

## XXVI SIMPÓSIO BRASILEIRO DE RECURSOS HIDRÍCOS

### **APLICAÇÃO DA REDE NEURAL LSTM PARA PREDIÇÃO DE NÍVEL DO RIO CAÍ NO RIO GRANDE DO SUL**

*Luciana da Silva Mieres<sup>1</sup>; Arthur da Fontoura Tschiedel<sup>2</sup>; Elaine Souza<sup>3</sup> & Claudineia Brazil<sup>4</sup>*

**Abstract:** The increasing frequency and intensity of extreme natural events, intensified by climate change, have caused serious socioeconomic damage, as exemplified by the recent floods in Rio Grande do Sul. In this context, investing in the disaster prevention phase becomes essential, particularly through the use of technologies such as Artificial Intelligence. In light of this, this study evaluates the application of the Long Short-Term Memory (LSTM) recurrent neural network to predict the water level of the Caí River in São Sebastião do Caí/RS, a municipality historically affected by floods. For this purpose, a 32-year time series of daily precipitation and river level data provided by the Brazilian National Water Agency (ANA) was used. The study was divided into stages, including data cleaning and processing, splitting into training (70%) and testing (30%) sets, normalization of these sets, and adaptation to the format required by the LSTM model. The training phase involved searching for the best hyperparameters, and prediction performance was evaluated using metrics such as MSE, RMSE, and NSE, demonstrating good results. However, it became clear that improved performance is linked to the need for an expanded dataset with a greater number of high river level samples. Therefore, it is proposed to enhance the historical series with indirect data, such as those obtained through remote sensing, to improve the model's ability to predict critical events. In summary, the LSTM model proved effective in forecasting river levels, showing potential to support early warning issuance and disaster mitigation.

**Resumo:** O aumento da frequência e intensidade de eventos naturais extremos, intensificado pelas mudanças climáticas, tem causado sérios danos socioeconômicos, como exemplificado pelas recentes inundações ocorridas no Rio Grande do Sul. Nesse contexto, torna-se fundamental investir na fase de prevenção de desastres, especialmente por meio do uso de tecnologias como a Inteligência Artificial. Diante disso, este estudo avalia a aplicação da rede neural recorrente *Long Short-Term Memory* (LSTM) para prever o nível do rio Caí, em São Sebastião do Caí/RS, município historicamente afetado por inundações. Para tanto, utilizou-se uma série temporal de 32 anos de dados diários de precipitação e nível do rio, fornecidos pela Agência Nacional de Águas (ANA). O estudo foi subdividido em etapas que consistiram na limpeza e tratamento dos dados, na subdivisão em conjunto de treinamento (70%) e teste (30%), na normalização desses conjuntos e na adaptação ao formato exigido pelo modelo LSTM. O treinamento envolveu a busca pelos melhores hiperparâmetros e a predição foi avaliada por métricas como MSE, RMSE e NSE, demonstrando bom desempenho. Entretanto, ficou evidenciado que melhores resultados estão relacionados à necessidade de ampliação do conjunto de dados que deve ser enriquecido com um número maior de amostras de

<sup>1</sup> Mestre em Sensoriamento Remoto (UFRGS) - Secretaria de Planejamento, Governança e Gestão do Rio Grande do Sul (SPGG-RS). Porto Alegre, RS - mieres.academico@gmail.com

<sup>2</sup> Doutor em Recursos Hídricos e Saneamento (UFRGS) – Water Flow Engenharia – Porto Alegre, RS – arthur.tschiedel@waterflow.eng.br

<sup>3</sup> Doutora em Ciências da Computação e Matemática Computacional (USP) – Instituto de Ciências Matemáticas e Computação - Universidade de São Paulo – São Carlos, SP – parros@icmc.usp.br

<sup>4</sup> Doutora em Recursos Hídricos e Saneamento (UFRGS) – Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia - Universidade Federal do Rio Grande do Sul – Porto Alegre, RS – neiabrazil@yahoo.com.br

níveis máximos do rio. Propõe-se, assim, ampliar a série histórica com dados indiretos, como os obtidos por sensoriamento remoto, para melhorar a capacidade do modelo em prever eventos críticos. Em resumo, a LSTM mostrou-se eficaz na previsão de níveis fluviais, com potencial para contribuir na emissão de alertas antecipados e mitigação de desastres.

