UNIVERSIDADE FEDERAL DO PARANÁ

ALEX BRUNO KRAEMER

ESTIMATIVA DE COTAS E VAZÕES DE CURTO PRAZO EM BACIAS HIDROGRÁFICAS USANDO REDES CONVLSTM

CURITIBA

2022

ALEX BRUNO KRAEMER

ESTIMATIVA DE COTAS E VAZÕES DE CURTO PRAZO EM BACIAS HIDROGRÁFICAS USANDO REDES CONVLSTM

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Ciências Geodésicas.

Orientador: Prof. Dr. Daniel Rodrigues dos Santos

CURITIBA

2022

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Kraemer, Alex Bruno
Estimativa de cotas e vazões de curto prazo em bacias hidrográficas usando redes ConvLSTM / Alex Bruno Kraemer. – Curitiba, 2022.
1 recurso on-line : PDF.

Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Ciências da Terra, Programa de Pós-Graduação em Ciências Geodésicas.
Orientador: Prof. Dr. Daniel Rodrigues dos Santos

1. Hidrologia - Modelos. 2. Redes Neurais Artificiais. 3. Vazante. I. Santos, Daniel Rodrigues dos. II. Universidade Federal do Paraná. Programa de Pós-Graduação em Ciências Geodésicas. III. Título.

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

FOLHA DE APROVAÇÃO

MINISTÉRIO DA EDUCAÇÃO
SETOR DE CIENCIAS DA TERRA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS GEODÉSICAS - 40001016002P6

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação CIÊNCIAS GEODÉSICAS da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **ALEX BRUNO KRAEMER** intitulada: **ESTIMATIVA DE COTAS E VAZÕES DE CURTO PRAZO EM BACIAS HIDROGRÁFICAS USANDO REDES CONVLSTM**, sob orientação do Prof. Dr. DANIEL RODRIGUES DOS SANTOS, que após terem inquirido o aluno e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 02 de Março de 2022.

Assinatura Eletrônica
03/03/2022 19:56:02.0
DANIEL RODRIGUES DOS SANTOS
Presidente da Banca Examinadora

Assinatura Eletrônica
03/03/2022 17:05:53.0
JOSÉ EDUARDO GONÇALVES
Avaliador Externo (INSTITUTO TECNOLOGICO SIMEPAR)

Assinatura Eletrônica
04/03/2022 09:27:35.0
HIDEO ARAKI
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica
03/03/2022 23:13:07.0
RODRIGO DE CAMPOS MACEDO
Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Dedico este trabalho à minha Família, em especial aos meus pais Alécio (*in memorian*) e Marlei e à minha irmã Brunele, com todo o meu amor e profunda gratidão, pela paciência e apoio para comigo.

AGRADECIMENTOS

A Deus, por tudo que Ele representa em minha vida!

Ao professor e amigo Daniel Rodrigues dos Santos, pela orientação, apoio e confiança em meu trabalho.

Ao professor e amigo Daniel Carvalho Granemann, o qual foi de extrema importância na escolha de minha carreira profissional.

A meu amigo Kauê de Moraes Vestena por seu auxílio e colaboração no desenvolvimento deste trabalho.

Ao professor e amigo José Eduardo (ZéDu) pesquisador do SIMEPAR, por seus ensinamentos em eletrônica, aeromodelismo e Hidrologia.

Ao SIMEPAR por me oferecer bolsa de estudo, bem como fornecer os dados das estações hidrológicas e de precipitação.

Aos professores do Programa de Pós-graduação em Ciências Geodésicas.

A CAPES pela bolsa ofertada durante os meses iniciais desta pesquisa.

A Lizandra Castagnoli, cujo apoio e incentivo foi essencial para a finalização deste trabalho.

A Daniel Andrade, Kuaê de Moraes Vestena e Stanlei E. P. Fontana pelo auxílio na correção textual.

Aos meus amigos/irmãos: Luiz H. Bossola e Douglas A. dos Santos.

Aos demais amigos: Marcos A. Brocardo Jr, Adriellen S. Câmpara, Daniel Andrade, José R. B. Pedrosa, José C. Bica.

“Nossas virtudes e nossos sentimentos
são inseparáveis, assim como força e matéria.
Quando se separam o homem deixa de existir.”

