

Técnicas de filtragem adaptativa e aprendizado de máquina para redução de erros sistemáticos em modelos climáticos

Diogo S. Magalhães, Lisandro Lovisoló, Michel P. Tcheou, Sin Chan Chou

Resumo—Neste trabalho, aplicam-se técnicas de filtragem adaptativa e aprendizado de máquina para reduzir o viés nas previsões do modelo climático *Brazilian Earth System Model* (BESM), desenvolvido pelo Instituto Nacional de Pesquisas Espaciais. A acurácia é avaliada com base nos dados de reanálise ERA5, do *European Centre for Medium-Range Weather Forecasts*. Utilizam-se os métodos *Affine Projection*, *Recursive Least Squares*, *Random Forest Regression* e *Multilayer Perceptrons* para pós-processar a temperatura da superfície do mar prevista pelo BESM. Os resultados mostram redução dos erros sistemáticos, contribuindo para metodologias de aprimoramento de modelos de previsão climática e eventos extremos com auxílio de técnicas computacionais eficientes.

Palavras-Chave—Aprendizado de máquina, Filtros adaptativos, Modelos climáticos, Temperatura de superfície do mar, INPE, BESM, ERA5.

Abstract—In this work, adaptive filtering and machine learning techniques are applied to reduce bias in the predictions of the *Brazilian Earth System Model* (BESM) climate model, developed by the National Institute for Space Research. Accuracy is evaluated based on ERA5 reanalysis data from the *European Centre for Medium-Range Weather Forecasts*. The *Affine Projection*, *Recursive Least Squares*, *Random Forest Regression* and *Multilayer Perceptrons* methods are used to post-process the sea surface temperature predicted by BESM. The results show a reduction in systematic errors, contributing to methodologies for improving climate prediction models and extreme events with the aid of efficient computational techniques.

Keywords—Machine learning, Adaptive filtering, Climate models, Sea surface temperature, INPE, BESM, ERA5

I. INTRODUÇÃO

Previsões climáticas são essenciais para fornecer informações estratégicas para o planejamento de diversos setores e atividades, como agricultura, produção de energia, planejamento urbano e gestão de riscos de desastres, entre outros [1]. Nesse contexto, previsões mais acuradas auxiliam formuladores de políticas e outras partes interessadas a tomar decisões com informações mais embasadas. O Painel Intergovernamental sobre Mudanças Climáticas (IPCC, *Intergovernmental Panel on Climate Change*) fornece avaliações periódicas das mudanças climáticas [2], visando antecipar variações de temperatura,

Diogo S. Magalhães, FEN, UERJ, Rio de Janeiro-RJ, e-mail: ddiogorj@gmail.com; Lisandro Lovisoló, FEN, UERJ, Rio de Janeiro-RJ, e-mail: lisandro@uerj.br; Michel P. Tcheou, FEN, UERJ, Rio de Janeiro-RJ, e-mail: mtcheou@uerj.br; Sin Chan Chou, INPE, Cachoeira Paulista-SP, e-mail: chou.chan@inpe.br. Este trabalho foi parcialmente apoiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) Código Financeiro 001 e pela Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

padrões de precipitação e eventos climáticos extremos, o que pode facilitar a formulação de estratégias, reduzir riscos potenciais e maximizar oportunidades para o desenvolvimento sustentável.

A complexidade do sistema climático terrestre introduz incertezas nas previsões em diferentes estágios – assimilação de dados insuficientes, dados com baixa qualidade, erros de representação e simplificações nos modelos podem comprometer a qualidade e a acurácia das previsões, reduzindo sua confiabilidade [3]. Diante desses desafios, abordagens computacionais têm sido propostas para mitigar erros sistemáticos de previsões climáticas, incluindo técnicas de aprendizado de máquina (*Machine Learning* - ML) e filtragem adaptativa [4], [5], [6], [7].

