o gráfico dos autovetores até a autotripla 19, para mostrar o quanto cada autotripla participa na composição da série principal. A figura 6, mostra os componentes.

Figura 6– Gráfico dos autovetores para as 19 autotriplas.

Eigenvectors



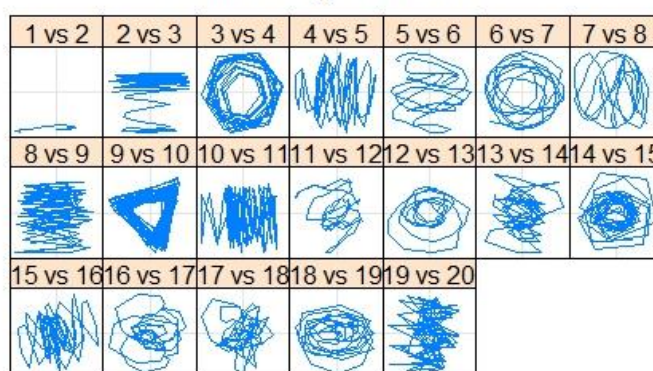
Fonte: Resultados da pesquisa.

Notamos que até o autovetor 19 tem uma contribuição de 0,3% na série, em virtude do autovetor 20 já conter muito a presença de ruído, temos que desconsiderar o restante dos autovetores mesmo que eles expliquem uma porcentagem ínfima da série. Outros trabalhos como Hassani et al. (2009) é possível separar muito bem a sazonalidade e tendência, explicando assim 100% da série nos primeiros autovetores. Entretanto isso não impede que a série não possa ser explicada com uma porcentagem menor, como exemplo em Hassani (2007), onde o último autovetor tem um poder de explicação de 0,010%, logo é possível supor que o restante dos autovetores tenham um poder de explicação mesmo que seja ínfimo.

Por fim, será apresentado também o gráfico com os pares de autovetores que permitirá identificar se há algum tipo de sazonalidade na série e de quantos meses é essa sazonalidade. O gráfico está presente na figura 7, a seguir:

Figura 7- Pares de autovetores para o estado do Pará

Pairs of eigenvectors



Fonte: Resultados da pesquisa.

Vemos pelos pares de autovetores que as autotriplas 1 e 2 são tendência, enquanto para autotriplas 3-4 vemos uma sazonalidade de 6 meses e para o par 9-10 temos uma sazonalidade de 3 meses.