Nikola Tesla

RESUMO

Nos últimos anos, o aprendizado de máquina (AM) é uma das linhas de pesquisa que tem se fortalecido na resolução de problemas cada vez mais complexos, estes problemas demandam de ferramentas computacionais sofisticadas e autônomas, reduzindo a necessidade de intervenção humana. Embora o advento de AM ofereça inúmeros benefícios, ainda há desafios em seu uso para a modelagem hidrológica, visto que, de acordo com HERATH et al. (2020), até o momento, nenhum modelo hidrológico consegue ter um desempenho igualmente satisfatório em toda a dinâmica dos fatores ambientais presentes em uma bacia hidrográfica. Atualmente, as inundações constituem o principal desastre natural no mundo e ocorrem durante eventos de alta pluviosidade aliado as características do relevo e uso do solo. Para tentar resolver este problema, neste trabalho é proposto o treinamento de um modelo de rede neural artificial (RNA) denominado ConvLSTM, empregando a abordagem distribuída. Para tal, foi necessário estabelecer os dados que influenciam diretamente no escoamento superficial d'água, sendo: (1) cota fluvial, (2) relevo, (3) *Curve Number* (CN) e (4) precipitação. Em seguida, processou-se os dados para se adequarem ao padrão de entrada da ConvLSTM, a qual utiliza o formato matricial. Para o treinamento do modelo, utilizou-se a plataforma Google Colab com processamento em nuvem, linguagem python 3 e as bibliotecas keras. Realizaram-se três experimentos: no primeiro, utilizou-se o modelo ConvLSTM, sem dados sintéticos; no segundo, acrescentaram-se dados sintéticos de chuva nas amostras de treino; no terceiro, empregaram-se dados sintéticos de cotas fluviais. Confrontando os resultados da estação de referência, observou-se que o acréscimo de dados sintéticos nas amostras de treino melhorou a capacidade do modelo em predizer os valores de cotas fluviais para a estação de referência.

Palavras-chave: Modelo Hidrológico, Modelo Distribuído, Previsão de vazão, Redes ConvLSTM, Redes Neurais Artificiais.

ABSTRACT

In recent years, machine learning (ML) is one of the research fields that has grown in importance for the resolution of increasingly complex problems. Such problems call for sophisticated and autonomous computational tools, thereby reducing human intervention. Even though the advent of ML brings countless benefits, there are many challenges concerning their use in hydrological modeling, as according to HERATH et al. (2020), currently, no hydrological model can perform equally well across the dynamic of all environmental factors present in a watershed. Floods are currently the world's major natural disaster that occurs during high rainfall associated with relief and land use characteristics. To address this issue, this paper proposes the training of an artificial neural network (ANN) model entitled ConvLSTM, with a distributed approach. To achieve this goal, it was necessary to establish the water surface runoff data, which are: (1) fluvial height; (2) relief; (3) Curve Number (CN); and (4) precipitation. Then, the data was prepared to fit the ConvLSTM input pattern, which employs the matrix format. Google Colab platform with cloud processing, python 3 language, and Keras libraries were used to implement the model training. Three experiments were carried out: in the first one, the ConvLSTM model was deployed without synthetic data; in the second one, synthetic rainfall data was added to the training samples; and finally, synthetic river level data was employed. By comparing the results at the reference station, one could observe that the addition of synthetic data in the training samples enhanced the model's power to predict the fluvial height of the reference station.

Keywords: Hydrological Model, Distributed Model, Streamflow Forecasting, ConvLSTM Networks, Artificial Neural Networks.

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE ABREVIATURAS E/OU SIGLAS

ANA – Agência Nacional De Águas e Saneamento Básico

AM – Aprendizado de Máquina

CN – Curve Number

CNS – Coeficiente de Nash-Sutcliffe

EMA – Erro Médio Absoluto

DL – Deep Learning

GPU – Unidade de Processamento Gráfico

INPE – Instituto Nacional de Pesquisas Espaciais

IAT – Instituto Água e Terra

LSTM – Long short-term memory

MDE – Modelo Digital de Elevação

MMS – média móvel simples

RNA – Redes Neurais Artificiais

RNC – Redes Neurais Convolucionais

RNR – Rede Neural Recorrente

SIMEPAR – Sistema de Tecnologia e Monitoramento Ambiental do Paraná

SRTM – Shuttle Radar Topographic Mission

SUMÁRIO

1 INTRODUÇÃO

1.1 CONSIDERAÇÕES INICIAIS

A humanidade, desde os seus primórdios, buscou compreender o dinamismo dos corpos d'água. Com o passar do tempo, a dinâmica dos fatores ambientais que regem o ciclo hidrológico foi sendo estudada progressivamente, culminando no desenvolvimento dos modelos hidrológicos, os quais desempenham um papel fundamental na aquisição das assinaturas de chuva e vazão em bacias hidrográficas.

Atualmente, as inundações constituem o principal desastre natural no mundo. Compondo uma parte natural do ciclo hidrológico, as inundações ocorrem durante eventos de alta pluviosidade aliado as características do relevo e uso do solo, causando impactos catastróficos nas esferas ambiental, econômica e social (HUA et al., 2020).

Entender e representar da melhor maneira possível o processo de escoamento da água em rios e bacias é fundamental para estudos de inundações. A predição de eventos hidrológicos é alcançada por meio da utilização de um sistema operacional de previsão de vazão. Este sistema fornece cenários futuros do escoamento superficial em um determinado local, apoiando a tomada de decisão e redução dos impactos negativos associados (FAN et al., 2020).