O presente trabalho segue nessa linha e propõe uma abordagem integrada para a redução de viés em modelos climáticos por meio da aplicação de filtros adaptativos e técnicas de ML. São investigados os filtros adaptativos *Affine Projection Algorithm* (APA) e *Recursive Least Squares* (RLS) [8], os modelos de ML *Random Forest Regression* (RFR) e redes neurais do tipo *Multilayer Perceptrons* (MLP) [9], [10]. Cada um desses métodos é aplicado para pós-processar a Temperatura da Superfície do Mar (*Sea Surface Temperature* - SST) prevista pelo modelo climático brasileiro BESM [11], utilizando como referência os dados do conjunto de reanálise ERA5 [12].

O BESM, desenvolvido pelo INPE, é um modelo acoplado oceano-atmosfera utilizado para prever variáveis como a SST e correntes oceânicas. A SST é uma variável-chave para a compreensão e modelagem do clima em regiões tropicais, e dispor de valores acurados da mesma impacta diretamente a qualidade de projeções de eventos extremos [13].

Neste trabalho, aplicam-se cada uma das técnicas mencionadas com a finalidade de corrigir erros de previsão climática. Para isso, a SST prevista pelo BESM é processada por cada uma das técnicas, utilizando como referência o conjunto de dados observacionais ERA5 reanálise climática global desenvolvida pelo *European Centre for Medium-Range Weather Forecast* (ECMWF) [14], [12].

As diferentes técnicas foram comparadas visando analisar aquela que apresenta maior capacidade de correções dos erros de previsão da SST do BESM. Identificada a melhor técnica, esta pode ser empregada para aprimorar previsões climáticas e, por consequência, conferir maior confiabilidade nas projeções climáticas e de eventos extremos.

Após esta introdução, as seções subsequentes detalham: o erro em modelos climáticos (Seção II), as técnicas utiliza-

das (Seção III), a metodologia empregada (Seção IV) e os resultados obtidos (Seção V), nesta ordem. Na Seção VI, são apresentadas as contribuições e conclusões, bem como sugestões de aprimoramentos futuros.

II. ERRO DE PREVISÃO EM MODELOS CLIMÁTICOS

O desenvolvimento de modelos climáticos remonta ao início do século XX, quando se postulou a possibilidade de prever o clima por meio de equações matemáticas baseadas nas leis físicas da atmosfera [13]. Contudo, a complexidade desses sistemas e as incertezas nas medições de variáveis físicas inviabilizaram aplicações práticas à época. Nas últimas décadas, avanços computacionais permitiram aprimorar significativamente os modelos climáticos, embora erros sistemáticos persistam em regiões críticas, como os trópicos [15].

Os erros em modelos climáticos podem ser causados por múltiplos fatores, tais como: resolução espacial limitada, em função de tamanhos de grade muito amplos; dados de entrada imprecisos, incluindo observações incompletas ou com viés instrumental; simplificação de processos físicos complexos; lacunas no conhecimento científico, especialmente em mecanismos de retroalimentação climática, cuja dinâmica ainda não é totalmente compreendida [13]. Assim, o uso de resultados não corrigidos em modelos ou avaliações de impacto climático pode frequentemente produzir resultados divergentes dos observados [16], [17], [13].

Neste contexto, a Figura 1 ilustra o erro que se observa na variável prognóstica SST; é possível observar a diferença (área hachurada) entre o valor previsto pelo modelo BESM e o valor de referência ERA5 para a coordenada 52.50°S, 44.25°E, no período de novembro de 2007 a outubro de 2008.

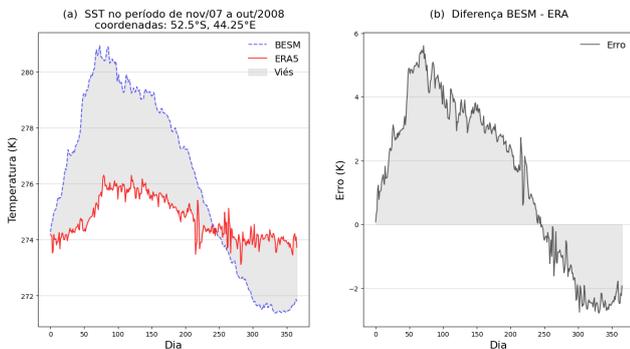


Fig. 1. (a) SST no intervalo de nov. de 2007 a out. de 2008 para os modelos BESM e ERA5 na posição 52.50°S, 44.25°E. (b) Apresenta a diferença BESM - ERA5, i.e., o erro de previsão da SST.

