




## PREVISÃO DO RISCO DO IBOVESPA POR ABORDAGEM HÍBRIDA AESM-MLCP

### PREDICTING IBOVESPA RISK USING A HYBRID AESM-MLCP APPROACH

### PREDICCIÓN DEL RIESGO DE IBOVESPA MEDIANTE UN ENFOQUE HÍBRIDO AESM-MLCP

 <https://doi.org/10.56238/levv17n59-032>

Data de submissão: 15/03/2026

Data de publicação: 15/04/2026

**Carlos Alberto Orge Pinheiro**

Doutor em Modelagem Computacional e Tecnologia Industrial

Instituição: Universidade do Estado da Bahia (UNEB)

E-mail: [capinheiro@uneb.br](mailto:capinheiro@uneb.br)

Lattes: <http://lattes.cnpq.br/0257392772572219>

#### RESUMO

Este estudo propõe e avalia um modelo híbrido para a previsão do risco diário do IBOVESPA, combinando a Análise Espectral Singular Multivariada – d AESM e redes neurais do tipo Memória de Longo e Curto Prazo – d MLCP. A pesquisa caracteriza-se como aplicada, quantitativa, explicativa, documental, computacional e empírico-analítica, utilizando dados diários dos índices IBOVESPA, IBrX-100 e IBrX-50, obtidos no portal ADVFN, no período de 15 de fevereiro de 2021 a 27 de março de 2026. A variável de risco foi construída a partir do logaritmo natural da razão entre a máxima e a mínima de cada pregão. Metodologicamente, foi estruturado em duas etapas: na primeira, realizou-se uma otimização automática dos parâmetros da AESM e da MLCP; na segunda, executou-se o treinamento final com a melhor configuração identificada. A AESM foi empregada como técnica de filtragem estrutural e extração de componentes comuns entre as séries, enquanto a MLCP foi utilizada como modelo preditivo para captar dependências temporais não lineares. Os resultados evidenciaram desempenho superior da abordagem híbrida em relação às estratégias preliminares, especialmente quanto à aderência da série prevista à série observada. A melhor configuração encontrada na etapa de treinamento foi  $r = 5$ ,  $L = 10$  e  $lookback = 30$ , alcançando correlação de 0,860537 no conjunto de teste, além de erro absoluto médio de 0,051964 e raiz do erro quadrático médio de 0,066677, calculados sobre a variável normalizada. Na etapa de teste a correlação foi de 0,90 com menores medidas de erro. Conclui-se que a integração entre AESM e MLCP constitui uma alternativa promissora para a previsão do risco em séries financeiras, embora persistam limitações na captura completa de eventos extremos e na ausência de validação temporal robusta.

**Palavras-chave:** Previsão de Risco. Análise Espectral Singular Multivariada. Memória de Longo e Curto Prazo. Séries Temporais Financeiras. Aprendizado.

#### ABSTRACT

This study proposes and evaluates a hybrid model for forecasting the daily risk of the IBOVESPA, combining Multivariate Singular Spectrum Analysis (MSSA) and Long Short-Term Memory (LSTM) neural networks. The research is characterized as applied, quantitative, explanatory, documental, computational, and empirical-analytical, using daily data from the IBOVESPA, IBrX-100, and IBrX-50 indices, obtained from the ADVFN portal, covering the period from February 15, 2021, to March

27, 2026. The risk variable was constructed using the natural logarithm of the ratio between the daily high and low prices. Methodologically, the study was structured into two stages: in the first, an automatic optimization of MSSA and LSTM parameters was performed; in the second, the final training was conducted using the best configuration identified. MSSA was employed as a structural filtering technique and for extracting common components across the series, while LSTM was used as a predictive model to capture nonlinear temporal dependencies. The results demonstrated superior performance of the hybrid approach compared to preliminary strategies, particularly in terms of adherence between the predicted and observed series. The best configuration identified during training was  $r=5$ ,  $L=10$  and  $\text{lookback} = 30$ , achieving a correlation of 0.860537 in the test set, along with a mean absolute error of 0.051964 and a root mean squared error of 0.066677, calculated on the normalized variable. In the testing phase, correlation reached 0.90 with lower error measures. It is concluded that the integration of MSSA and LSTM constitutes a promising alternative for risk forecasting in financial time series, although limitations remain in fully capturing extreme events and in the absence of robust temporal validation.

**Keywords:** Risk Forecasting. Multivariate Singular Spectrum Analysis. Long Short-Term Memory. Financial Time Series. Machine Learning.