Contudo, de acordo com Yassen et al. (2015), os fenômenos e o padrão característico da vazão não são facilmente previsíveis, isso se deve ao fato de que a vazão de um rio é caracterizada por alta complexidade, não estacionariedade, dinamismos e não linearidade.

Dessa forma, até o momento, nenhum modelo hidrológico consegue ter um desempenho igualmente satisfatório em toda a dinâmica dos fatores ambientais presente em uma bacia hidrográfica, estimulando o surgimento de diferentes linhas de pesquisa, buscando diferentes modelos e estratégias de modelagem (HERATH et al., 2020).

Não obstante, Faceli et al. (2011), destacam que *"a crescente complexidade dos problemas a serem tratados computacionalmente demandam ferramentas computacionais sofisticadas e autônomas, reduzindo a necessidade de intervenção humana"*. A utilização de aprendizado de máquina (AM) é uma das linhas de pesquisa que tem se fortalecido nas últimas décadas, de acordo com a qual, diferentes algoritmos, adaptações e formas de treinamento são continuamente propostos.

As redes neurais artificiais (RNA) são um método de aprendizado de máquina e são aplicadas na modelagem hidrológica desde a década de 1990 (XIANG et al., 2020). Segundo esse método, o processo de aprendizado consiste em reconhecer num conjunto de dados um possível padrão incorporado, ou seja, "identificar" o processo que gerou os dados, que, uma vez reconhecido, pode ser usado para estimar valores em cenários futuros.

No treinamento das RNA aplicadas a modelagem hidrológica utilizam-se séries temporais de dados, contudo, as séries temporais das estações hidrológicas não são contínuas, visto que existem lacunas causadas por falhas nos sensores que mensuram o nível do rio (cota fluvial). Essas falhas resultam em dados inválidos e/ou discrepantes (*outliers*). Deste modo, para a geração de séries temporais coerentes, faz necessário realizar a consistência de dados fluviométricos, esse processo consiste em uma filtragem para remover leituras espúrias e preencher com dados sintéticos.

Consequentemente, à qualidade do treinamento da RNA está diretamente atrelada a qualidade dos dados de entrada, e, portanto, os dados sintéticos que serão inseridos na série temporal devem simular com rigor o fenômeno físico envolvido, evitando-se aumentar a incerteza dos dados de entrada.

1.2 CONTEXTUALIZAÇÃO DO PROBLEMA

Os órgãos/empresas de energia hidrelétrica, de abastecimento e defesa civil são os mais interessados em saber o comportamento hídrico futuro. Neste contexto, a previsão de vazão dos afluentes de reservatórios permite o planejamento do atendimento da demanda energética, controle de cheias e disponibilidade hídrica para consumo humano (FAN et al., 2020).

Segundo Herath et al. (2020), até o momento, nenhum modelo consegue ter um bom desempenho para caracterizar toda a dinâmica dos fatores ambientais presentes em uma bacia hidrográfica. Diante disso, a abordagem distribuída vem ganhando relevância na previsão hidrológica, uma vez que busca reconhecer o padrão entre as variáveis e suas relações espaciais por meio do georreferenciamento dos parâmetros que caracterizam fisicamente uma bacia hidrográfica.

Outro fator importante a ressaltar é a existência de lacunas nas séries históricas das estações hidrológicas. Essas lacunas são causadas por erros de leitura dos sensores, provocando a

falta de leituras e dados inválidos (*outliers*). Este problema pode ser minimizado através de processos de filtragem e geração de dados sintéticos, visando-se obter a completude da série histórica.

Sabe-se que as bacias hidrográficas são separadas por condicionantes geomorfológicas e que eventos de pluviosidade não interferem diretamente umas às outras, em outras palavras, se chover em uma determinada bacia e não chover na bacia adjacente, estas terão comportamentos diferentes para o escoamento fluvial, e isto, é um item relevante a ser observado na preparação dos dados de treinamento das redes neurais convolucionais.

1.3 ESTADO DA ARTE

Nesta pesquisa, para a revisão sistemática de literatura, empregou-se a metodologia *Methodi Ordinatio*, proposta por Pagani et al. (2017). Esta encontra justificativa no aumento das publicações científicas, o qual dificulta à seleção de material bibliográfico para o embasamento da pesquisa, assim, a principal característica da *Methodi ordinatio* é a classificação dos artigos quanto a sua relevância científica por meio de uma equação ponderada, denominada *InOrdinatio*.

Para cumprir com a tarefa de geração de previsões de vazão, diversas metodologias foram desenvolvidas ao longo das últimas décadas. Com o avanço dos recursos de *hardware* e algoritmos computacionais, modelagens baseadas em RNA têm apresentado resultados promissores. Dentre as arquiteturas de RNA, a *Long short-term memory* (LSTM) tem se destacado na modelagem de séries temporais, sendo indicada a séries temporais longas, consoante as quais as redes recorrentes comuns possuem dificuldades em encontrar padrões. Xiang et al. (2020) atestaram que um modelo LSTM mostra poder preditivo suficiente e pode ser utilizado para melhorar a precisão da previsão de enchentes a curto prazo (eventos repentinos).