III. MODELOS PARA CORREÇÃO DE VIÉS

Este trabalho explora técnicas de filtragem adaptativa e ML para redução de erros de previsão como os ilustrados na Figura 1. Os modelos utilizados neste trabalho compartilham como princípio fundamental a minimização de uma função de custo baseada no erro entre a saída de um modelo e a referência observacional [9], [10], [8]. Especificamente, os modelos visam minimizar o erro entre a previsão do modelo climático BESM e os dados do conjunto de reanálise ERA5. Tal tarefa pode ser formulada como um problema de otimização, em que

o objetivo é ajustar os parâmetros do modelo de correção para reduzir a discrepância entre a previsão corrigida e os dados de referência.

Os filtros adaptativos utilizados, *Affine Projection* (AP) e *Recursive Least Squares* (RLS), são métodos lineares iterativos que atualizam os coeficientes a cada nova amostra de entrada, buscando minimizar uma função custo baseada em erro quadrático [8]. O filtro AP é uma extensão do algoritmo LMS e emprega múltiplas amostras passadas para formar um subespaço afim, no qual o vetor de coeficientes é ajustado por projeção. Isso resulta em uma convergência mais rápida e estável em ambientes com sinais correlacionados. Já o RLS utiliza uma estratégia recursiva para minimizar a soma ponderada dos erros quadráticos passados, aplicando maior peso às observações mais recentes por meio de um fator de esquecimento. Sua principal vantagem é a velocidade de convergência, mesmo sob alta dispersão espectral do sinal de entrada, embora exija maior complexidade computacional.

Os modelos de ML empregados, *textitMultilayer Perceptron* (MLP) e *Random Forest Regression* (RFR), seguem o mesmo princípio de redução de erro através de uma função custo [9], [10]. O MLP é uma rede neural com múltiplas camadas e funções de ativação não lineares, treinada para minimizar uma função de erro a partir da qual os pesos são ajustados via retropropagação. Este modelo é capaz de aprender representações complexas dos dados, o que o torna adequado para capturar relações não lineares entre as variáveis de entrada e saída. O RFR, por sua vez, é um modelo baseado em árvores de decisão, no qual árvores são treinadas usando subconjuntos dos dados (com amostragem e seleção de variáveis aleatórias), e suas previsões são combinadas por média. Essa estratégia reduz a variância do modelo e melhora sua capacidade de generalização.

Apesar das diferenças estruturais, todos os métodos considerados aprendem, com base em dados históricos, parâmetros que lhes permitem reduzir o erro sistemático presente nas previsões do modelo climático. A escolha das abordagens empregadas deve-se à sua eficiência computacional e estabilidade numérica.

IV. MÉTODO

O fluxo de processamento aplicado para reduzir os erros de previsões é ilustrado na Figura 2. Cada um dos modelos é treinado (considerando o mesmo conjunto de amostras), utilizando como entrada os dados da SST do modelo BESM (período: nov/2007 a out/2008) e, como saída (alvo) dos modelos, os dados observacionais do ERA5 no mesmo período. Após o treinamento, o modelo é aplicado para processar as previsões do BESM do período subsequente (nov/2008 a out/2009), resultando na previsão ajustada por cada modelo.