## RESUMEN

Este estudio propone y evalúa un modelo híbrido para la predicción del riesgo diario del IBOVESPA, combinando el Análisis Espectral Singular Multivariante (AESM) y redes neuronales de Memoria a Largo y Corto Plazo (LSTM). La investigación se caracteriza como aplicada, cuantitativa, explicativa, documental, computacional y empírico-analítica, utilizando datos diarios de los índices IBOVESPA, IBrX-100 e IBrX-50, obtenidos del portal ADVFN, en el período comprendido entre el 15 de febrero de 2021 y el 27 de marzo de 2026. La variable de riesgo fue construida a partir del logaritmo natural de la razón entre el máximo y el mínimo de cada jornada. Metodológicamente, el estudio se estructuró en dos etapas: en la primera, se realizó una optimización automática de los parámetros de AESM y LSTM; en la segunda, se llevó a cabo el entrenamiento final con la mejor configuración identificada. El AESM se empleó como técnica de filtrado estructural y extracción de componentes comunes entre las series, mientras que el LSTM se utilizó como modelo predictivo para captar dependencias temporales no lineales. Los resultados evidenciaron un rendimiento superior del enfoque híbrido en comparación con las estrategias preliminares, especialmente en términos de adherencia entre la serie prevista y la observada. La mejor configuración encontrada en la etapa de entrenamiento fue  $r=5$ ,  $L=10$  Y  $\text{lookback} = 30$ , alcanzando una correlación de 0.860537 en el conjunto de prueba, además de un error absoluto medio de 0.051964 y una raíz del error cuadrático medio de 0.066677, calculados sobre la variable normalizada. En la etapa de prueba, la correlación alcanzó 0.90 con menores medidas de error. Se concluye que la integración entre AESM y LSTM constituye una alternativa prometedora para la predicción del riesgo en series temporales financieras, aunque persisten limitaciones en la captura completa de eventos extremos y en la ausencia de una validación temporal robusta.

**Palabras clave:** Predicción de Riesgo. Análisis Espectral Singular Multivariante. Memoria a Largo y Corto Plazo. Series Temporales Financieras. Aprendizaje Automático.

## 1 INTRODUÇÃO

A previsão de risco e volatilidade em mercados financeiros permanece como um desafio da literatura econômica e quantitativa, em razão da natureza não linear, ruidosa e dinâmica das séries temporais financeiras. Em especial, índices como o Índice da Bolsa de Valores de São Paulo – IBOVESPA apresentam comportamentos caracterizados por mudanças abruptas, dependência temporal complexa e forte influência de fatores exógenos, o que dificulta a modelagem por métodos tradicionais (CONT, 2001; POON; GRANGER, 2003; MACIEL; BALLINI, 2017)).

Modelos clássicos, baseados em estruturas lineares ou em pressupostos de estacionariedade, frequentemente apresentam limitações na captura de padrões não lineares e de eventos extremos. Nesse contexto, abordagens baseadas em decomposição de séries, como a Análise Espectral Singular – AES e sua extensão multivariada ou Análise Espectral Singular Multivariada – AESM, têm sido utilizadas como ferramentas de filtragem e extração de estrutura, permitindo separar componentes de tendência e ruído (Hassani et al., 2020). Contudo, apesar de sua eficácia na redução de ruído, tais métodos tendem a produzir representações suavizadas, o que limita sua capacidade preditiva em cenários com variabilidade.

Paralelamente, modelos baseados em aprendizado, como as redes neurais do tipo Memória de Longo e Curto Prazo – MLCP, têm demonstrado capacidade superior na modelagem de dependências temporais complexas e não lineares, sendo amplamente aplicados na previsão de séries financeiras (Fischer & Krauss, 2021). Ainda assim, tais modelos podem ser sensíveis ao ruído presente nos dados, o que comprometer sua estabilidade e desempenho.

Diante desse cenário, emerge a possibilidade de integração entre técnicas de decomposição estrutural e modelos de aprendizado, combinando as vantagens das abordagens. Pesquisas recentes indicam que estratégias híbridas, que utilizam métodos de filtragem como etapa prévia ao aprendizado, podem melhorar a qualidade da informação de entrada e, conseqüentemente, o desempenho preditivo (Zhang et al., 2022; Qin et al., 2023).

Neste contexto, o artigo tem como objetivo geral propor e avaliar um modelo híbrido baseado na combinação entre AESM e MLCP para a previsão do risco diário do IBOVESPA. Como objetivo específico, busca-se investigar o impacto da utilização de componentes estruturais extraídas via AESM e de seus resíduos na capacidade preditiva de modelos MLCP, bem como identificar a melhor configuração de parâmetros para essa integração.

A hipótese subjacente é que a utilização do AESM como etapa de filtragem estrutural, aliada à capacidade do MLCP de capturar dependências temporais não lineares, resulta em melhoria significativa na previsão do risco em comparação a abordagens isoladas.

Para atingir tais objetivos, o artigo está estruturado da seguinte forma: a seção seguinte apresenta o referencial teórico, abordando a previsão de risco em séries financeiras, os fundamentos

da AESM e das redes MLCP; na sequência, a seção de metodologia descreve os dados, a construção da variável de risco; a seção de resultados apresenta as análises empíricas e a avaliação de desempenho da combinação proposta; por fim, a seção de considerações finais sintetiza os principais achados, limitações e sugestões para pesquisas futuras.

## 2 REFERENCIAL TEÓRICO

A previsão de risco e volatilidade em mercados financeiros constitui um dos principais desafios da literatura econômica e quantitativa, em razão da natureza não linear, não estacionária e ruidosa das séries temporais financeiras. Índices do mercado de ações como o IBOVESPA apresentam comportamento caracterizado por mudanças abruptas, dependência temporal complexa e elevada sensibilidade a choques externos, dificultando a modelagem por métodos tradicionais (GIANTSIDI; TARANTOLA, 2025; CHUNG; ESPINOZA; QUISPE, 2025; HUSSAIN; BASHIR; REHMAN, 2024).