Berkhahn et al. (2019) apresentaram um modelo artificial baseado em redes neurais para a previsão dos níveis máximos de água durante um evento de inundação repentina, visando prever as enchentes urbanas em tempo real. O modelo foi testado com eventos de chuva sintética em duas áreas com diferentes tamanhos e declives.

Em Hua et al. (2020) utilizaram-se técnicas de geoprocessamento para mapear áreas com risco de inundação durante eventos de precipitação, ocasião em que os eventos de enchentes anteriores foram comparados com o banco de dados de vulnerabilidade a enchentes para validar o

resultado modelado. Para tanto, empregaram-se dados geoespaciais de elevação, declividade, textura e drenagem do solo, relevo, precipitação, distância do rio principal, uso do solo e escoamento superficial.

Herath et al. (2020) adotaram uma abordagem de seleção de modelo quantitativa para selecionar um modelo ideal com complexidade em vez do paradigma “mais simples melhor”. Incorporaram a heterogeneidade espacial das propriedades de captação e variáveis climáticas na modelagem chuva-vazão, mantendo a parcimônia dos modelos induzidos, combinando os pontos fortes dos modelos baseados em física e modelos de ciência de dados.

Em ElSaadani et al. (2021) foi empregado um algoritmo *Deep Learning* (DL) que combina as propriedades das Redes Neurais Convolucionais (RNC) e das LSTM. Este algoritmo é denominado ConvLSTM e foi aplicado na previsão da umidade do solo, usando como entrada séries temporais de umidade do solo especializadas em matrizes de duas dimensões. O método proposto mostrou-se superior aos modelos RNC e LSTM, além de permitir o fornecimento de previsões discretas.

No QUADRO 1 está esquematizada a comparação entre os autores citados.

QUADRO 1 - COMPARAÇÃO DE METODOLOGIAS DO ESTADO DA ARTE

Autor	**Método**	**Vantagens**	**Desvantagens**
Xiang et al. (2020)	LSTM-based seq2seq	Trabalhar com séries temporais longas	Imprecisão da distribuição espacial da chuva
Berkhahn et al. (2019)	RNA + Chuva sintética	Maior robustez no treinamento da RNA	Não cita como a chuva sintética foi gerada
Hua et al. (2020)	Geoprocessamento	Controle sobre os processos realizados	Trabalho “manual/convencional”
Herath et al. (2020)	Combinação dos modelos baseados em física e ciência de dados	Modela a heterogeneidade espacial das variáveis	Necessita um maior volume de dados e modelagem explícita
ElSaadani et al. (2021)	ConvLSTM	Possibilita previsões discretas	Requer computadores mais robustos e elevada disponibilidade de memória RAM

FONTE: O autor (2022).

1.4 OBJETIVOS

1.4.1 Objetivo geral

O presente estudo tem por objetivo explorar a previsão horária de cotas fluviais em estações hidrológicas, incorporando dados sintéticos no treinamento de redes ConvLSTM.

1.4.2 Objetivos específicos

- Identificar uma arquitetura de RNA adequada para trabalhar com dados espaciais e séries temporais;
- Definir critérios para a sintetização dos dados espaciais;
- Adaptar e implementar uma função energia para o treinamento do modelo ConvLSTM;
- Calcular o valor das cotas fluviais para as estações hidrológicas;
- Avaliar o potencial do método proposto.

1.4.3 Contribuições

Esta pesquisa traz as seguintes contribuições:

- Utilizar RNA para aferição de dados fluviométricos;
- Empregar a arquitetura ConvLSTM para a previsão horária de cotas fluviais, utilizando quatro tipos de dados espaciais;
- Empregar dados sintéticos no treinamento da RNA.

