

Previsão Futura de Séries Temporais Meteorológicas no Cais do Porto de Porto Alegre Usando Aprendizado Profundo com Redes Neurais LSTM

Lucas dos Santos Azevedo¹, Rafael Manica², Eduardo Puhl²

¹Instituto de Pesquisas Hidráulicas – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.029 – 91501-970 – Porto Alegre – RS – Brazil Porto Alegre – RS – Brazil

lucas.azevedo@inf.ufrgs.br, manica@iph.ufrgs.br, eduardo.puhl@ufrgs.br

Abstract. *This study evaluates an LSTM model for intraday forecasting (up to 100 hours ahead) of meteorological variables at the Porto Alegre harbor. The model was trained with 100-lag input windows using a combined dataset of API-REDEMETS/SBPA–SBPK records (306,250 entries; 2003–2024) and in situ harbor measurements (12,589 entries; 2021–2024). In benchmark tests, the LSTM outperformed the Tomorrow.io API (RMSE 11.93 vs. 22.37; MAE 9.48 vs. 20.59) and WRF (Jan 2022) across all variables. The results demonstrate the feasibility and advantage of locally tuned LSTM models for short- to medium-term forecasting in port environments, offering a reliable alternative to commercial approaches.*

Resumo. *Este estudo avalia um modelo LSTM para previsão intradiária (até 100 horas à frente) de variáveis meteorológicas no cais do porto de Porto Alegre. O modelo foi treinado com janelas de 100 lags, utilizando um conjunto de 306.250 dados da API-REDEMETS/SBPA–SBPK (2003–2024) e 12.589 medições in loco no porto (2021–2024). Nos testes comparativos, o LSTM superou a API Tomorrow.io (RMSE 11,93 vs. 22,37; MAE 9,48 vs. 20,59) e o WRF (jan/2022) em todas as variáveis analisadas. Os resultados demonstram a viabilidade e a vantagem de modelos LSTM ajustados localmente para previsões de curto e médio prazo em ambientes portuários, configurando-se como alternativa confiável a abordagens comerciais.*

1. Introdução

O transporte hidroviário constitui um eixo estratégico da infraestrutura logística do sul do Brasil e um vetor de desenvolvimento econômico, integração regional e competitividade internacional do Rio Grande do Sul. Esse sistema é apoiado por uma malha que abrange os rios Guaíba, Jacuí, Taquari, Caí, Sinos e Gravataí, além das lagoas dos Patos e Mirim [PortosRS 2023b, de Mendonça and Ballestrin 2021].

No centro dessa rede, o porto de Porto Alegre atua como *hub* fluvio-marítimo, conectando cadeias produtivas aos mercados externos — em especial via Porto de Rio Grande e o Atlântico [Cortés 2012, Nascimento et al. 2018, PortosRS 2023a]. Contudo, a plena eficiência operacional está condicionada à variabilidade meteorológica e a eventos adversos, como tempestades, nevoeiros e flutuações de nível d’água, que impactam a segurança, a regularidade e o desempenho logístico.

Nesse contexto, a integração de dados meteorológicos confiáveis e de sistemas avançados de previsão torna-se crucial para o planejamento e a tomada de decisão portuária, potencializada pelo avanço em sensoriamento, automação de estações e modelagem numérica, que elevam a precisão e o alcance das previsões [Bauer et al. 2015, Peterson and Vose 1997, Vose et al. 2012].

Motivado por esse cenário aplicado, este estudo investiga modelos de aprendizado profundo para previsão intradiária de variáveis meteorológicas — tais como temperatura, ponto de orvalho, pressão, umidade relativa e vento — no ambiente portuário de Porto Alegre, com vistas a mitigar riscos e sustentar a eficiência operacional.

2. Metodologia

Este estudo empregou Redes Neurais Recorrentes do tipo *Long Short-Term Memory* (LSTM) para previsão de variáveis meteorológicas na região do cais do porto de Porto Alegre. Foram utilizadas duas bases complementares:

[(i)]

1. dados da API da REDEMET referentes às estações SBPA (Porto Alegre) e SBPK (Pelotas), totalizando 306.250 registros horários entre janeiro de 2003 e agosto de 2024, contemplando direção e velocidade do vento, pressão atmosférica ao nível da estação, umidade relativa, ponto de orvalho e temperatura de superfície;
2. dados coletados *in loco* no cais do porto, fornecidos pelo Programa de Gestão Ambiental do Porto de Porto Alegre (PGA-POA), compreendendo 12.589 registros horários distribuídos em seis variáveis, correspondentes a cerca de 380 dias no período de outubro de 2021 a março de 2024.