A. Conjunto Amostra

As variáveis do BESM e do ERA5 estão dispostas em uma grade global com 720 divisões na latitude e 1440 na longitude com uma resolução espacial de 0.25°, cada célula resultante (das 720×1440) possui a SST diária em um período de 365 dias (366 no caso de anos bissextos). Note-se que apenas aproximadamente 60% da grade é válida, uma vez que

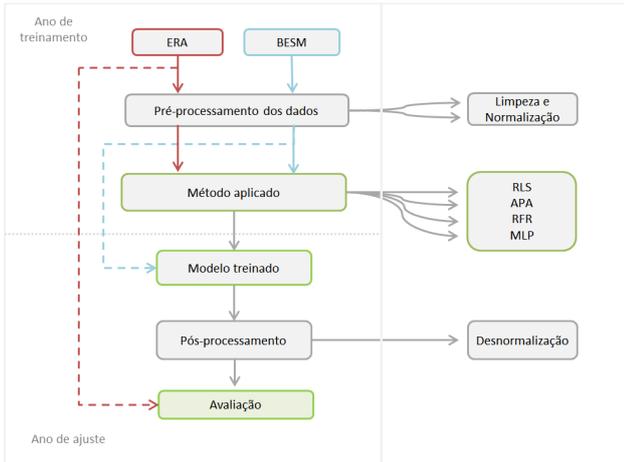


Fig. 2. Fluxo do processamento dos dados.

as células predominantemente terrestres não apresentam SST. Devido a limitações computacionais, utiliza-se uma estratégia de amostragem espacialmente balanceada, selecionando aleatoriamente 500 células. O tamanho da amostra foi definido utilizando inferência estatística do conjunto de dados, com margem de erro de 5% e intervalo de confiança de 95% [18], buscando que as células que compõem o conjunto de amostras sejam representativas do conjunto total. Dessa forma, reduz-se custo computacional para testar as técnicas propostas, enquanto se captura a variabilidade espacial e temporal dos dados.

B. Pré-processamento

O pré-processamento de dados é uma etapa fundamental no desenvolvimento de modelos de ML. A remoção de ruído e de *outliers*, bem como a padronização dos dados, é uma prática recomendada para melhorar a qualidade de modelos de aprendizado. A padronização da variável SST dos modelos BESM (período: 2007-2008 e 2008-2009) e ERA5 (período: 2007 - 2008) foi obtida por meio da normalização [19]

$$z_i = \frac{x_i - \mu}{\sigma}, \quad (1)$$

onde z_i é o valor da variável padronizada, x_i é o valor individual da variável analisada, μ é a média de x no conjunto amostral ($X = x_i, i = 1 \dots 365$) e σ seu desvio padrão.

1) *Conjunto de Entrada – Vetor de Características*: As amostras de entrada são obtidas usando a técnica da janela deslizante. Obtêm-se vetores/subconjuntos de tamanho fixo (janela), deslizando uma janela sobre um conjunto de dados maior (uma série temporal de previsões) incrementalmente [20]. Seja $X = \{x_1, x_2, \dots, x_L\}$ o conjunto de dados, e seja W o tamanho da janela. A operação de janela deslizante obtém na posição i o vetor de entrada

$$X_i = \{x_{i-W+1}, x_{i-W+2}, \dots, x_{i-1}, x_i\}. \quad (2)$$

A janela desliza sobre o conjunto de dados, um elemento de cada vez, para cada posição i , a janela captura o subconjunto W_i . A escolha do tamanho da janela (de 4 dias) baseou-se em testes preliminares que compararam janelas de 4, 15 e 30 dias.

A janela menor apresentou maior sensibilidade a variações de curto prazo, capturando flutuações rápidas na SST.

Assim, numa dada célula e instante temporal (dia), a entrada é composta das 3 últimas previsões do BESM, da previsão corrente e dos dados de reanálise do ERA5 correspondentes. O modelo ajusta a previsão do BESM para o dia corrente.

2) *Mudança de domínio*: Neste estudo, são utilizadas duas abordagens para a correção de erros sistemáticos da SST do BESM: (i) a análise convencional no domínio do tempo, aplicada individualmente a cada célula; e (ii) a transformação para o domínio da frequência espacial por meio da DCT-2D em blocos 8×8 [21], [22]. A primeira abordagem preserva as variações temporais locais, mas pode ignorar correlações espaciais entre as previsões em células adjacentes. A segunda tenta capturar padrões espectrais espaciais.