A literatura recente destaca que a previsibilidade em mercados financeiros é limitada e dependente do conjunto de variáveis e do horizonte temporal considerado (GU; KELLY; XIU, 2020; ZHANG et al., 2022). Além disso, a volatilidade tende a ocorrer em *clusters*, com períodos de instabilidade seguidos por relativa calma, o que reforça a necessidade de modelos capazes de capturar dependências temporais não lineares.

Diante dessas características, abordagens híbridas que combinam técnicas de filtragem e aprendizado de máquina têm sido propostas como alternativas promissoras para melhorar a capacidade preditiva em ambientes complexos e ruidosos (QIN et al., 2023).

A AES é uma técnica não paramétrica utilizada para decomposição de séries temporais em componentes interpretáveis, tais como tendência, periodicidade e ruído. Sua extensão multivariada, conhecida como AESM, permite a análise conjunta de múltiplas séries, capturando estruturas comuns entre elas.

Os fundamentos teóricos da AES e AESM são amplamente discutidos na literatura clássica, com destaque para Hassani (2007), Golyandina e Zhigljavsky (2013) e Golyandina et al. (2018), sendo considerados referências essenciais na área.

O procedimento para AESM envolve quatro etapas principais: *embedding*, decomposição, agrupamento e reconstrução.

Na etapa de *embedding*, constrói-se uma matriz trajetória a partir de uma janela de tamanho  $L$ , sendo o número de colunas definido por:

$$K = N - L + 1 \quad (1)$$

onde  $N$  representa o número de observações da série.

Na etapa de decomposição, a matriz trajetória  $X$  é decomposta por meio da decomposição em valores singulares, conforme:

$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} \mathbf{U}_i \mathbf{V}_i^T \quad (2)$$

em que  $\lambda_i$  são os autovalores associados às componentes principais e  $\mathbf{U}_i$ ,  $\mathbf{V}_i$  representam os autovetores.

Na etapa de reconstrução, a série é reconstituída a partir de um subconjunto de componentes:

$$\tilde{X} = \sum_{i \in I} \mathbf{X}_i \quad (3)$$

onde  $I$  representa o conjunto de índices selecionados, com  $I = \{1, \dots, r\}$ .

No contexto deste artigo, a AESM é utilizada como ferramenta de filtragem estrutural, permitindo separar a série em uma componente suave e um componente residual, facilitando a modelagem subsequente.

Em relação às Redes Neurais Recorrentes – RNR elas são projetadas para modelar dados sequenciais, incorporando dependência temporal entre observações. No entanto, tais redes apresentam limitações na captura de dependências de longo prazo, devido ao problema do desaparecimento do gradiente.

Para contornar essa limitação, Hochreiter e Schmidhuber (1997) propuseram o modelo MLCP, que introduz mecanismos de memória controlados por portas (*gates*).

A dinâmica do MLCP é definida por um conjunto de equações que controlam o fluxo de informação ao longo do tempo.

A porta de esquecimento é definida por:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (4)$$

A porta de entrada é dada por:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (5)$$

O estado candidato da célula é:

$$\tilde{C}_t = \tanh (W_c x_t + U_c h_{t-1} + b_c) \quad (6)$$

A atualização do estado da célula ocorre por:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (7)$$

Por fim, a saída da célula é dada por:

$$h_t = o_t \odot \tanh (C_t) \quad (8)$$

onde  $\sigma$  representa a função sigmoide e  $\odot$  denota o produto elemento a elemento.

Modelos MLCP têm sido amplamente aplicados na previsão de séries financeiras devido à sua capacidade de capturar padrões não lineares e dependências temporais complexas (FISCHER; KRAUSS, 2018; GU; KELLY; XIU, 2020).

A integração entre MSSA e MLCP permite combinar a capacidade de filtragem estrutural com o aprendizado temporal não linear.

A decomposição da série pode ser representada por:

$$X_t = S_t + R_t \quad (9)$$

onde  $S_t$  representa a componente estrutural e  $R_t$  o resíduo.

Essas componentes são utilizadas como entrada para o modelo MLCP, permitindo capturar tanto padrões de longo prazo quanto variações de maior frequência. Estudos recentes indicam que abordagens híbridas podem melhorar o desempenho preditivo em ambientes com alto nível de ruído (ZHANG et al., 2022; QIN et al., 2023).

### 3 METODOLOGIA

Quanto à sua natureza, essa pesquisa pode ser classificada como aplicada, pois busca produzir conhecimento com finalidade prática, voltado à construção e avaliação de um modelo de previsão do risco diário do IBOVESPA a partir de técnicas de decomposição de séries temporais e aprendizado. Segundo Gil (2008), a pesquisa aplicada volta-se à solução de problemas concretos, preservando o rigor científico na escolha dos procedimentos e na interpretação dos resultados.

No que se refere à abordagem do problema, o estudo adota uma abordagem quantitativa, uma vez que utiliza dados numéricos, transformação matemática das séries, estimação computacional de modelos e avaliação de desempenho por métricas estatísticas. De acordo com Prodanov e Freitas

(2013), a pesquisa quantitativa caracteriza-se pelo uso da mensuração, da modelagem e do tratamento estatístico como elementos centrais da investigação.