**Palavras-Chave** – Inundações, Rede neural, LSTM

## INTRODUÇÃO:

Desastres, sejam de origem natural ou antropogênica, afetam países em todo o mundo, mas é nos países pobres e em desenvolvimento que os impactos tendem a ser mais severos, devido a fatores como infraestrutura precária, fragilidade institucional e desigualdade social. O relatório *World Risk Report* (2023) aponta que a vulnerabilidade de uma sociedade está diretamente ligada à sua exposição aos perigos e à sua capacidade de enfrentamento. Nesse contexto, a compreensão dos conceitos de perigo, vulnerabilidade e risco torna-se fundamental. Perigo refere-se ao potencial de um fenômeno causar danos; vulnerabilidade é a fragilidade da população frente ao evento; e risco é o produto entre ambos.

A classificação dos desastres é feita em naturais, tecnológicos e híbridos, sendo os híbridos causados por interações entre fenômenos naturais e ações humanas. No Brasil, a Secretaria Nacional de Proteção e Defesa Civil padronizou essa classificação através da COBRADE, que segmenta os desastres naturais em cinco categorias: geológicos, hidrológicos, meteorológicos, climatológicos e biológicos. O foco deste estudo recai sobre os desastres hidrológicos, especificamente as inundações, dada sua relevância em termos de impacto humano e frequência no território nacional. Entre 1991 e 2022, o Brasil registrou mais de 57 mil desastres naturais, sendo 28,4% do tipo hidrológico. As inundações destacaram-se pelo elevado número de desalojados e desabrigados (mais de 3,6 milhões de pessoas) e pelo total de 20,15 milhões de afetados. O estado do Rio Grande do Sul, notadamente o município de São Sebastião do Caí, tem sido um dos mais atingidos, em especial por inundações associadas à bacia hidrográfica do Rio Caí. Esse histórico, somado à disponibilidade de dados, justificou a escolha do município como área de estudo.

A gestão de desastres naturais no Brasil tem se estruturado por meio de políticas públicas voltadas às diferentes etapas do ciclo de proteção e defesa civil: prevenção, mitigação, preparação, resposta e recuperação. Dentre essas, a prevenção tem ganhado destaque como estratégia essencial para minimizar danos humanos e materiais. O governo estadual do Rio Grande do Sul lançou uma publicação dedicada ao mapeamento de desastres naturais, enquanto o governo federal instituiu o Atlas de Desastres no Brasil, uma plataforma digital que consolida dados para planejamento de ações preventivas.

Destaca-se que, quanto à previsão de inundações, tradicionalmente utiliza-se modelos hidrológicos, que são representações matemáticas do ciclo da água. Esses modelos, como os discutidos por Fagundes (2021) e Brunner et al. (2021), requerem variáveis como precipitação, evapotranspiração, infiltração e escoamento, podendo utilizar dados de sensoriamento remoto e modelos digitais de elevação. Apesar de eficazes, esses modelos demandam alta complexidade e calibração, o que pode dificultar sua aplicação em larga escala, especialmente em contextos em que não há capacidade técnica e institucional para desenvolvimento de tal modelagem. Como alternativa, surgem abordagens mais simplificadas baseadas em inteligência artificial (IA), incluindo técnicas de aprendizado de máquina (*machine learning*) e aprendizado profundo (*deep learning*). Estudos recentes vêm demonstrando que essas técnicas podem ser promissoras para a previsão de inundações, oferecendo soluções ágeis e de menor complexidade operacional.

Diversas aplicações de Inteligência Artificial (IA) foram desenvolvidas com foco em desastres hidrológicos. Sambati et al. (2019) utilizaram a técnica KNN para prever alagamentos na Região Metropolitana de São Paulo, incorporando dados de precipitação, descargas elétricas, mapas de susceptibilidade e informações extraídas de redes sociais. Já Schimdt et al. (2021) propuseram o ClimateGam, que utiliza redes neurais adversárias para simular visualmente como locais urbanos seriam afetados por inundações, com o intuito de conscientizar a população sobre os riscos relacionados às mudanças climáticas. Outro destaque é o sistema de alerta de inundações em tempo real desenvolvido pela empresa Google, em parceria com o Serviço Geológico do Brasil, que utiliza IA para fornecer alertas por meio da aplicação Google Maps a mais de 60 localidades brasileiras. Essas iniciativas demonstram o potencial da IA em substituir, ou ao menos complementar, os modelos hidrológicos tradicionais.

O estudo em questão investiga se é possível prever as cotas de inundação no município de São Sebastião do Caí por meio da modelagem de dados de precipitação e nível fluviométrico utilizando Redes Neurais Recorrentes (RNN), especificamente a arquitetura LSTM (*Long Short-Term Memory*). Os resultados indicam que a técnica é eficaz, oferecendo simplicidade operacional, sem a necessidade de calibração de variáveis hidrológicas complexas. Assim, o uso de LSTM representa uma alternativa viável e eficiente para a previsão de inundações, contribuindo significativamente para estratégias de prevenção e mitigação de desastres naturais.