A integração dessas duas fontes permitiu combinar a representatividade local das medições do cais com a robustez temporal das séries históricas da API-REDEMET, assegurando consistência espacial e estatística para a modelagem preditiva.

A escolha do modelo LSTM decorreu de uma comparação inicial com diferentes abordagens, incluindo o *Multilayer Perceptron* (MLP), um modelo *naive* baseado na maior frequência observada e a regressão logística multivariada. O LSTM apresentou desempenho superior na captura das dependências temporais de longo prazo, sendo selecionado como abordagem final para o desenvolvimento preditivo.

A estruturação dos dados considerou janelas deslizantes de cem registros horários passados (*lags*) para a previsão de cem horas futuras (*leads*), respeitando a natureza sequencial das séries temporais e preservando a integridade espacial. Do total de 318.839 registros, 309.839 foram utilizados para o treinamento e 9.000 registros (equivalentes a pouco mais de um ano) foram destinados à validação.

A definição dos melhores hiperparâmetros foi realizada por meio de busca em grade (*grid search*), o que permitiu identificar a configuração mais eficiente. O modelo final foi estruturado com uma camada de entrada contendo 256 neurônios, seguida de sete camadas ocultas do tipo LSTM, cada uma associada a uma camada de *dropout* com taxa de 0,1, e finalizada por uma camada densa de saída com ativação linear. Utilizou-se o otimizador Adam e a função de perda baseada no erro quadrático médio (MSE). O treinamento foi conduzido com até 500 épocas, com emprego de técnicas de regularização por meio de *callbacks*, como *EarlyStopping*, *ReduceLROnPlateau* e *ModelCheckpoint*,

a fim de evitar sobreajuste e assegurar a estabilidade do processo de aprendizado. As métricas utilizadas na avaliação incluíram o erro médio absoluto (MAE), o erro quadrático médio (MSE) e a raiz do erro quadrático médio (RMSE).

A validação do modelo foi realizada em três etapas complementares. Inicialmente, o desempenho foi aferido em relação aos próprios dados de referência, provenientes das estações SBPA, SBPK e das medições locais do cais do porto, assegurando consistência interna do modelo em relação às séries utilizadas no treinamento e validação. Em seguida, as previsões do LSTM foram comparadas com aquelas disponibilizadas pela API Tomorrow.io, a partir de simulações de cem horas futuras para a região de Pelotas (SBPK), onde havia dados atualizados após o desastre climático de 2024, diferentemente da estação de Porto Alegre. Por fim, o desempenho do modelo foi comparado às simulações geradas pelo modelo atmosférico WRF, fornecido pelo PGA-POA. Nesta etapa, foram analisadas 168 simulações referentes ao período de 1º a 16 de janeiro de 2022, confrontando as previsões do LSTM e do WRF com medições obtidas na estação do cais do porto de Porto Alegre.

3. Resultados

A Tabela 1 consolida os resultados das três fases do estudo, com as variáveis dispostas em colunas (cada uma contendo RMSE/MAE) e as métricas organizadas por fase em linhas.

Na **Fase 1** (LSTM puro), foram reportados os erros desnormalizados para as *leads* 1 e 100. Os resultados mostram o padrão físico esperado de degradação com o horizonte: os valores de RMSE/MAE aumentam da lead 1 para a lead 100, especialmente para a temperatura (por exemplo, 2,56/1,97 °C para 4,84/3,74 °C) e o ponto de orvalho (2,26/1,73 °C para 4,43/3,50 °C). Para a pressão atmosférica, a degradação é mais contida (4,02/3,06 hPa e 5,92/4,64 hPa), refletindo a estabilidade barométrica. Já a umidade relativa praticamente dobra de valor (10,33/8,04% para 17,20/13,45%), assim como a velocidade do vento (2,75/2,21 para 4,00/2,88 nós), evidenciando persistência temporal. A direção do vento revelou-se a variável mais desafiadora, com erros altos em ambos os horizontes (124,81/92,52 para 135,47/109,62 graus), devido ao caráter angular e à variabilidade acentuada (módulo 360°).