2 REVISÃO DE LITERATURA

2.1 BACIAS HIDROGRAFICAS

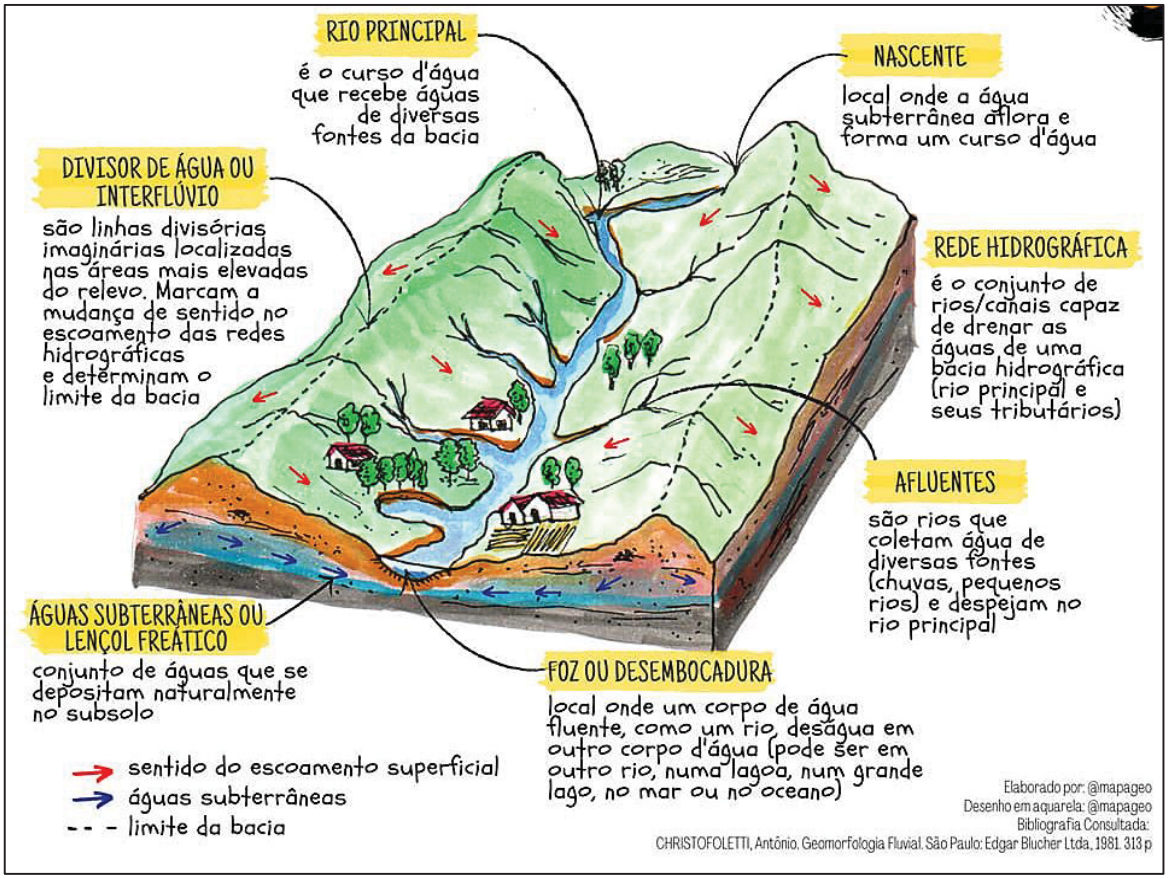
À medida que a economia se desenvolve e se diversifica, maior é a necessidade de uma gestão hídrica eficiente. Tundisi (2013 apud POLETO et al., 2019) constata que a gestão dos recursos hídricos é uma componente primordial no planejamento territorial e econômico, tornando-se um componente estratégico de grande valia nas últimas décadas. Dessa forma, a utilização da bacia hidrográfica como unidade de gestão decorre da necessidade de avaliar as causas e efeitos no ambiente determinados por ações antropizantes.

Cunha e Guerra (2003, apud SILVEIRA, 2017) definem que as bacias hidrográficas são porções da superfície terrestre delimitadas por condicionantes geomorfológicas (divisores de água ou interflúvios), que drenam a água, sedimentos e solutos para uma saída comum (exutório ou foz) por meio de um conjunto de canais de escoamento interligados (cursos de água, rios).

Em síntese, a ideia de bacia hidrográfica está na existência de nascentes, divisores de águas e características dos cursos de água, nas quais a água se desloca das partes mais altas para as mais baixas (POLETO, 2019). Na FIGURA 1 são ilustrados e comentados os elementos naturais do meio físico que compõem uma bacia hidrográfica.

É sobre o território definido como bacia hidrográfica que as atividades se desenvolvem, pois todas as áreas urbanas, industriais, agrícolas ou de preservação pertencem a alguma bacia hidrográfica. Portanto, no ponto exutório (foz) estarão representados todos os processos que definem a dinâmica da bacia.

FIGURA 1 - ELEMENTOS NATURAIS QUE COMPÕEM UMA BACIA HIDROGRÁFICA



FONTE: Christofoletti (1981 apud BLOG DE GEOGRAFIA, 2020).

2.2 MODELOS HIDROLÓGICOS

Modelos matemáticos são muito utilizados para resolver um problema comum em domínios científicos, que é a representação das relações entre variáveis físicas. A abordagem convencional para representar tais relações é usar modelos baseados em conhecimento científico, ou seja, modelos baseados em teoria. Uma abordagem alternativa é usar um conjunto de exemplos de treinamento envolvendo variáveis de entrada e saída para treinar um modelo de ciência de dados a extrair automaticamente as relações entre as variáveis (FAN et al., 2020).

A Hidrologia tem por finalidade a compreensão dos processos responsáveis pelo movimento, distribuição e qualidade da água em todo o planeta. Nesse contexto, a modelagem hidrológica é utilizada como ferramenta para obtenção de conhecimento mais aprofundado a

respeito dos fenômenos físicos envolvidos e na previsão de cenários (Karpatne et al., 2017; Moraes, 2003; Marinho Filho et al., 2012).