## ÁREA DE ESTUDO:

O Rio Grande do Sul é um dos Estados brasileiros mais atingidos por desastres naturais, sobretudo os de natureza hidrometeorológica. Recentemente, foi acometido pelo maior desastre dessa natureza e de sua totalidade de municípios, 452 decretaram situação de calamidade pública ou de emergência, representando 90,9% desse total (Rio Grande do Sul, 2024). Um município que se destaca em relação aos demais por seu histórico em ocorrência de desastres do tipo inundação é São Sebastião do Caí, localizado na porção leste do território estadual, especificamente na bacia hidrográfica do Rio Caí. Adicionalmente, ressalta-se ao longo do período analisado, foi um dos municípios que mais registrou danos humanos devido a inundações. Diante disso e considerando o monitoramento dessa bacia hidrográfica com disponibilidade de dados hidrometeorológicos, o presente estudo concentrou sua análise nesse município. A Figura 1 ilustra sua localização e inserção na bacia hidrográfica do rio Caí, bem como a identificação das estações pluviométricas e fluviométrica utilizadas para consolidação da base de dados.

Figura 1 – Localização da área de estudo



Fonte: ANA, IBGE e SEMA-RS – Elaborado pelos autores

## MATERIAIS E MÉTODOS:

Primeiramente convém destacar que o presente estudo foi desenvolvido em cinco diferentes etapas, quais sejam: 1) Avaliação das estações pluviométricas e fluviométrica, consolidando a base de dados; 2) Análise exploratória dos dados, buscando principalmente a identificação de *outliers*; 3) Tratamento e transformação desses dados; 4) Desenvolvimento do modelo de predição baseado na arquitetura de rede neural recorrente LSTM e sua aplicação ao grupo de treinamento e de teste; 5) Avaliação final dos resultados a partir do emprego de medidas de acurácia.

### **Etapas 1: Consolidação Da Base De Dados**

Para a composição da base de dados utilizada neste estudo, foram considerados registros de precipitação de quatro estações pluviométricas e dados de nível do rio Cai de uma estação fluviométrica, todos obtidos no portal Hidroweb da Agência Nacional de Águas (ANA). Embora as séries dessas estações abranjam o período de 1950 a 2024, a ocorrência de falhas operacionais gerou lacunas significativas nos dados. Com base na análise da disponibilidade temporal e considerando a recomendação de Abreu et al. (2018) para uso de séries com no máximo 10% de falhas anuais, foram selecionadas as estações que apresentaram ao menos 90% de completude, visando reduzir a necessidade de tratamento de dados ausentes. Foram escolhidas, assim, as estações pluviométricas identificadas pelos códigos 2950019, 2950033, 2951022 e 2951027, e a estação fluviométrica até o ano de 2004, período em que não apresentou falhas. Com o intuito de garantir a consistência da série para calibração do modelo LSTM e minimizar vieses, definiu-se o recorte temporal de 1970 a 2004, totalizando 35 anos de dados diários, conforme recomendado pela OMM para estudos climáticos. A representatividade desse intervalo foi validada por meio da comparação com a normal climatológica do Instituto Nacional de Meteorologia (INMET) referente ao período 1991 a 2020, da estação de Caxias do Sul, próxima à área de estudo, demonstrando compatibilidade nas médias mensais. A



completude da série adotada é ilustrada na Figura 2 sendo que, para trabalhos futuros, recomenda-se testar o modelo com dados mais recentes obtidos por fontes indiretas, como sensores remotos ou combinações de estimativas (ensemble).

Figura 2. Disponibilidade temporal de dados das estações pluviométricas e fluviométrica (% ano) para o período de 1970 a 2004 – 35 anos de dados diários de precipitação e de nível

		Estação	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Precipitação	2950019	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
	2950033	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	
	2951022	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	93	100	100	100	100	100	100	
	2951027	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	92	100	100	72	100	100	100	100	100	100	100	100	100	100	98	99	
Nível	87170000	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Fonte: Elaborado pelos autores

## Etapas 2: Análise Exploratória

A análise do comportamento da precipitação e do nível fluviométrico na bacia hidrográfica do rio Caí ao longo do período de 1970 a 2004 revela que a região apresenta volumes anuais acumulados de chuva superiores a 1.000 mm, com predominância de faixas entre 1.400 e 1.600 mm em três das estações, e entre 1.600 e 2.000 mm na estação 2950033. Registros de chuvas diárias máximas anuais acima de 80 mm são frequentes, com destaque para as estações 2951022 e 2951027, que frequentemente superam os 100 mm. Em relação ao nível do rio Caí, observou-se que em 32 dos 35 anos analisados, os níveis máximos anuais ultrapassaram os 10 metros, com predominância entre 10 e 12 metros. A distribuição mensal da precipitação mostra maiores volumes nos meses de junho, julho e outubro, seguidos por janeiro e fevereiro, enquanto os níveis máximos do rio concentram-se entre julho e outubro, além de dezembro e fevereiro, indicando uma correlação entre os períodos de maior precipitação e elevação do nível do rio.

