PREVISÃO MENSAL DA VELOCIDADE DO VENTO NO SUDESTE DA AMÉRICA DO SUL: UMA ABORDAGEM BASEADA EM LSTM COM DADOS DE REANÁLISE ERA5

CLÁUDIO DE O. DUARTE[1]; JULIANA A. ANOCHI[2]; MARILIA H. SHIMIZU[2]; GERÔNIMO G. Z. LEMOS[2]; MARILTON S. DE AGUIAR[1]

[1]*Universidade Federal de Pelotas – {coduarte,marilton}@inf.ufpel.edu.br*
[2]*Instituto Nacional de Pesquisas Espaciais – {juliana.anochi,marilia.shimizu,geronimo.lemos}@inpe.br*

1. INTRODUÇÃO

O Acordo de Paris, adotado em 2015 e em vigor desde 2016, constitui o principal tratado internacional de combate às mudanças climáticas (SALAWITCH et al., 2017). Sucessor do Protocolo de Quioto, o acordo busca limitar a elevação da temperatura média global a menos de 2°C (GLANEMANN et al., 2020). Ainda assim, eventos meteorológicos extremos, como ondas de calor, tempestades, secas e grandes inundações, têm se tornado mais frequentes e intensos (STOTT et al., 2016). O desastre de 2024 no Rio Grande do Sul ilustra bem esse cenário: as inundações sem precedentes reforçam a urgência de ações de prevenção, adaptação e mitigação frente às mudanças climáticas projetadas (MARENGO et al., 2024). Tais fenômenos ameaçam não só os ecossistemas, mas também trazem impactos profundos à saúde humana, à economia e à segurança alimentar (LUSTED, 2017), consolidando a previsão climática como instrumento indispensável ao planejamento estratégico de diversos setores.

Nesse contexto, destaca-se a velocidade do vento como uma das variáveis climáticas de maior relevância prática, por sua influência direta na geração de energia eólica, na agricultura e na segurança do transporte aéreo e marítimo (DHIMAN; DEB, 2020). As previsões climáticas tradicionais empregam modelos matemáticos que representam o comportamento da atmosfera, um processo computacionalmente custoso e dependente de grande poder de processamento. Técnicas de aprendizado de máquina, como as redes *Long Short-Term Memory* (LSTM), dispensam a resolução explícita das equações físicas da atmosfera, aprendendo diretamente os padrões presentes nos dados históricos. Por serem capazes de captar dependências em séries temporais, essas redes são particularmente adequadas à modelagem de variáveis meteorológicas como a velocidade do vento, despontando como alternativa mais leve aos modelos numéricos tradicionais (HOCHREITER; SCHMIDHUBER, 1997).

O sudeste da América do Sul (SESA) configura-se como uma região de especial interesse para esse tipo de análise, dada sua relevância agrícola e energética e a necessidade de estudos que aprofundem a compreensão dos padrões de vento na área. Este trabalho tem, portanto, como objetivo desenvolver um modelo baseado em redes LSTM para a previsão mensal da velocidade do vento nessa região, avaliando sua capacidade de reproduzir os padrões espaciais e temporais observados.

Os resultados desta pesquisa devem contribuir para a compreensão dos padrões climáticos e para o aumento da confiabilidade da previsão do clima, beneficiando de modo mais direto os setores de energia eólica e agrícola, além dos recursos hídricos, da gestão de desastres e da sociedade em geral. Previsões mais acuradas e espacialmente detalhadas permitem um planejamento mais eficaz no uso de múltiplos recursos, atenuando os impactos de eventos

climáticos extremos, e abrem caminho para investigações futuras em climatologia e previsão do clima.

2. METODOLOGIA

O conjunto de dados utilizado no experimento foi o ERA5, quinta geração de dados de reanálise do ECMWF (*European Centre for Medium-Range Weather Forecasts*), do qual foi extraída apenas a variável velocidade do vento (HERSBACH et al., 2020). O *dataset* abrange 86 anos, de janeiro de 1940 a dezembro de 2025, totalizando 1032 meses de amostra, com resolução espacial de 0,25° x 0,25°. O domínio de estudo compreende o sudeste da América do Sul (SESA), delimitado pelas latitudes 14°S e 40°S e pelas longitudes 67°W e 43°W, incluindo Uruguai e Paraguai em sua totalidade, além de partes da Argentina, Bolívia e Brasil, conforme ilustrado na Figura 1. A região é composta por 105 latitudes e 97 longitudes, totalizando 10185 pontos de grade.