2.2.1 Previsão de vazões

De acordo com Fan et al. (2020), "*uma previsão de vazão consiste na estimativa do escoamento em um determinado local de um curso de água com uma definida antecedência temporal*". O conhecimento do comportamento do fluxo de um rio visa dar suporte à gestão hídrica, apoiando na tomada de decisão e na redução de impactos negativos associados a eventos hidrológicos, como, por exemplo, a previsão de inundações e estiagem, possibilitando o gerenciamento dos reservatórios de abastecimento.

Ainda de acordo com Fan et al. (2020), as previsões podem ser classificadas em conformidade à escala temporal em que foram calculadas, sendo:

a) **De curto a médio prazo:** quando o horizonte de previsão é de algumas horas até cerca de duas semanas. Sua previsibilidade é dependente das condições iniciais da atmosfera;
b) **Sub sazonais:** até aproximadamente 45 dias. Considera as condições iniciais da atmosfera, monitoramento das condições terrestre, oceânica e cobertura de gelo, da estratosfera e outras fontes;
c) **Sazonais:** até 9 meses. Também considera a energia oceânica.

As previsões de curto a médio prazo são aplicadas na identificação de eventos repentinos, como antecipação de impactos causados por inundações em áreas habitadas por meio da emissão de alertas, plano de assistência humanitária e evacuações.

Já as previsões sub sazonais e sazonais visam antecipar eventos hidrológicos mais lentos e de maior duração. São muito úteis no gerenciamento de reservatórios para eventos de estiagem, por meio do monitoramento de previsões, revisão dos planos de contingência e publicidade das decisões de planos estratégicos.

Uma das principais técnicas utilizadas para a previsão de vazão com antecedência superior ao tempo de concentração[1] de uma bacia hidrográfica, que inclui a previsão de precipitação como

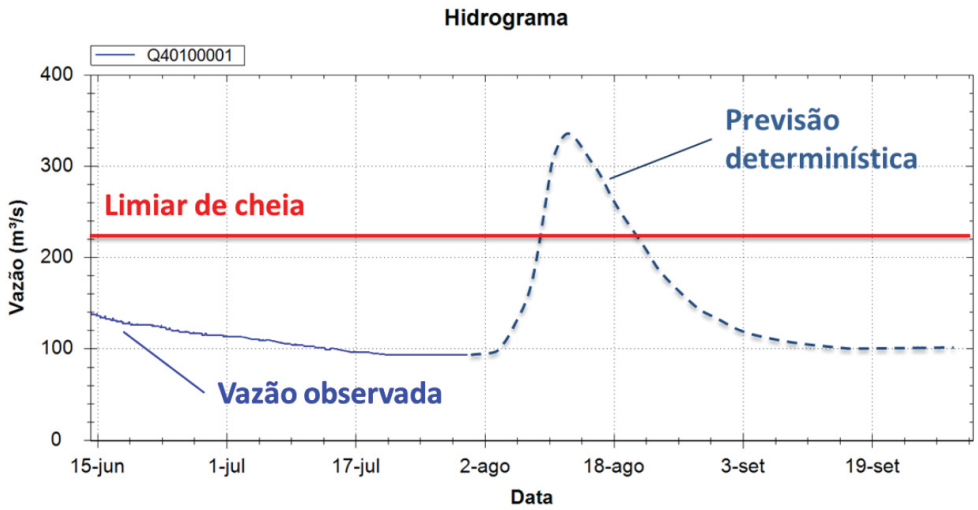
[1] Tempo de concentração é o intervalo de tempo para que toda a bacia contribua para o escoamento superficial da seção estudada. De maneira mais palpável, é o tempo para que a gota de água que cai no ponto mais distante (calculado hidraulicamente e não geograficamente) chegue até a seção que define o limite da bacia. FONTE: hidromundo.com.br

um dado de entrada no modelo hidrológico (FAN et al., 2020). Em síntese, utilizar um panorama futuro da precipitação enseja presumir o comportamento do escoamento superficial da bacia hidrográfica, caracterizando uma previsão determinística da vazão.

A principal dificuldade em determinar previsões hidrológicas determinísticas é a existência de diversas incertezas no sistema de previsão, como imperfeições nos modelos hidrológicos e meteorológicos, e/ou dados observados insuficientes.

FAN et al. (2020) ressaltam que as previsões determinísticas podem ser representadas através de um hidrograma, o qual mostra a trajetória prevista para a vazão e o limiar de cheia, conforme exemplo na FIGURA 2.

FIGURA 2 - APRESENTAÇÃO DO RESULTADO DE UMA PREVISÃO DETERMINÍSTICA



FONTE: FAN et al. (2020)

2.2.2 Tipos de modelos hidrológicos

Usualmente, na literatura, são encontradas formas de classificação para modelos hidrológicos. Para Tucci (2005 apud FAN et al., 2020), esses modelos podem ser classificados de acordo com sua estrutura, variabilidade espacial, fenômenos aleatórios e séries temporais, logo, segundo a:

- **Estrutura:** modelos empíricos, conceituais e de base física;
- **Representatividade espacial:** modelos concentrados e distribuídos;
- **Fenômenos aleatórios:** modelos determinísticos e estocásticos;

- **Séries temporais:** modelos de eventos e modelos de séries contínuas.