Com o objetivo de compreender a distribuição dos dados de precipitação máxima anual e nível máximo anual nas estações analisadas, foram calculadas medidas estatísticas descritivas, como média, mediana, desvio padrão, coeficiente de variação, obliquidade e curtose, além de testes de normalidade. Os resultados revelaram distribuições assimétricas: a precipitação apresentou assimetria à direita e maior dispersão, com coeficiente de variação acima de 20%, enquanto os dados de nível mostraram assimetria à esquerda e menor variação, em torno de 10,5%. A curtose indicou que os dados de precipitação são mais concentrados que os de nível, os quais demonstraram tendência à normalidade. O teste de *Shapiro-Wilk* confirmou que apenas os dados de nível seguem distribuição normal, ao passo que os dados de precipitação não, o que é comum em séries de máximos que se ajustam melhor à distribuição de *Gumbel*. Também foram identificados *outliers* em todos os conjuntos, sendo que os dados de nível apresentaram valores extremos abaixo do limite inferior. Por fim, a análise de correlação indicou relação positiva entre precipitação e nível, com destaque para a estação 2950019, que apresentou a correlação mais forte com os dados de nível. A Tabela 1 apresenta a estatística descritiva calculada para o conjunto de dados.

Tabela 1. Resumo da análise exploratória dos dados de precipitação e nível máximos do rio Caí

Estação	Mínimo	1º Quartil	Mediana	Média	3º Quartil	Máximo	Desvio Padrão	Coeficiente de variação (%)	p-valor	Obliquidade (skewness)	Curtose
2950019	38,00	72,10	79,60	82,85	93,00	129,40	22,04	26,61	0,0460	0,4646	3,1882
2951022	60,50	74,55	91,00	93,16	102,25	168,70	24,29	26,08	0,0016	1,3031	4,4917
2951027	50,20	73,35	92,40	98,22	114,70	217,30	32,54	33,13	0,0015	1,4863	6,1855
2950033	51,50	69,75	75,30	81,32	90,50	159,30	21,57	26,53	0,0004	1,6645	6,5286
87170000	825,00	1.099,00	1.171,00	1.163,77	1.248,00	1.374,00	125,96	10,52	0,5821	-0,4235	3,0594

Fonte: Elaborado pelos autores

### Etapa 3: Tratamento dos Dados

Em relação ao conjunto de dados, destaca-se que o número de dados faltantes tem baixa representatividade em relação ao total de cada um dos conjuntos, sendo encontrado em duas das 4 estações pluviométricas, representando menos de 0,5% na primeira e 1,1% na segunda. Na estação 2951022 os dados ausentes correspondem a dias com falhas no registro de chuva no mês de março de 1997. Já na estação 2951027 apresentou maior número de dias com falhas, ocorridas em dezembro (30 dias) de 1988, em abril (30), maio (30), setembro (30) e outubro (14) de 1991, setembro (3) e outubro (4) de 2003 e janeiro (2) e fevereiro (1) de 2004. Buscando estabelecer uma série temporal mais completa, realizou-se a atribuição dos valores ausentes por valores que pudessem representar a precipitação provável no dia. Para tanto, aplicou-se uma técnica de interpolação utilizada na hidrologia, a qual é baseada no cálculo da média ponderada pelo inverso da distância, em que se considera que a precipitação em uma estação pluviométrica pode ser calculada como a média ponderada das precipitações registradas em estações próximas (Collischonn e Dornelles, 2013). Para tanto foi necessário acessar os dados de uma nova estação, próxima tanto à estação 2951022 quanto à 2951027. A escolha dessa estação foi baseada na completude de dados no período necessário para a correção dos dados faltantes.

Para atribuição dos valores de precipitação aplicou-se a Equação 1, baseada em Colischon e Dorneles (2013), onde  $P_m$  é a chuva no dia,  $NP$  é o número de postos pluviométricos com dados disponíveis,  $P_j$  é a chuva observada na estação  $j$  e  $b$  um expoente igual a 2, o que identifica o método como interpolação ponderada pelo inverso da distância ao quadrado. A partir da aplicação da técnica de interpolação, uma nova base de dados foi obtida, sem valores ausentes para dados de chuva.

$$P_m = \frac{\sum_{j=1}^{NP} \frac{P_j}{(d_{ij})^b}}{\sum_{j=1}^{NP} \frac{1}{(d_{ij})^b}} \quad (1)$$