C. Configuração dos Modelos

A configuração de cada modelo foi realizada respeitando suas particularidades inerentes e buscando os hiperparâmetros que resultaram no menor erro quadrático médio (MSE) da previsão corrigida em relação aos dados de referência ERA5. O MSE foi utilizado como função objetivo para a otimização, garantindo que o modelo minimizasse a diferença quadrática entre os valores corrigidos e os observados.

A Tabela I resume os intervalos utilizados na busca em grade dos hiperparâmetros dos filtros adaptativos, ajustados individualmente para cada amostra, não sendo possível individualizá-los.

A Tabela II apresenta os hiperparâmetros explorados durante a busca em grade para os modelos MLP e RFR, adotando `random_state` fixo igual a 30, com o objetivo de garantir reprodutibilidade dos experimentos.

TABELA I
HIPERPARÂMETROS AVALIADOS PARA OS FILTROS ADAPTATIVOS

Filtro	Ordem	Parâmetro	Intervalo
RLS	[2, 4, 6]	Fator de esquecimento λ	0.79 – 0.99 (passo 0.02)
APA	[2, 4, 6]	Taxa de aprendizado μ	0.01 – 0.20 (passo 0.02)

TABELA II
HIPERPARÂMETROS AVALIADOS PARA OS MODELOS DE ML

Modelo	Parâmetro	Valores Testados
MLP	Neurônios (1ª camada)	[2, 4, 6, 8]
	Neurônios (2ª camada)	[4, 8]
	Função de ativação	ReLU, tanh
	Taxa de aprendizado	10^{-3} a 10^{-2}
RFR	Nº de estimadores	[30, 50, 70, 90, 110, 130, 150]
	Profundidade máxima	[1, 2, 3, 4, 5]

A busca em grade para o modelo MLP resultou na configuração de duas camadas escondidas com 4 e 8 neurônios, respectivamente, utilizando a função de ativação *ReLU* e uma taxa de aprendizado de 10^{-3} . Para o modelo RFR, o hiperparâmetro Nº de estimadores, que define o tamanho da floresta, foi definido como 50 e a profundidade máxima igual a 5.

D. Pós-processamento

Após o processamento de correção de viés, a variável da SST do BESM ajustada (padronizada) deve retornar ao seu domínio original por meio de

$$\hat{x}_i^{modelo} = x_i^{modelo} * \sigma^{ERA5} + \mu^{ERA5} \quad (3)$$

onde, \hat{x}_i^{modelo} é o valor da saída do modelo despadronizada, x_i^{modelo} é o valor da variável do modelo padronizada, μ^{ERA5} é a média da amostra ERA5 usada no treinamento e σ^{ERA5} é o desvio padrão da amostra ERA5 usada no treinamento do modelo.

A IDCT-2D (transformada inversa) deve ser aplicada caso o ajuste seja realizado no domínio da frequência retornando a SST ajustada ao seu domínio original.

E. Avaliação do modelo

Para avaliar o desempenho da melhoria na variável prognóstica SST do modelo BESM e quantificar a redução do erro, o MSE (*Mean Squared Error*), o MAE (*Mean Absolute Error*) e o RMSE (*Root Mean Squared Error*) [23] são calculados usando os dados observacionais do ERA5 como alvo.

Para avaliar o desempenho dos modelos, emprega-se a função de distribuição cumulativa (CDF) [18] das medidas de desempenho. A CDF fornece a probabilidade acumulada de que uma variável aleatória seja menor ou igual a um dado valor, i.e.:

$$F_X(x) = P(X \leq x) \quad (4)$$

onde $F_X(x)$ representa a probabilidade acumulada de X até o valor x . Com isso, é possível analisar a distribuição das medidas de desempenho e seus valores máximos.