Figura 1 – Domínio de Estudo.



Foi treinada uma única rede LSTM para todos os pontos de grade simultaneamente, o que permite ao modelo aprender de forma conjunta os padrões espaciais e temporais do domínio. Os dados foram organizados em janelas deslizantes de três meses consecutivos para a entrada da rede, intervalo escolhido por corresponder à sazonalidade trimestral típica da região, definindo-se a velocidade do vento do mês imediatamente seguinte como o valor a ser previsto. Essa estrutura decorre do próprio funcionamento da LSTM, que aprende a partir de sequências, e não de observações isoladas. Os dados também foram normalizados com o *StandardScaler* (PEDREGOSA et al., 2011), ajustado apenas nos dados de treino e depois aplicado aos conjuntos de validação e teste, evitando assim o vazamento de informação durante o treinamento.

O particionamento seguiu a ordem cronológica, sem embaralhamento, de modo que observações futuras não influenciassem o aprendizado do modelo. O conjunto de teste corresponde às amostras de janeiro de 2022 a dezembro de 2025. Já o conjunto de treinamento e validação, referente ao período de 1940 a 2021, foi avaliado por validação cruzada com o *Time Series Split* (PEDREGOSA et al., 2011), dividido em três *folds* sucessivos: no primeiro, o treinamento abrangeu 04/1940 a 07/1960 e a validação, 10/1960 a 02/1981; no segundo, o treinamento foi de 04/1940 a 10/1980 e a validação, de 03/1981 a 07/2001; e no terceiro, o treinamento estendeu-se de 04/1940 a 05/2001, com validação de 08/2001 a 10/2021. Essa técnica foi escolhida por preservar a ordem cronológica da série temporal ao longo da busca, evitando que dados futuros fossem usados para validar previsões de períodos anteriores.

O modelo foi implementado em TensorFlow e Keras, com uma única camada LSTM e função de ativação tangente hiperbólica. O treinamento utilizou o otimizador Adam, a função de perda MSE (erro quadrático médio) e *early stopping*

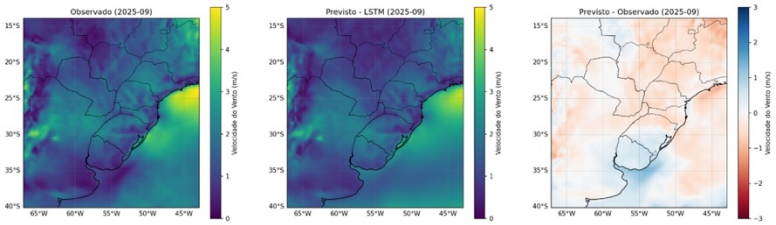
com paciência de 10 épocas, num total de até 100 épocas. A busca de hiperparâmetros foi realizada com a biblioteca Optuna, configurada para 20 *trials*, variando a taxa de aprendizado entre 0,0001 e 0,01 e o *batch size* entre 4 e 8, com o número de neurônios por camada mantido fixo em 32. A melhor combinação encontrada, avaliada nos três *folds* descritos acima, resultou em taxa de aprendizado de 0,0003965 e *batch size* de 8.

O desempenho do modelo foi avaliado por meio das métricas RMSE (raiz do erro quadrático médio) e MAE (erro absoluto médio). O processamento dos experimentos foi realizado em uma placa de vídeo NVIDIA GeForce GTX 1650 de 4GB.

3. RESULTADOS E DISCUSSÃO

O treinamento do modelo de previsão mensal mostrou que ele foi capaz de identificar os principais padrões climáticos do domínio, como a região dos Andes, a serra do Rio Grande do Sul e o litoral de Santa Catarina, Paraná e São Paulo. A Figura 2 ilustra esse comportamento para setembro de 2025: à esquerda, a velocidade observada; ao centro, a prevista pelo modelo LSTM; à direita, o erro de previsão (Previsto−Observado), concentrado sobre o oceano, sobretudo na costa sudeste do Brasil, e próximo de zero sobre o continente. Mesmo utilizando apenas a variável velocidade do vento como entrada, o modelo alcançou, no melhor conjunto de hiperparâmetros, RMSE de 0,6593 m/s e MAE de 0,4717 m/s, valores que indicam boa aderência às observações da reanálise ERA5.

Figura 2 – Comportamento previsto para setembro de 2025.