Quanto aos objetivos, a pesquisa apresenta natureza explicativa, pois busca compreender em que medida a combinação entre a AESM e redes neurais do tipo MLCP pode melhorar a previsão do risco diário do IBOVESPA. Em conformidade com Gil (2008), pesquisas explicativas procuram identificar fatores que contribuem para a ocorrência de fenômenos, aprofundando a compreensão de suas relações.

Em relação aos procedimentos técnicos, trata-se de uma pesquisa documental com delineamento computacional-experimental. É documental porque utiliza dados secundários obtidos em base pública de mercado, sem tratamento analítico prévio pelo pesquisador, característica compatível com a definição proposta por Gil (2008). Simultaneamente, apresenta delineamento computacional-experimental por desenvolver, testar e comparar rotinas algorítmicas implementadas em linguagem de programação, com o objetivo de produzir previsões a partir de séries temporais financeiras.

Adicionalmente, a pesquisa possui caráter empírico-analítico, uma vez que parte de uma base observacional, constrói variáveis de interesse, estima modelos e compara resultados a partir de critérios quantitativos. Nessa perspectiva, combina elementos de modelagem aplicada e avaliação experimental controlada, ainda que não se trate de experimento no sentido clássico, mas de experimento computacional sobre dados históricos.

A estratégia metodológica foi organizada em duas etapas principais. Na primeira etapa, foi realizada uma otimização automática de parâmetros, com o objetivo de identificar a melhor combinação entre parâmetros do AESM e da rede MLCP. Na segunda etapa, a melhor configuração encontrada foi utilizada para o teste ou para geração das previsões a serem comparadas no conjunto de teste.

A lógica central da pesquisa consistiu em combinar um método de decomposição estrutural multivariada com um método de aprendizado sequencial. Em termos operacionais, a AESM foi utilizada como etapa de filtragem e extração de estrutura das séries temporais, enquanto MLCP foi utilizada como modelo preditivo propriamente dito. Dessa forma, a modelagem buscou explorar simultaneamente a capacidade de redução de ruído e separação estrutural e a capacidade de aprender dependências temporais não lineares.

Todo o desenvolvimento empírico foi realizado na linguagem R, em ambiente compatível com a execução de rotinas estatísticas, manipulação de dados, decomposição espectral e aprendizado profundo. Os principais pacotes utilizados foram: *readxl*, para leitura da base em formato Excel; *dplyr*, para organização, transformação e manipulação dos dados; *Rssa*, para execução da AES e da AESM; *keras* e *reticulate*, para construção, treinamento e previsão com redes MLCP em interface com o *TensorFlow*; *ggplot2*, para construção dos gráficos comparativos entre valores observados e previstos.

A escolha do ambiente em R dá-se por sua ampla utilização em pesquisas quantitativas e pela disponibilidade de bibliotecas consolidadas para séries temporais, modelagem estatística e integração com aprendizado profundo.

### 3.1 AMOSTRA E MANIPULAÇÃO DOS DADOS

A amostra utilizada no estudo é não probabilística, de natureza intencional e documental, composta por observações históricas diárias de mercado disponíveis publicamente. Segundo Prodanov e Freitas (2013), amostras não probabilísticas são apropriadas quando a seleção decorre da disponibilidade e da pertinência dos dados ao problema de pesquisa.

Os dados foram obtidos no portal ADVFN, abrangendo o período de 15 de fevereiro de 2021 a 27 de março de 2026. A utilização conjunta dos índices IBOVESPA, IBrX-100 e IBrX-50 justifica-se pelo interesse em capturar não apenas o comportamento do IBOVESPA isoladamente, mas também a estrutura multivariada do mercado brasileiro, permitindo ao MSSA identificar componentes comuns entre séries correlacionadas.

Inicialmente, os dados históricos dos índices foram organizados em planilha eletrônica e importados para o ambiente R. Após a importação, os registros foram ordenados cronologicamente de forma crescente, preservando a sequência temporal necessária à modelagem.

Para cada índice foram coletados, em periodicidade diária, os valores de: abertura; fechamento; máxima do dia; mínima do dia. Embora os dados de abertura e fechamento tenham sido preservados na base original, a variável central utilizada na modelagem foi construída a partir dos valores de máxima do dia e de mínima do dia, por serem mais diretamente associados à amplitude diária de variação do mercado de ações.

A variável de interesse do estudo foi definida como uma medida diária de risco intradiário, construída a partir da razão entre a máxima e a mínima observadas em cada pregão. Para cada índice, o risco foi calculado por meio do logaritmo natural dessa razão, conforme:

$$risk_t = \log \left( \frac{High_t}{Low_t} \right) \quad (10)$$

em que:

$risk_t$  representa o risco no dia  $t$ ;  $High_t$  representa o valor máximo do índice no dia  $t$ ;  $Low_t$  representa o valor mínimo do índice no dia  $t$ .

A partir dessa formulação, foram obtidas três séries de risco:

$ib\_risk_t$ , para o IBOVESPA;  $xx\_risk_t$ , para o IBrX 100;  $xl\_risk_t$ , para o IBrX 50.

A escolha dessa medida decorre do interesse em capturar a amplitude relativa intradiária do mercado, preservando a comparabilidade ao longo do tempo.