V. RESULTADOS

Cada um dos modelos foi treinado com a variável SST dos modelos BESM e ERA5, utilizando o mesmo conjunto de dados, (período: nov/2007 a out/2008), representando o que se deseja corrigir e o valor de referência respectivamente. O modelo obtido é então aplicado às previsões do modelo BESM em período posterior (período: nov/2008 a out/2009) retornando o conjunto de previsões ajustadas. A Tabela III compara os índices de desempenho do modelo BESM relativamente ao ERA5 (última linha) com os índices após a aplicação das diferentes técnicas testadas para reduzir o erro sistemático da SST do BESM, nos domínios do tempo e da frequência, com e sem normalização dos dados de entrada.

Na Tabela III, é possível verificar que há uma efetiva redução do erro de previsão com as diferentes técnicas. Em particular, destacam-se os resultados obtidos com o pré-processamento dos dados via DCT-2D e normalização, empregando os modelos MLP e RFR, que apresentaram os menores valores de MAE, MSE e RMSE.

Visando analisar o desempenho em cada célula, e não somente o desempenho médio, analisa-se a persistência $P = (N_{melhoria}/N_{total}) \times 100\%$. Ela quantifica a proporção de células nas quais a correção é efetiva (isto é, o erro é menor que o do BESM). A Figura 3 apresenta as persistências para as quatro técnicas testadas quando se aplica o pré-processamento via DCT-2D e normalização. Observe que, em todos os casos, as

TABELA III
COMPARAÇÃO DE MÉTRICAS DOS MODELOS

Domínio	Método	MAE [K]	MSE [K ²]	RMSE [K]
Tempo	RLS	1.389	3.988	1.997
	APA	1.343	3.807	1.951
	MLP	1.147	2.439	1.562
	RFR	0.715	0.998	0.999
Tempo normalizado	RLS	1.258	4.215	2.053
	APA	1.165	3.626	1.904
	MLP	0.670	0.843	0.918
	RFR	0.713	0.976	0.988
Frequência	RLS	1.361	3.884	1.971
	APA	1.349	3.769	1.941
	MLP	1.136	2.409	1.552
	RFR	0.699	0.952	0.975
Frequência normalizada	RLS	0.947	1.921	1.386
	APA	0.947	1.921	1.386
	MLP	0.653	0.811	0.901
	RFR	0.695	0.925	0.962
Referência	BESM	1.429	4.099	2.024

persistências são maiores do que 50%, evidenciando a eficácia dos modelos na correção da previsão, com destaque para as técnicas MLP e RFR.

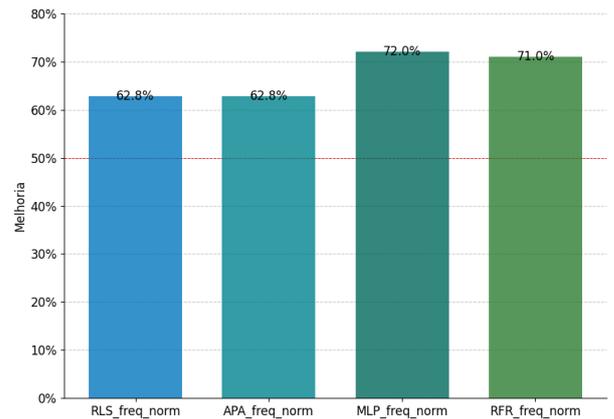


Fig. 3. Persistência do método: proporção de pontos no conjunto de dados onde o método aplicado oferece um desempenho superior ao BESM

A Figura 4 apresenta o resultado do método na amostra de coordenadas geográficas 52.50°S, 44.25°E, no período de novembro de 2008 a outubro de 2009, apresentando a SST do modelo BESM, juntamente com as previsões corrigidas usando os métodos MLP e RFR no domínio da frequência e a referência ERA5.