O desempenho variou de acordo com a estação do ano. Nos meses de outono e inverno austral (março, abril, maio, agosto e setembro) o modelo apresentou o melhor desempenho, com previsões mais próximas dos valores observados tanto sobre o continente quanto, em menor grau, sobre o oceano. Já nos meses de transição e verão, o desempenho foi mais heterogêneo: em fevereiro, verificou-se subestimação da velocidade do vento em todo o domínio; em junho e outubro, essa subestimação concentrou-se sobre o oceano; em julho ocorreu o padrão inverso, com bom desempenho sobre o continente e superestimação sobre o oceano; e novembro apresentou comportamento misto na região oceânica, com sub e superestimação ocorrendo simultaneamente. De modo geral, a área oceânica foi o principal ponto de dificuldade do modelo, independentemente da estação do ano.

Essa maior dificuldade sobre o oceano pode ser atribuída tanto à natureza dos dados quanto às características físicas do vento nessas áreas: a reanálise ERA5 conta com menor densidade de observações de superfície sobre o oceano, o que reduz a confiabilidade da própria referência usada para treinar e avaliar o modelo; o vento oceânico apresenta maior variabilidade e turbulência pela ausência de obstáculos físicos, dificultando a captura de padrões consistentes pela LSTM; e o conjunto de entrada não inclui variáveis de relevo, que auxiliam o modelo sobre o continente, mas não têm equivalente sobre o oceano.

Em conjunto, esses resultados indicam que o modelo foi capaz de reproduzir tanto os padrões espaciais quanto os temporais da velocidade do vento no SESA, cumprindo o objetivo proposto neste trabalho, com desempenho mais consistente sobre áreas continentais e maior incerteza sobre o oceano.

4. CONCLUSÕES

Este trabalho apresentou o desenvolvimento de um modelo de previsão mensal da velocidade do vento para o sudeste da América do Sul (SESA), baseado em uma única rede LSTM treinada simultaneamente para todos os pontos de grade do domínio, a partir de dados de reanálise do ERA5.

Os resultados indicam que o modelo foi capaz de reproduzir tanto os padrões espaciais quanto os temporais da velocidade do vento observados na região, cumprindo o objetivo proposto neste trabalho. No melhor conjunto de hiperparâmetros, o modelo alcançou RMSE de 0,6593 m/s e MAE de 0,4717 m/s, com desempenho mais consistente nos meses de outono e inverno austral. A principal limitação ocorreu na região oceânica, onde a velocidade do vento foi superestimada e subestimada com maior intensidade, provavelmente em razão da menor densidade de observações do ERA5 sobre o mar, da maior turbulência do vento oceânico e da ausência de variáveis de relevo no conjunto de entrada; sobre o continente, o desempenho foi consistentemente melhor.

Como principal contribuição, o trabalho evidencia a viabilidade do uso de uma única rede LSTM, treinada de forma conjunta para todo o domínio espacial, na previsão mensal da velocidade do vento no SESA. Como perspectivas de continuidade, sugere-se a inclusão de variáveis relacionadas ao relevo e à rugosidade da superfície, além da avaliação de fontes de dados com maior densidade de observações sobre o oceano, como caminhos possíveis para reduzir as limitações identificadas e ampliar a robustez das previsões.

5. REFERÊNCIAS BIBLIOGRÁFICAS

DHIMAN, H. S.; DEB, D. A Review of Wind Speed and Wind Power Forecasting Techniques. **arXiv**, 2020. Disponível em: https://arxiv.org/abs/2009.02279.

GLANEMANN, N.; WILLNER, S. N.; LEVERMANN, A. Paris Climate Agreement passes the cost-benefit test. **Nature Communications**, v. 11, n. 1, 2020.

HERSBACH, H. et al. The ERA5 global reanalysis. **Quarterly Journal of the Royal Meteorological Society**, v.146, n.730, p.1999-2049, 2020.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory**. Neural Computation**, v.9, n.8, p.1735-1780, 1997.

LUSTED. **Extreme Weather Events.** Nova York: Greenhaven Publishing LLC, 2017.

MARENGO, J.A. et al. O maior desastre climático do Brasil: chuvas e inundações no estado do Rio Grande do Sul em abril-maio 2024. **Estudos Avançados**, v.38, n.112, p.203-228, 2024.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. **Journal of Machine Learning Research**, v.12, p.2825-2830, 2011.

SALAWITCH, R.J. et al. **Paris climate agreement: Beacon of hope.** Springer Nature, 2017.

STOTT, P.A. et al. Attribution of extreme weather and climate-related events. **WIREs Climate Change**, v.7, n.1, p.23-41, 2016.