Após a construção das séries de risco, foi aplicada a técnica de AESM. Nessa etapa, as séries  $ib\_risk_t$ ,  $xx\_risk_t$  e  $xl\_risk_t$  foram organizadas em estrutura multivariada, permitindo extrair componentes comuns e reduzir o efeito de ruídos específicos.

A aplicação do MSSA envolveu os parâmetros centrais:

- $L$ , correspondente ao comprimento da janela de *embedding*;
- $r$ , correspondente ao número de componentes reconstruídas.

A matriz trajetória foi construída a partir da janela  $L$ , sendo o número de colunas dado conforme equação (01).

A decomposição foi feita por meio da decomposição em valores singulares, com posterior reconstrução das componentes escolhidas. O produto final dessa etapa foi a obtenção de três componentes suavizadas:  $ib\_mssa_t$ ;  $xx\_mssa_t$ ;  $xl\_mssa_t$ .

Em seguida, foram calculados os resíduos entre a série original e a componente reconstruída, conforme:

$$res_{ib,t} = ib\_risk_t - ib\_mssa_t \quad (11)$$

$$res_{xx,t} = xx\_risk_t - xx\_mssa_t \quad (12)$$

$$res_{xl,t} = xl\_risk_t - xl\_mssa_t \quad (13)$$

Esses resíduos foram mantidos por representarem a parcela de alta frequência não capturada pela componente estrutural.

Após a decomposição, as séries utilizadas como entrada do modelo foram normalizadas por meio da transformação min-max, com o objetivo de colocá-las em escala comparável e facilitar o treinamento da rede neural.

A transformação utilizada pode ser expressa como:

$$x_t^* = \frac{x_t - \min(x)}{\max(x) - \min(x)} \quad (14)$$

em que:

$x_t^*$  representa o valor normalizado;  $x_t$  representa o valor original;  $\min(x)$  e  $\max(x)$  representam os valores mínimo e máximo da série.

Essa padronização foi aplicada às variáveis de entrada e à variável-alvo.

Como a rede MLCP opera sobre dados sequenciais, foi necessário converter a base em janelas temporais de comprimento fixo, denominadas *lookback*. Cada observação de entrada passou a ser composta por uma sequência de *lookback* dias anteriores, contendo as seguintes seis variáveis:  $ib_{mssa_t}$ ;  $xx_{mssa_t}$ ;  $xl_{mssa_t}$ ;  $res_{ib,t}$ ;  $res_{xx,t}$ ;  $res_{xl,t}$ .

Assim, para cada instante de previsão, o modelo recebia uma matriz de dimensão:

$$lookback \times 6 \quad (15)$$

e produzia como saída a previsão do valor futuro de *ib\_risk*.

Em termos práticos, cada sequência continha a memória recente do sistema multivariado, permitindo à LSTM aprender padrões dinâmicos tanto da componente estrutural quanto das variações residuais.

A arquitetura final utilizada foi composta por: uma camada de entrada compatível com o número de passos temporais e variáveis; uma primeira camada LSTM com retorno de sequências; uma camada de *dropout* para redução de sobreajuste; uma segunda camada LSTM e uma camada densa final com uma unidade de saída.

A função de perda adotada foi o erro quadrático médio – EQM (à média dos quadrados das diferenças entre os valores observados e os valores previstos pelo modelo), e o algoritmo de otimização utilizado foi o Adam. Além disso, empregou-se o mecanismo de *early stopping*, de modo a interromper o treinamento quando não houvesse melhoria adicional na perda de validação após determinado número de épocas.

A etapa de otimização foi concebida para tornar adaptável a outras séries e contextos, evitando a fixação arbitrária de parâmetros.

Foram testadas automaticamente combinações entre:

- $r = 2,3,4,5,6,7,8,9$ ;
- $L = 10,15,20$ ;
- $lookback = 10,15,20,30$ .

Ao todo, foram avaliadas 96 combinações. Para cada combinação: aplicou-se AESM às três séries de risco; reconstruíram-se as componentes estruturais; calcularam-se os resíduos; normalizaram-se as variáveis; construíram-se as sequências temporais; treinou-se a rede MLCP; geraram-se previsões para o conjunto de teste; calcularam-se métricas de desempenho. Essa etapa foi realizada no período de treinamento.

A melhor configuração foi escolhida com base, prioritariamente, no maior valor da correlação entre série observada e série prevista, considerando-se adicionalmente os menores valores de EQM.

A base foi dividida cronologicamente em:

- 80% dos dados para treinamento;
- 20% dos dados para teste.

Essa divisão respeitou a ordem temporal da série, evitando embaralhamento e preservando a lógica de previsão fora da amostra. O procedimento está alinhado à literatura de previsão fora da amostra, segundo a qual a avaliação de modelos preditivos em séries temporais deve respeitar a estrutura temporal dos dados, utilizando-se um menor parcela das observações históricas para estimação e observações posteriores para teste de desempenho preditivo (TASHMAN, 2000).

O desempenho do modelo final foi avaliado por meio de três métricas.