Em vista disso, esta pesquisa encaixa-se em modelos empíricos, distribuídos, determinísticos e séries contínuas. Deste modo, nos próximos tópicos serão descritas as características destas classificações.

2.2.3 Modelos empíricos

Consistem em modelos que utilizam séries de medidas (mensurações) para identificar a estrutura e os parâmetros da modelagem, por isso, são também conhecidos como modelos baseados em dados. Esses modelos buscam a correlação entre variáveis hidrológicas, sem necessidade de conhecimento anterior sobre o processo físico que pode ser responsável pela existência de correlação (FAN et al., 2020). Por exemplo, a relação entre as variáveis hidrológicas de entrada e saída pode ser obtida com a utilização de redes neurais artificiais.

Muitas vezes, os modelos empíricos são chamados de modelos “caixa preta” devido à falta de interpretabilidade das camadas intermediárias que interpretam os dados de entrada e os convertem em uma saída, ou seja, é muito difícil descrever fisicamente os processos intermediários, por este motivo parte da comunidade hidrológica reluta em utilizar modelos baseado em dados.

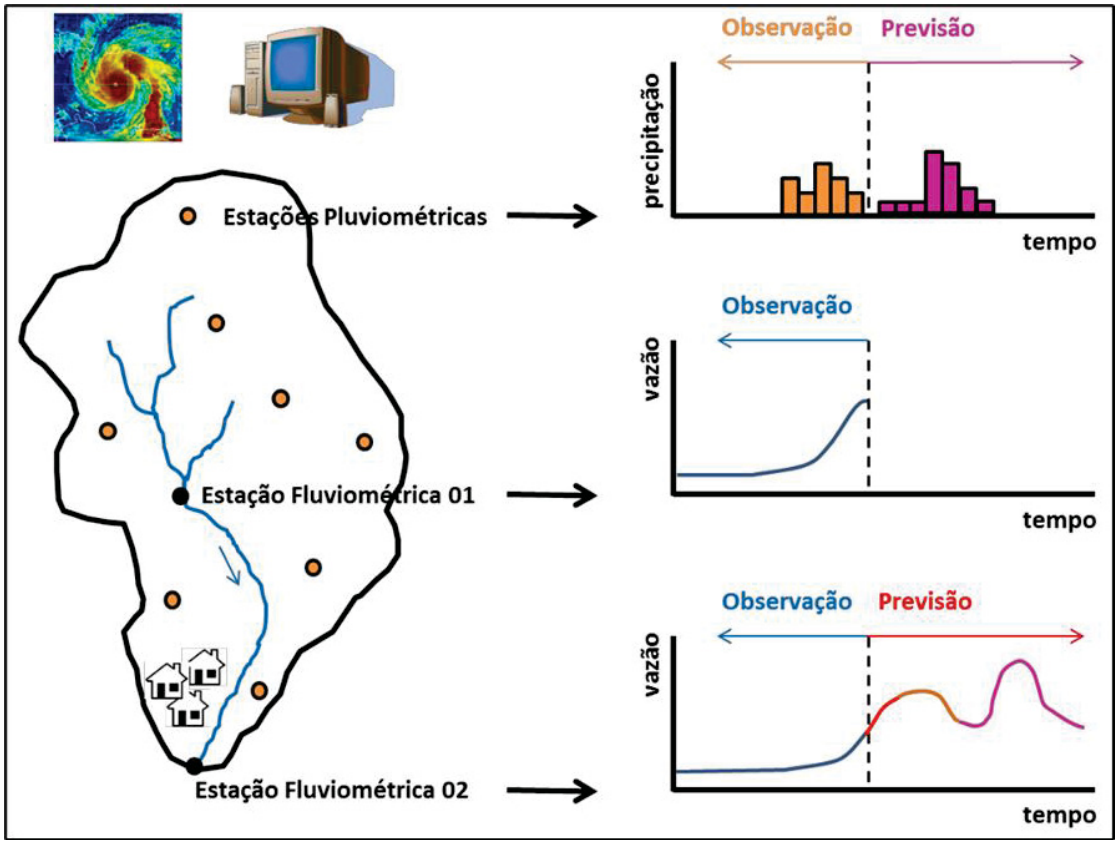
A principal vantagem de um modelo empírico é que ele pode ser usado sem uma explicação teórica para o fenômeno. No entanto, a desvantagem está em ser “fortemente dependente da qualidade dos dados utilizados”.