### Etapa 4: Desenvolvimento do Modelo LSTM

Para o desenvolvimento do modelo proposto, utilizou-se o ambiente *Google Colaboratory*, escolhido pela facilidade de acesso a recursos computacionais com GPU e compatibilidade com a linguagem Python. Antes da modelagem, foi realizado o pré-processamento dos dados, etapa essencial que incluiu a organização dos dados em um único arquivo (.CSV) com colunas de precipitação (em milímetros) e nível (em centímetros), totalizando 12.784 registros e 63.920 valores. Esse arquivo foi armazenado no *Google Drive* para integração com o *Colaboratory*. Em seguida, os dados foram divididos sequencialmente em conjuntos de treino (70%) e teste (30%) para preservar a ordem temporal, evitando aleatoriedade que poderia comprometer a coerência da série. Foi aplicada a normalização dos dados para o intervalo entre 0 e 1, ajustando diferentes unidades de medida à mesma escala e evitando distorções no processo de aprendizado da rede neural LSTM, já que variáveis com magnitudes maiores poderiam indevidamente receber mais peso durante o treinamento. A normalização do conjunto foi baseada na Equação 2.

$$x_{(normalizado)} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2)$$

Após a normalização dos dados, foi necessário reestruturar os conjuntos de treino e teste para o formato tridimensional (3D) exigido pela rede neural LSTM, convertendo as matrizes 2D iniciais, com amostras nas linhas e atributos (chuva e nível) nas colunas, em matrizes 3D que incorporam também o número de passos temporais, que é o número de dias que o modelo irá considerar para a

previsão (*time steps*). Essa adaptação considerou o tempo de concentração da bacia hidrográfica, utilizando 2, 5 e 7 dias como parâmetros para avaliar a influência do histórico de dados na previsão do nível do rio. A seguir, foi estruturado o modelo LSTM, cuja performance foi avaliada por meio de diferentes combinações de hiperparâmetros: número de neurônios na camada LSTM (*units*), tamanho do lote (*batch size*) e número de épocas (*epochs*), com ativação sigmoide na camada densa para capturar a não-linearidade dos dados. Para automatizar essa etapa, criou-se uma classe com funções que implementam o modelo e testam variações desses hiperparâmetros, selecionando como melhor modelo aquele que apresentou o menor valor de erro na validação (*val\_loss*<sup>5</sup>). Ressalta-se que para esta etapa utilizou-se a linguagem *Python*. A Tabela 2 apresenta os resultados dos hiperparâmetros

Tabela 2: Hiperparâmetros e respectivos valores testados para identificar o melhor modelo a ser utilizado na predição dos valores de nível

Hiperparâmetro	Valores testados	Melhores valores (utilizados no modelo de predição)
Units	4 a 78	20
Batch Size	8, 16, 32	8
Epochs	50 - 100	50

Fonte: Elaborado pelos autores

Como camada de entrada (*input\_shape*) foram utilizadas as matrizes 3D de treinamento (denominada *X\_train*) e de teste (*X\_test*) sendo adotado o otimizador *Adam*<sup>6</sup>. Como medida de avaliação da função de perda foi definido o Erro Quadrático Médio (*Mean Squared Error - MSE*), indicado para avaliar a acurácia de modelos que buscam resolver problemas de regressão, ressaltando que sua avaliação se baseia na diferença entre os valores reais (observados) e os preditos (simulados), conforme apresenta a Equação 3

$$MSE = \frac{1}{n} \sum_{t=0}^{n-1} (y_{obs.} - y_{predito})^2 \quad (3)$$

Findado o treinamento da rede e definido o melhor modelo, o próximo passo foi sua aplicação à predição dos conjuntos de treino e de teste a fim de serem obtidos os valores de nível relacionados ao treino (denominado como *train\_pred*) e ao teste (*test\_pred*). Por fim, para a análise dos resultados procedeu-se a inversão da normalização trazendo os resultados preditos para a escala real dos dados de nível (em centímetros), facilitando a comparação com os dados observados para o período em análise. Outra medida utilizada na avaliação dos resultados foi a Raiz do Erro Quadrático Médio (*Root Mean Square Error - RMSE*), que consiste na raiz quadrada do MSE, conforme Equação 4

<sup>5</sup> *Val\_loss*: é a função de perda da validação tem o objetivo de monitorar como o modelo está generalizando a predição para dados que ele não viu durante o treinamento. Ela é calculada do mesmo modo que a *Loss*. Esta por sua vez é a métrica que avalia o quão bem o modelo está se ajustando aos dados de treinamento. Nesse estudo a medida adotada é a *mean squared error (mse)*. A documentação pode ser acessada em: [https://keras.io/api/losses/regression\\_losses/#meansquarederror-class](https://keras.io/api/losses/regression_losses/#meansquarederror-class)