Erro absoluto médio:

$$EAM = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (16)$$

Raiz do erro quadrático médio:

$$REQM = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (17)$$

Correlação linear entre observado e previsto:

$$COR = corr(y_t, \hat{y}_t) \quad (18)$$

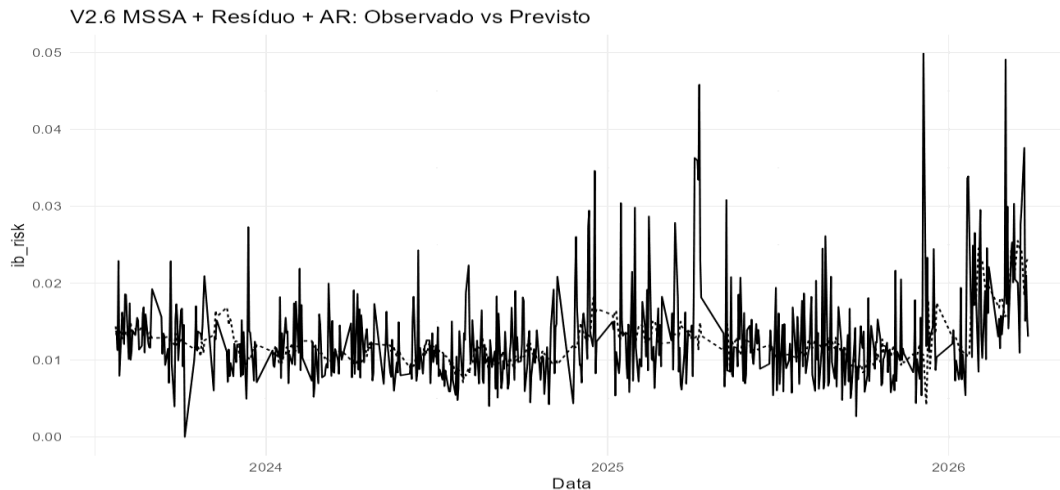
em que  $y_t$  representa o valor observado;  $\hat{y}_t$  representa o valor previsto.

Além das métricas numéricas, a avaliação incluiu análise gráfica da aderência entre a série observada e a série prevista.

## 4 RESULTADOS

Inicialmente, foram avaliadas abordagens baseadas em decomposição estrutural por AESM, bem como extensões com modelos autorregressivos, definido no R como V2.6 MSSA+AR. Os resultados obtidos nessa etapa indicaram desempenho limitado em termos de aderência à série observada.

Figura 1 – Comparação entre a série de risco observada e a série prevista pelo modelo MSSA–Resíduo + AR no conjunto de teste



Fonte: Elaborado pelo autor.

Os ótimos ( $L$  e  $r$ ) foram definidos conforme a literatura (HASSANI, 2007). A Figura 1 demonstra que o modelo não apresenta aderência à dinâmica da série.

Observa-se que a série prevista (linha tracejada) não acompanha as oscilações ao longo do tempo (linha sólida), não capturando tanto movimentos ascendentes quanto descendentes, ao contrário trata-se de uma linha suavizada.

No caso da AESM mais AR, verificou-se que a série reconstruída apresentava suavização, resultando em previsões com baixa sensibilidade às variações de curto prazo. Modelos complementares, como estruturas autorregressivas, não produziram ganhos relevantes na captura da dinâmica não linear da série. A aplicação resultou em métricas com coeficientes de correlação tipicamente inferiores a 0,30, além de previsões com comportamento estável e EAM e RSQM acima 0,09.

De forma geral, essas abordagens produziram resultados caracterizados por baixa capacidade de antecipação de movimentos e aderência limitada à amplitude da série.

#### 4.1 RESULTADOS DA OTIMIZAÇÃO DE PARÂMETROS

A etapa de otimização automática para AESM e MLCP avaliou 96 combinações de parâmetros, considerando diferentes valores para o número de componentes para AESM ( $r$ ), o tamanho da janela de *embedding* ( $L$ ) e o comprimento da sequência temporal de entrada (*lookback*).

Tabela 1 – Combinações de parâmetros em AESM–MLCP ordenadas por correlação

<b>r</b>	<b>L</b>	<b>lookback</b>	<b>COR</b>	<b>REQM</b>	<b>EAM</b>
5	10	30	0,8555	0,0795	0,0642
5	10	15	0,8300	0,0728	0,0559
7	20	20	0,8201	0,0763	0,0570
3	15	20	0,8200	0,0863	0,0695
8	20	20	0,8140	0,0752	0,0570

Fonte: Elaborado pelo autor.

Observa-se que a melhor configuração foi obtida para  $r = 5$ ,  $L = 10$  e  $lookback = 30$ , com coeficiente de correlação de aproximadamente 0,86. Esse resultado indica que a utilização de uma janela menor na AESM, associada a uma memória temporal mais longa no modelo MLCP, favorece a captura da dinâmica da série.

Considerando que a base de dados compreende o período de 15 de fevereiro de 2021 a 27 de março de 2026, o conjunto de teste corresponde aproximadamente ao intervalo mais recente da série, situado entre março de 2025 a março de 2026, preservando a ordem cronológica e evitando vazamento de informação futura. Para o R, o modelo foi definido PIPELINE LSTM + MSSA, terminologia em inglês para MLCP e AESM. Nesse contexto, as métricas de desempenho foram calculadas exclusivamente sobre esse subconjunto final, representando, portanto, uma avaliação fora da amostra (*out-of-sample*).