Referências bibliográficas

- Abreu, R. J.; Souza, R. M.; Oliveira, J. G. Applying singular spectrum analysis and ARIMA-GARCH for forecasting EUR/USD exchange rate. *Revista de Administração Mackenzie*, V.20, N.4, 2019.
- Álvarez-Díaz, M.; Rosselló-Nadal, J. Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics*, V.16, N.1, 2010.
- Anuário Estatístico do Turismo. Ministério do Turismo, Brasília, 2019. Disponível em: Página Inicial — Ministério do Turismo (www.gov.br). Acessado em: 10/10/2022.
- Assaf, A. G.; Barros, C. P.; Gil-Alana, L. A. Persistence in the Short- and Long-Term Tourist Arrivals to Australia. *Journal of Travel Research*, V.50, N.2, 2011.
- Beneki, C.; Eeckels, B.; Leon, C. Signal extraction and forecasting of the UK tourism income time series: a Singular Spectrum Analysis approach. *Journal of Forecasting*, V.31, N.5, p.391-400, 2012.
- Box, G.; Jenkins, G. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- Broomhead, D. S.; King, G. P. On the qualitative analysis of experimental dynamical systems. In S. Sarkar, editor, *Nonlinear Phenomena and Chaos*, Adam Hilger, Bristol, p. 113–144, 1986.
- Broomhead, D. S.; Jones, R.; King, G.P.; Pike, E. R. Singular system analysis with application to dynamical systems. In E.R. Pike and L.A. Lugaito, editors, *Chaos, Noise and Fractals*, IOP Publishing, Bristol, p.15-27, 1987.
- Camargo *et al.*; Análise da sustentabilidade do Turismo ecológico no município de Bonito, Mato Grosso do Sul na promoção do desenvolvimento regional. *Sociedade & Natureza*, V.23, N.1, Uberlândia, p.65-75, 2011.
- Cang, S. A non-linear tourism demand forecast combination model. *Tourism Economics*, V. 17, N.1, p. 5-10, 2011.
- Chu, F. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management*, V.29, 2008.
- Coshall, J. T.; Charlesworth, R. A management orientated approach to combination forecasting of tourism demand. *Tourism Management*, V.32, 2011.
- Diebold, F. X.; Mariano, R. S. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, V.20, N.1, p.134-144, 1995.
- Elsner, J. B., Tsonis, A. A. **Singular Spectral Analysis: A new tool in time series analysis**. 1 ed. Nova Iorque: Plenum Press, 1996.
- Ete, A. A.; Suhartono, R. M. A. SSA and ARIMA for Forecasting Number of Foreign Visitor Arrivals to Indonesia. *Inferensi*, V.3, N.1, 2020.
- Golyandina, N.; Korobeynikov, A. *Basic Singular Spectrum Analysis and forecasting with R*, Elsevier, 2012.
- Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. *Analysis of Time Series Structure: SSA and Related Techniques*. In: GOLYANDINA N.; NEKRUTKIN V.; ZHIGLJAVSKY A. **Basic SSA**. 1. ed. Boca Raton: Chapman & Hall/CRC, cap. 1, p. 28-105, 2001.
- Gounopoulos, D.; Petmezas, D.; Santamaria, D. Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists origin. *Annals of Tourism Research*, V.39, N.2, p. 641-666, 2012.
- Harvey, D. I.; Leybourne, S. J.; Newbold, P. Tests for forecast encompassing. *Journal of Business and Economic Statistics*, V.16, p.254–259, 1998.
- Hassani, H. Singular spectrum analysis: methodology and comparison. **Journal of Data Science**, v.5, n.2, p. 239-257, 2007.

Hassani, H.; Heravi, S.; Zhigljavsky, A. Forecasting European industrial production with singular spectrum analysis. *International Journal of Forecasting*, V.25, 2009.

Hassani, H.; Webster, A.; Silva, E. S.; Heravi, S. Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis. *Tourism Management*, V. 46, 2015.

Houssein, H.; Silva, E. S.; Antonakakis, N.; Filis, G.; Gupta, R. Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, V.63, 2017.

Jackman, M., & Greenidge, K. Modelling and forecasting tourist flows to Barbados using structural time series models. *Tourism and Hospitality Research*, V.10. N.1, 2010.

Kim, S. S.; Wong, K. K. Effects of news shock on inbound tourist demand volatility in Korea. *Journal of Travel Research*, V.44, N.4, p.457-466, 2006

Menezes, M.; Teixeira Júnior, L.; Souza, R.; Cassiano, K.; Pessanha, J.; Souza, R. HYDROELECTRIC ENERGY FORECAST. *International Journal of Energy and Statistics*, 2013.

Porto, B. M.; Philippi, D. A.; Leite, V. A. W. Previsão dos modelos univariados e rede neural da demanda turística do estado de Mato Grosso do Sul. *Caderno Virtual de Turismo*, V.19, N.3, 2019.

Vautard, R.; Yiou, P.; Ghil, M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D*, v.58, p. 95-126, 1992.

Wan, S. K. ; Wang, S. H.; Woo, C. K. Aggregate vs. disaggregate forecast: case of Hong Kong. *Annals of Tourism Research*, V.42, p.434-438, 2013.

World Tourism Organization. *World Tourism Barometer*. V. 17, Issue 4, novembro de 2019.

Zuin, L.F.S.; Zuin, P. B. PRODUÇÃO DE ALIMENTOS TRADICIONAIS Contribuindo para o desenvolvimento local/regional e dos pequenos produtores rurais. *Revista Brasileira de Gestão e Desenvolvimento Regional*, V.4, N.1, Taubaté-SP, Brasil, p.109-127, 2008.