Para analisar se a melhoria pontual observada na Figura 4 é consistente (isto é, é observada em larga escala no conjunto das 500 amostras/células), opta-se pela apresentação das CDFs do erro absoluto. Tal análise é apresentada na Figura 5 considerando a aplicação das diferentes técnicas com processamento no domínio da frequência de dados normalizados. Observa-se que os erros absolutos da previsão do BESM ajustado estão concentrados em valores menores que os originais, apresentando melhor desempenho nos modelos MLP e RFR.

VI. CONCLUSÕES

Neste trabalho, foram aplicados filtros adaptativos AP e RLS, e os modelos de ML MLP e RFR com o objetivo

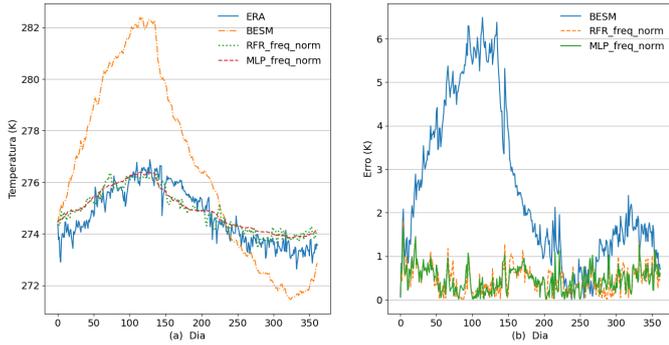


Fig. 4. (a) SST dos modelos ERA5, BESM e dos métodos RFR e MLP no domínio da frequência normalizada. (b) erro de previsão da SST, entre os modelos BESM, RFR, MLP e os dados do ERA5.

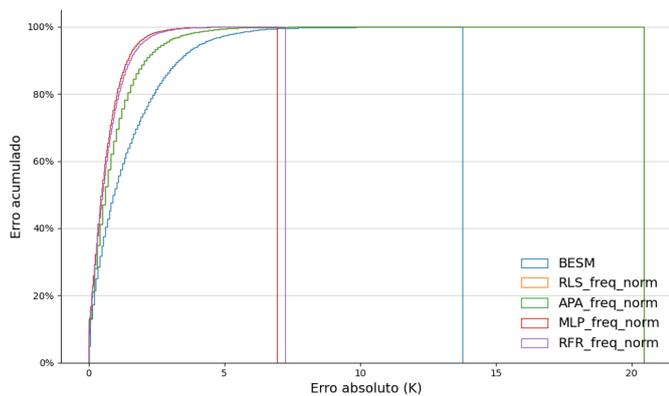


Fig. 5. Erro acumulado dos modelos com melhor desempenho em relação ao BESM

de reduzir o erro na previsão da SST do modelo BESM em relação à reanálise climática fornecida pelo ERA5. Cada um dos modelos foi utilizado para ajustar as previsões da variável prognóstica SST, sendo treinados com um conjunto de 500 amostras do BESM e utilizando o ERA5 como valor de referência, compreendendo o período de novembro de 2007 a outubro de 2008. O modelo pré-processado e ajustado foi então utilizado para corrigir as previsões do período subsequente, de novembro de 2008 a outubro de 2009. O desempenho foi avaliado utilizando métricas como o Erro Absoluto Médio e seu desvio padrão, o Erro Médio Quadrático e a Raiz do Erro Médio Quadrático. Os resultados demonstraram uma redução considerável no viés da variável SST do modelo BESM com a aplicação do método, com destaque para o MLP normalizado no domínio da frequência espacial (via DCT-2D). Os resultados indicam que a metodologia apresentada tem potencial para reduzir erros de previsão climática via pós-processamento e, assim, melhorar a acurácia dos prognósticos. A melhoria na acurácia da variável SST do modelo BESM permite, entre outros ganhos, melhores previsões de eventos climáticos extremos e clima, o que pode ajudar a evitar desastres e aumentar a produtividade de diferentes setores econômicos, gerando uma contribuição significativa para a sociedade.