Com base na melhor configuração identificada, os resultados obtidos no conjunto de teste foram:

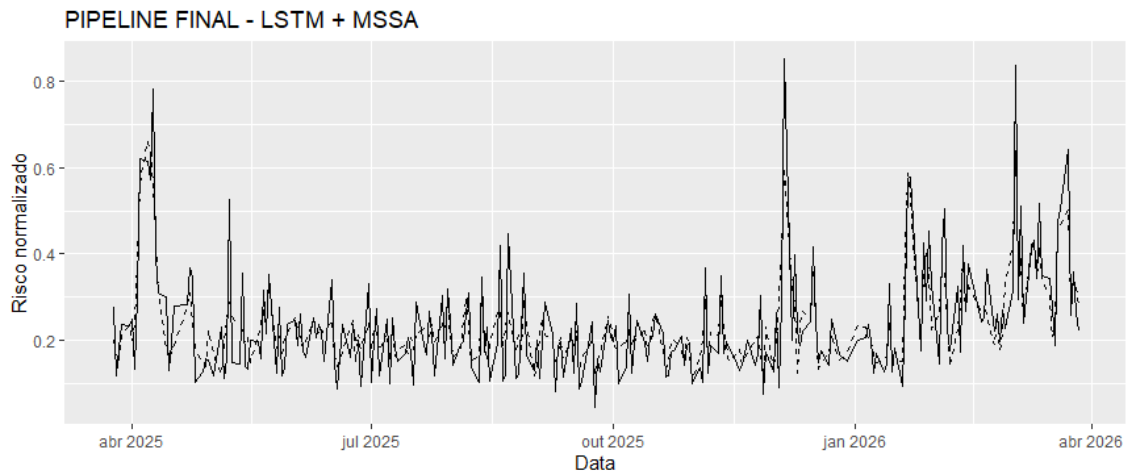
- EAM = 0,051964
- RSQM = 0,066677
- COR = 0,895537

É importante destacar que essas métricas foram calculadas sobre a variável de risco previamente normalizada no intervalo  $[0, 1]$ . Dessa forma, os valores de erro representam desvios médios e quadráticos dentro dessa escala padronizada.

O EAM indica que, em média, a diferença entre valores observados e previstos é de aproximadamente 0,052 unidades na escala normalizada. Já o REQM, por penalizar mais fortemente erros de maior magnitude, indica um desvio típico de aproximadamente 0,067. Esses valores sugerem que o modelo apresenta boa precisão, considerando a variabilidade da série. COR de 0,895537 evidencia associação linear entre valores observados e previstos, indicando que o modelo é capaz de reproduzir adequadamente a direção e o comportamento geral da série ao longo do tempo.

A Figura 2 demonstra que o modelo apresenta aderência à dinâmica da série.

Figura 2 – Comparação entre a série de risco observada e a série prevista pelo modelo MSSA–LSTM no conjunto de teste



Fonte: Elaborado pelo autor.

Diferentemente das abordagens preliminares, observa-se que a série prevista (linha tracejada) acompanha as oscilações ao longo do tempo (linha sólida), capturando tanto movimentos ascendentes quanto descendentes.

Além disso, nota-se que o modelo deixa de apresentar comportamento suavizado, passando a refletir variações mais próximas da série real. Entretanto, ainda se observa que eventos de maior magnitude tendem a ser parcialmente subestimados, indicando limitação na captura completa da amplitude dos extremos.

## 5 CONSIDERAÇÕES FINAIS

Os resultados obtidos evidenciam que a abordagem híbrida AESM–MLCP apresenta desempenho superior em relação às estratégias previamente testadas. A utilização da AESM como etapa de filtragem contribuiu para a redução de ruído, enquanto o modelo MLCP foi capaz de capturar padrões temporais não lineares presentes na série.

A melhora observada no coeficiente de correlação, que atingiu aproximadamente 0,90, representa um avanço em relação às abordagens anteriores, que apresentavam desempenho substancialmente inferior.

Apesar disso, o modelo ainda apresenta limitações, especialmente na representação de eventos extremos, que tendem a ser suavizados. Esse comportamento é consistente com a literatura, que aponta a dificuldade de previsão precisa de movimentos abruptos em mercados financeiros.

De forma geral, os resultados confirmam que a integração entre técnicas de decomposição e aprendizado profundo constitui uma abordagem promissora para a previsão de risco em séries financeiras, ainda que desafios permaneçam na modelagem de eventos extremos.

Essa pesquisa apresenta limitações que devem ser consideradas na interpretação dos achados. A avaliação do modelo foi realizada a partir de uma única divisão temporal da amostra, com separação

entre dados de treinamento e teste. Embora essa estratégia preserve a ordem cronológica e evite vazamento de informação, não permite avaliar plenamente a estabilidade do modelo em diferentes janelas temporais, o que poderia ser obtido por meio de validações do tipo *walk-forward*.

A variável de interesse foi construída a partir da razão entre máxima e mínima diária, representando uma medida específica de risco intradiário. Ainda que essa abordagem capture a amplitude de variação do mercado, outras métricas de risco, como volatilidade baseada em retornos ou medidas derivadas de modelos econométricos, não foram consideradas, o que pode limitar a generalização dos resultados.