<sup>6</sup> *Adam (Adaptive Moment Estimation)* é um otimizador baseado em gradiente que busca ajustar os parâmetros peso e *bias* durante o processo de treinamento, de modo a minimizar a função de perda (*loss*). A documentação pode ser encontrada em <https://keras.io/api/optimizers/adam/>

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (y_{obs.} - y_{predito})^2} \quad (4)$$

Ressalta-se, ainda, que além das medidas supracitadas, foi também utilizado o Coeficiente de Nash-Sutcliffe (*Nash-Sutcliffe Efficiency - NSE*), medida comumente adotada na avaliação previsões de variáveis hidrológicas. Esse coeficiente também avalia a diferença entre os valores observados e os preditos, conforme exemplifica a Equação 5. Como critérios deste coeficiente aponta-se as seguintes classes: (i) Bom: ( $NSE > 0,75$ ); (ii) Qualificado ( $0,36 < NSE < 0,75$ ) e; (iii) Não Qualificado ( $NSE < 0,36$ ).

$$NSE = 1 - \frac{\sum_{i=1}^n (y_{obs.} - y_{predito})^2}{\sum_{i=1}^n (y_{obs.} - y_{média\ obs.})^2} \quad (5)$$

## RESULTADOS:

Os resultados consistem na Etapa 5 – Avaliação do Modelo. A rede neural foi treinada com 70% dos dados históricos (de 01/01/1970 a 01/07/1994) e testada com os 30% restantes (de 02/07/1994 a 31/12/2004), utilizando as métricas referenciadas na metodologia, além da análise de erros percentuais por *time step* (2, 5 e 7 dias). Os resultados indicaram que o desempenho do modelo variou conforme o número de passos adotado, sendo o *time step* de 2 dias o que apresentou os piores resultados, tanto em termos de erro quanto na proximidade entre os valores previstos e observados. Já os *time steps* de 5 e 7 dias apresentaram melhores desempenhos, com destaque para o de 7 dias, que obteve os menores valores de erro e maior precisão nas previsões.

Ao analisar o RMSE na escala original dos dados de nível (em centímetros), observou-se que o *time step* de 2 dias apresentou erro médio de 167 cm, com 28,7% dos dias superando esse valor. Já o passo de 5 dias reduziu esse erro para 70,1 cm (14,2% dos dias com erro igual ou superior), e o de 7 dias apresentou o menor erro médio, de 67,7 cm (12,6% dos dias). Essa evolução também se refletiu no NSE, que foi insatisfatório para o *time step* de 2 dias, mas melhorou significativamente para os passos de 5 e 7 dias, com destaque para o último, que obteve os coeficientes mais altos tanto no conjunto geral quanto nas divisões por treino e teste, indicando maior precisão do modelo nas previsões de nível do rio. A Tabela 3 apresenta as medidas de erro para cada *time step*

Tabela 3: Resumo das medidas de erro conforme os diferentes passos (*time steps*) adotados

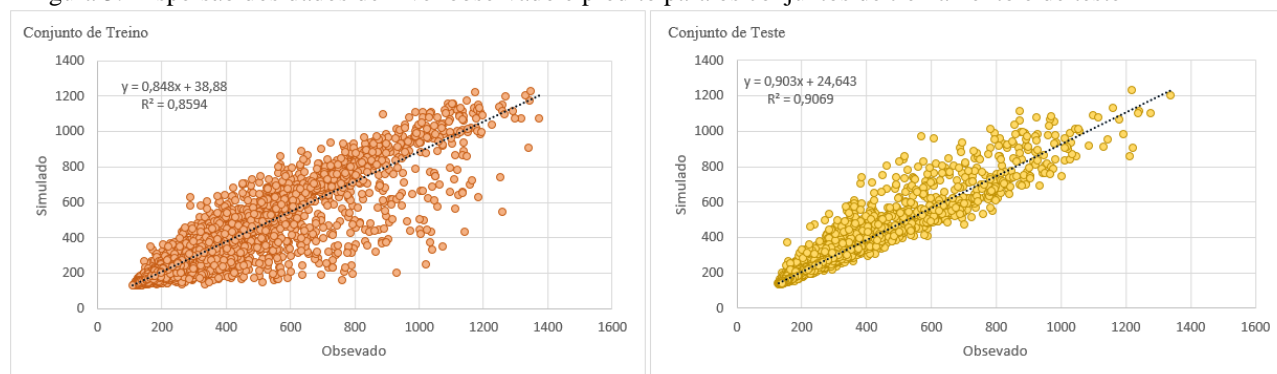
	Geral			Treino			Teste		
Medida de erro	Predições com <i>Time Step</i> 7 dias	Predições com <i>Time Step</i> 5 dias	Predições com <i>Time Step</i> 2 dias	Predições com <i>Time Step</i> 7 dias	Predições com <i>Time Step</i> 5 dias	Predições com <i>Time Step</i> 2 dias	Predições com <i>Time Step</i> 7 dias	Predições com <i>Time Step</i> 5 dias	Predições com <i>Time Step</i> 2 dias
MSE	0,0035	0,0037	0,017	0,0033	0,0035	0,017	0,0039	0,0042	0,018
RMSE	0,059	0,061	0,13	0,057	0,059	0,13	0,062	0,065	0,13
NSE	0,87	0,86	0,26	0,86	0,85	0,26	0,91	0,90	0,25