Na **Fase 2** (estação SBPK), compararam-se o desempenho do LSTM e da API Tomorrow.io nas leads 1 e 99. A parametrização local do LSTM mostrou vantagem ampla sobre o Tomorrow.io em três das seis variáveis (velocidade do vento, direção do vento e umidade relativa). Para as variáveis temperatura e, principalmente, pressão na lead 99, observaram-se exceções, com o Tomorrow.io apresentando erros próximos ou menores (por exemplo, pressão 8/6 vs. 4/4). Esse comportamento pode ser atribuído à assimilação de preditores exógenos e às diferenças de latência. Ainda assim, no agregado do estudo (médias globais), o LSTM apresentou RMSE de 11,93 e MAE de 9,48, valores melhores que os 22,37 e 20,59 do Tomorrow.io, confirmando o *fit* regional superior (essas médias são apresentadas apenas como comparação sumária, por envolverem unidades distintas).

Tabela 1. Resultados consolidados das três fases (variáveis em colunas; métricas/fases em linhas).

Fase / Cenário	Temp. (°C)		Pto. orvalho (°C)		Pressão (hPa)		Umidade (%)		Vel. vento (nós)		Dir. vento (graus)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Fase 1 — LSTM, lead 1	2,56	1,97	2,26	1,73	4,02	3,06	10,33	8,04	2,75	2,21	124,81	92,52
Fase 1 — LSTM, lead 100	4,84	3,74	4,43	3,50	5,92	4,64	17,20	13,45	4,00	2,88	135,47	109,62
Fase 2 — LSTM, lead 1	2	2	2	2	7	6	4	3	3	2	14	10
Fase 2 — Tomorrow.io, lead 1	6	6	10	8	14	12	23	20	5	5	89	88
Fase 2 — LSTM, lead 99	4	3	3	3	8	6	3	2	5	4	6	5
Fase 2 — Tomorrow.io, lead 99	3	3	3	3	4	4	26	25	7	6	44	38
Fase 3 — LSTM (média, 168 sims)	1,18	0,90	1,08	0,84	2,53	2,07	5,40	4,05	1,86	1,45	101,35	66,73
Fase 3 — WRF (média, 168 sims)	1,35	0,96	1,75	1,32	3,90	3,81	8,98	7,36	2,72	2,29	107,03	73,08

Na **Fase 3** (SBPA/POA, 01–16 jan 2022), foram analisadas as médias de RMSE e MAE em 168 simulações (cada uma com 100 *leads*). O LSTM superou o modelo atmosférico WRF em todas as variáveis: temperatura (1,18/0,90 vs. 1,35/0,96), ponto de orvalho (1,08/0,84 vs. 1,75/1,32), umidade relativa (5,40/4,05 vs. 8,98/7,36), pressão (2,53/2,07 vs. 3,90/3,81), velocidade do vento (1,86/1,45 vs. 2,72/2,29) e direção do vento (101,35/66,73 vs. 107,03/73,08). Esses resultados sugerem que, para horizontes de até 100 horas no cais de Porto Alegre, o aprendizado baseado em séries locais oferece vantagem operacional frente à modelagem físico-numérica geral.

4. Conclusões

Os resultados demonstram a viabilidade e a vantagem de modelos LSTM ajustados localmente para previsões de curto e médio prazo em ambientes portuários, configurando-se como uma alternativa confiável a abordagens comerciais.

Referências

- Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55.
- Cortés, A. O. (2012). The port of barranquilla: Challenges and recommendations.
- de Mendonça, D. and Ballestrin, L. (2021). Dossiê ocupar desde o sul? *Revista Sul Americana de Ciência Política*, 1(2):32.
- Nascimento, A., Ribeiro, H., and Wahrhaftig, A. (2018). Comparative analysis of port performances between italy and brazil.
- Peterson, T. C. and Vose, R. S. (1997). An overview of the global historical climatology network temperature database. *Bulletin of the American Meteorological Society*, 78(12):2837–2849.
- PortosRS (2023a). Custos e eficiência do transporte hidroviário no rio grande do sul. <https://www.portosrs.com.br>.
- PortosRS (2023b). Principais hidrovias e sua importância para o transporte de cargas. <https://www.portors.gov.br>.
- Vose, R. S. et al. (2012). Monitoring and understanding changes in heatwaves, cold waves, floods, and droughts in the united states: State of knowledge. *Bulletin of the American Meteorological Society*, 94(6):821–834.