A modelagem concentrou-se exclusivamente em informações endógenas de mercado, não incorporando variáveis macroeconômicas, indicadores de sentimento ou eventos exógenos. Considerando que mercados financeiros são fortemente influenciados por fatores externos, a ausência dessas variáveis pode restringir a capacidade do modelo de antecipar determinados movimentos.

Por fim, embora tenha sido estruturado de forma adaptativa, a escolha dos intervalos de parâmetros testados ainda representa uma delimitação do espaço de busca. Combinações fora desse conjunto não foram exploradas, o que pode implicar a existência de configurações potencialmente superiores não avaliadas nessa pesquisa.

Diante dessas limitações, diversas possibilidades de aprofundamento podem ser exploradas em trabalhos futuros. Uma primeira direção consiste na adoção de procedimentos de validação mais robustos, como o *walk-forward*, permitindo avaliar a estabilidade do modelo ao longo do tempo e em diferentes condições do mercado de ações.

Outra possibilidade refere-se à ampliação da variável de interesse, incluindo outras medidas de risco e volatilidade, como retornos logarítmicos, volatilidade realizada ou estimativas baseadas em modelos GARCH, possibilitando comparação entre diferentes definições de risco.

Adicionalmente, a incorporação de variáveis exógenas, tais como indicadores macroeconômicos, taxas de juros, câmbio e índices internacionais, pode contribuir para enriquecer a capacidade preditiva do modelo, permitindo capturar efeitos externos ao mercado doméstico.

No âmbito metodológico, novas arquiteturas de redes neurais podem ser exploradas, como modelos baseados em mecanismos de atenção (*attention*) ou arquiteturas híbridas mais complexas, que combinem diferentes abordagens de aprendizado.

Por fim, a aplicação proposta a outros ativos financeiros, setores específicos ou mercados internacionais pode contribuir para avaliar sua capacidade de generalização e robustez em diferentes contextos, ampliando a validade externa dos resultados obtidos.



## REFERÊNCIAS

- CHUNG, Víctor; ESPINOZA, Jenny; QUISPE, Renán. Forecasting financial volatility under structural breaks: a comparative study of GARCH models and deep learning techniques. *Journal of Risk and Financial Management*, v. 18, n. 9, p. 494, 2025. DOI: 10.3390/jrfm18090494.
- CONT, Rama. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, v. 1, n. 2, p. 223-236, 2001. DOI: 10.1080/713665670.
- FISCHER, Thomas; KRAUSS, Christopher. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, v. 270, n. 2, p. 654–669, 2018.
- GIANTSIDI, Sofia; TARANTOLA, Claudia. Deep learning for financial forecasting: a review of recent trends. *International Review of Economics & Finance*, 2025. DOI: 10.1016/j.iref.2025.104719.
- GIL, Antonio Carlos. *Métodos e técnicas de pesquisa social*. 6. ed. São Paulo: Atlas, 2008.
- GOLYANDINA, Nina; ZHIGLJAVSKY, Anatoly. *Singular Spectrum Analysis for Time Series*. Berlin: Springer, 2013.
- GOLYANDINA, Nina; NEKRUTKIN, Vladimir; ZHIGLJAVSKY, Anatoly. *Analysis of time series structure: SSA and related techniques*. Boca Raton: CRC Press, 2018.
- GU, Shihao; KELLY, Bryan; XIU, Dacheng. Empirical asset pricing via machine learning. *The Review of Financial Studies*, v. 33, n. 5, p. 2223–2273, 2020.
- HASSANI, Hossein. Singular spectrum analysis: methodology and comparison. *Journal of Data Science*, v. 5, n. 2, p. 239–257, 2007.
- HOCHREITER, Sepp; SCHMIDHUBER, Jürgen. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735–1780, 1997.
- HUSSAIN, Muntazir; BASHIR, Usman; REHMAN, Ramiz Ur. Exchange rate and stock prices volatility connectedness and spillover during pandemic induced-crises: evidence from BRICS countries. *Asia-Pacific Financial Markets*, v. 31, p. 183-203, 2024. DOI: 10.1007/s10690-023-09411-0
- MACIEL, Leandro dos Santos; BALLINI, Rosangela. Value-at-risk modeling and forecasting with range-based volatility models: empirical evidence. *Revista Contabilidade & Finanças*, v. 28, n. 75, p. 361-376, 2017. DOI: 10.1590/1808-057x201704140.
- POON, Ser-Huang; GRANGER, Clive W. J. Forecasting volatility in financial markets: a review. *Journal of Economic Literature*, v. 41, n. 2, p. 478-539, 2003. DOI: 10.1257/002205103765762743
- PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. *Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico*. 2. ed. Novo Hamburgo: Feevale, 2013.
- TASHMAN, Leonard J. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, v. 16, n. 4, p. 437-450, 2000. DOI: 10.1016/S0169-2070(00)00065-0



QIN, Yong; et al. Hybrid deep learning models for financial time series forecasting. *Applied Soft Computing*, 2023.

ZHANG, Yue; et al. Financial time series prediction using hybrid models: a review and new directions. *Expert Systems with Applications*, 2022.