Fonte: Elaborado pelos autores

Com vistas ao entendimento do desempenho do modelo em cada um dos conjuntos, a Figura 3 apresenta os gráficos de dispersão dos valores de nível preditos em relação aos observados. Percebe-se que o conjunto de teste apresentou melhor desempenho que o de treinamento, descartando-se assim a possibilidade de ter ocorrido *overfitting* ou *underfitting*.



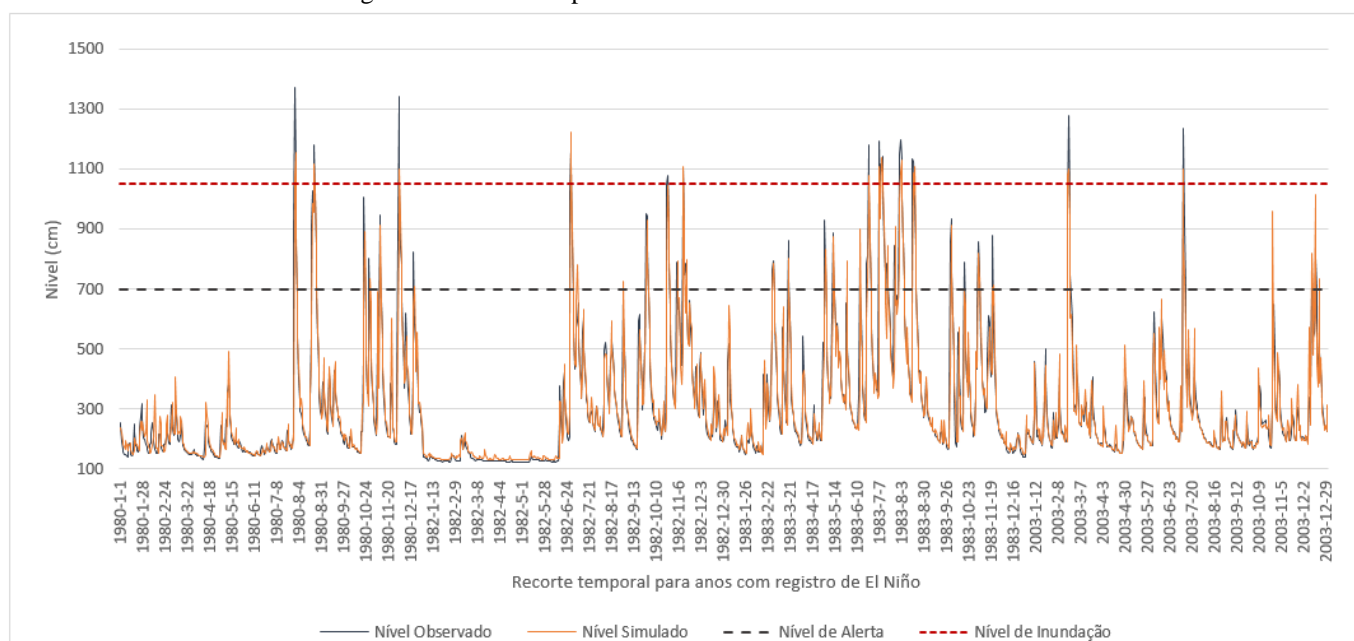
Figura 3. Dispersão dos dados de nível observado e predito para os conjuntos de treinamento e de teste



Fonte: Elaborado pelos autores

Tendo em vista o objetivo deste estudo, realizou-se a análise dos anos em que houve a ocorrência do fenômeno ENOS relacionado ao El Niño, uma vez que os *outliers* máximos ocorreram todos nesses anos. De modo geral verifica-se um bom desempenho do modelo para os anos em questão, contudo segue subestimando os níveis acima da cota de inundação, conforme ilustra a Figura 4

Figura 4: Recorte temporal referente aos anos de El Niño.