Futuramente, pretende-se aplicar a proposta em outras variáveis climáticas e aumentar a quantidade de variáveis de entrada do modelo, considerando não somente maior tempo de obser-

vação mas diferentes variáveis prognósticas conjuntamente. Adicionalmente, experimentos estão sendo conduzidos com um conjunto de amostra maior e períodos temporais distintos.

REFERÊNCIAS

- [1] C. P. Weaver et al. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. *Wiley Interdisciplinary Reviews: Climate Change*, pages 467–476, 2012.
- [2] Intergovernmental Panel on Climate Change. Climate change 2022: Mitigation of climate change - summary for policymakers, 2022. <https://www.ipcc.ch/2022/04/04/ipcc-ar6-wgiii-pressrelease/>.
- [3] S. R. Weart. The idea of anthropogenic global climate change in the 20th century. *Wiley Interdisciplinary Reviews: Climate Change*, pages 67–81, 2010.
- [4] P. Hoffmann et al. Bias and variance correction of sea surface temperatures used for dynamical downscaling. *J. Geophys. Res. Atmos.*, 2016.
- [5] M. P. Tcheou et al. Reducing forecast errors of a regional climate model using adaptive filters. *Applied Sciences*, 11:8001, 2021.
- [6] H. H. Farfán et al. Correção de viés em previsão climática global da temperatura da superfície do mar por meio de filtros adaptativos no domínio da frequência. *Brazilian Journal of Climatology*, 2024.
- [7] S. Zhang et al. A machine learning bias correction on large-scale environment of high-impact weather systems in e3sm atmosphere model. *Journal of Advances in Modeling Earth Systems*, 2024.
- [8] P. S. R. Diniz. *Adaptive Filtering: Algorithms and Practical Implementation*. Springer, Cham, 5 edition, 2019.
- [9] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2 edition, 2009.
- [10] A. C. Müller and S. Guido. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, 2016.
- [11] S. F. Veiga et al. The brazilian earth system model ocean-atmosphere (besm-oa) version 2.5: evaluation of its cmip5 historical simulation. *Geoscientific Model Development*, 2019.
- [12] Copernicus Climate Change Service (C3S). What is the copernicus climate change service's era5 reanalysis dataset and what can it do for you?, 2024.
- [13] J. Shukla. Predictability in the midst of chaos: A scientific basis for climate forecasting. *Science*, 282:728–731, 1998.
- [14] H. Hersbach et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, pages 1999–2049, 2020.
- [15] J. D. Neelin et al. Tropical air-sea interaction in general circulation models. *Climate Dynamics*, 7:73–104, 1992.
- [16] P. Tavares et al. Um conjunto de dados de projeções de mudanças climáticas de alta resolução para a américa do sul com correção de viés. *Derbyana*, 45, 2024.
- [17] Copernicus Climate Change Service. What is bias correction?, 2021. <https://climate.copernicus.eu/sites/default/files/2021-01/infosheet7.pdf>.
- [18] K. M. Ramachandran and C. P. Tsokos. *Mathematical Statistics with Applications*. Academic Press, Tampa, FL, 2010.
- [19] D. S. Moore, G. P. McCabe, and B. A. Craig. *Introduction to the Practice of Statistics*. W. H. Freeman and Company, New York, 6 edition, 2009.
- [20] M.-Y. Kao, editor. *Encyclopedia of Algorithms*. Springer, 2008. <https://link.springer.com/referencework/10.1007/978-0-387-30162-4>.
- [21] P. S. R. Diniz, E. A. B. da Silva, and S. L. Netto. *Processamento digital de sinais: projeto e análise de sistemas*. Bookman, Porto Alegre, 2 edition, 2014.
- [22] C. W. Kok. Fast algorithm for computing discrete cosine transform. *IEEE Transactions on Signal Processing*, 45(3):757–760, 1997.
- [23] T. O. Hodson. Root mean square error (rmse) or mean absolute error (mae): when to use them or not. *Geoscientific Model Development*, 15:5481–5487, 2022.