Fonte: Elaborado pelos autores

Quando se analisa apenas os dias em que houve registro de nível de alerta (acima de 700 cm), verifica-se que o modelo teve baixo desempenho, apresentando valores simulados inferiores ao nível de alerta, embora tenha acertado o nível de alerta em 87% das vezes em que houve nível de alerta para os dados observados. O mesmo comportamento é observado quando analisa-se os dias em que foram registrados níveis de inundação, nesse caso apresentando um desempenho ainda pior. Considerando que, do total de dias da série temporal, apenas 551 apresentaram níveis de alerta (o que representa 4,3% do total de dias) e, nos casos de inundação esse valor diminui ainda mais uma vez que apenas em 112 dias da série temporal houve registro de níveis de inundação (representando 0,9%), atribui-se a redução do desempenho do modelo à escassez de amostras de dias com níveis máximos. Acredita-se que o enriquecimento da série temporal, estendendo-a para um período ainda

maior, mas sobretudo que englobe uma quantidade maior de registros de níveis de inundação, repercutirá em uma melhora nos resultados referentes aos níveis de inundação e de alerta preditos.

## CONCLUSÃO:

A rede neural LSTM demonstrou bom desempenho na predição de níveis do rio, acertando 87% das ocorrências associadas a níveis de alerta durante o período testado, com melhor acurácia para valores não extremos. Apesar da subestimação em eventos mais severos, o modelo apresentou boas métricas de avaliação, especialmente o coeficiente de Nash-Sutcliffe, indicando sua adequação para futuras aplicações na previsão de níveis de inundação. Considerando a limitação de dados primários, sugere-se a ampliação da série temporal com dados indiretos, como os obtidos via sensoriamento remoto ou reanálises. Além disso, destaca-se a agilidade e simplicidade do uso da LSTM em comparação a modelos hidrológicos tradicionais, tornando-a uma ferramenta promissora para previsões rápidas e eficientes.

## REFERÊNCIAS:

- ABREU, M.C; CECÍLIO, R.A.; PRUSKI, F.F.; SANTOS, G.R.; ALMEIDA, L.T.; ZANETTI, S.S. *Critérios para escolha de distribuições de probabilidade em estudos de eventos extremos de precipitação*. Revista Brasileira de Meteorologia. <https://doi.org/10.1590/0102-7786334004>. Rio de Janeiro. 2018.
- BRUNER, M., SLATER, L., TALLAKSEN, L.M. e CLARK, M. *Challenges in modeling and predicting floods and droughts: A review*. *Wires Wiley Interdisciplinary Reviews*. Volume 8, Issue 3. 2021. Disponível em: <https://wires.onlinelibrary.wiley.com/doi/10.1002/wat2.1520> Consulta em: 13 de abril de 2024.
- COLLISCHONN, W; DORNELLES, F. *Hidrologia para engenharia e ciências ambientais*. Porto Alegre. Associação Brasileira de Recursos Hídricos – ABRH. 2013
- FAGUNDES, M. *Previsão hidrológica como ferramenta para auxiliar no critério de fechamento da trilha do Rio do Boi (SC)*. 2021. 148f Dissertação (Mestrado em Recursos Hídricos e Saneamento Ambiental) – Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021
- RIO GRANDE DO SUL. Secretaria de Planejamento, Governança e Gestão. *Mapa Único do Plano Rio Grande*. Disponível em: <https://mup.rs.gov.br/> Consulta em: 08 de junho de 2025
- SAMBATI, S. MARTINS, R. G.; VILELA, R. B.; COTACALLAPA, M.; PESSOA, A. S. A.; DIAS, J.; BRESSIANI, D.; FERNANDES, G. *Previsão de riscos de alagamentos e inundações com uso de inteligência artificial*. Revista de Informática Aplicada, Volume 15, Número 1. 2019
- SCHMIDT, V.; LUCCIONI, A.; TENG, M.; ZHANG, T.; REYNAUD, A. S. R.; COSNE, G.; JURAVER, A.; VARDANYAN, V.; HERNANDEZ-GARCIA, A.; BENGIO, Y. *ClimateGAN: Raising Climate Change Awareness by Generating Images of Floods*. 2021. Disponível em: <https://arxiv.org/pdf/2110.02871.pdf> Consulta em 20 de janeiro de 2024
- WORLD RISK REPORT. Bündnis Entwicklung Hilft / IFHV (2023): *WeltRisikoBericht 2023*. Berlin: Bündnis Entwicklung Hilft. 2023. Disponível em: [https://weltrisikobericht.de/wp-content/uploads/2023/10/WRR\\_2023\\_english\\_online161023.pdf](https://weltrisikobericht.de/wp-content/uploads/2023/10/WRR_2023_english_online161023.pdf) Acesso em 20 de março de 2024