2.2.4 Modelos distribuídos

Cabe ressaltar a diferença entre os modelos concentrados e os modelos distribuídos. De acordo com Tucci (2005 apud FAN et al., 2020), se o modelo é do tipo concentrado, a dinâmica da bacia hidrográfica é representada de modo uniforme, então as previsões se baseiam na chuva observada, e não usam os dados de vazão existentes em locais intermediários da bacia, ou seja, utiliza a chuva média na bacia. Já se os modelos hidrológicos são do tipo distribuídos, estes consideram etapas intermediárias do ciclo hidrológico na bacia, como a propagação de vazões e caso possua mais de uma estação fluviométrica na bacia é possível utilizar os dados de vazões intermediárias observadas.

Fan et al. (2020) exemplificam o funcionamento de um modelo distribuído. Na FIGURA 3 está ilustrado esse exemplo, no qual a vazão na localidade da Estação Fluviométrica 02 é prevista por meio de uma combinação da onda de cheia da Estação Fluviométrica 01, chuva observada pela rede telemétrica e chuva prevista no horizonte de previsão por um modelo meteorológico.

FIGURA 3 - EXEMPLO DO FUNCIONAMENTO DE UM MODELO HIDROLÓGICO DISTRIBUÍDO



FONTE: FAN et al. (2020).

2.2.5 Modelos determinísticos

São os modelos que produzem respostas idênticas para o mesmo conjunto de entradas, mesmo quando uma variável de entrada tiver caráter aleatório. Para melhor entendimento, cabe descrever o seu antagônico, o modelo estocástico, neste uma ou mais variáveis envolvidas na modelagem têm um comportamento aleatório, possuindo distribuição de probabilidade. Por exemplo, o nível futuro de um reservatório depende da vazão afluente futura, que é uma variável aleatória com uma dada distribuição de probabilidade. Caso os conceitos de probabilidade sejam

negligenciados durante a elaboração de um modelo, este será denominado determinístico (Almeida et al., 2017, p. 132-133).

2.2.6 Modelos de séries contínuas

Dependem de como são representadas as séries temporais, se estas forem contínuas o modelo resultará em previsões periódicas, por outro lado, se a série temporal constituir numa determinada época (intervalo de tempo), o modelo será caracterizado como modelo de eventos, isto é, modela um evento específico.

2.3 REDES NEURAIS ARTIFICIAIS

A capacidade que alguns seres vivos possuem para aprender a partir de suas experiências é considerada essencial para se ter um comportamento inteligente. Este mesmo princípio é utilizado para diferenciar os algoritmos de AM dos demais algoritmos. Na literatura existem várias definições para o aprendizado de máquina. Mitchel (1997 apud FACELI et al., 2011, p. 3) descreve o aprendizado de máquina como a "*capacidade de melhorar o desempenho na realização de alguma tarefa por meio da experiência*". Atividades como memorizar, observar e explorar situações para aprender e organizar o conhecimento podem ser consideradas atividades relacionadas ao aprendizado.

> As redes neurais artificiais são modelos computacionais inspirados no sistema nervoso dos seres vivos. As quais possuem capacidade de aquisição e manutenção do conhecimento e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (sinapses artificiais), sendo as mesmas representadas por vetores/matrizes de pesos sinápticos (SILVA et al., 2010, p. 24).

2.3.1 Neurônio artificial

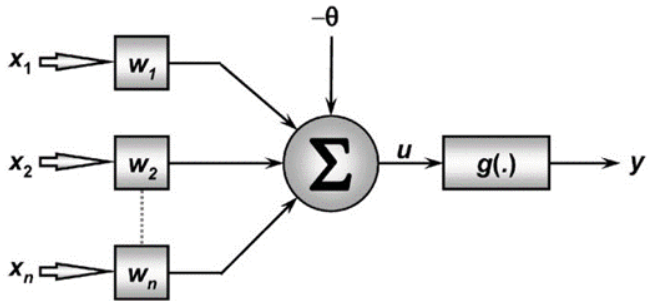
Em 1943, os cientistas McCulloch e Pitts desenvolveram um modelo de neurônio simples e que engloba as principais características de uma rede neural biológica.

Os neurônios artificiais empregados nos modelos de RNA são não-lineares, fornecem saídas tipicamentes contínuas e realizam funções simples, como coletar os sinais existentes em suas

entradas, agregá-los de acordo com sua função operacional e produzir uma resposta, levando em consideração sua função de ativação inerente (SILVA et al. 2010).

O modelo de neurônio artificial mais comumente utilizado é apresentado na FIGURA 4.

FIGURA 4 - ESTRUTURA DO NEURÔNIO ARTIFICIAL



FONTE: DSA (2020); SILVA et al. (2010).

Analisando-se a FIGURA 4, percebe-se que um neurônio artificial é composto por sete elementos básicos, onde:

a) **Sinais de entrada {X1, X2, X3, …, Xn}:** são os sinais ou medidas advindas do meio externo, usualmente normalizados para melhorar a eficiência computacional dos algoritmos de aprendizagem. Esses sinais representam os valores assumidos pelas variáveis de uma aplicação específica;
b) **Pesos sinápticos {W1, W2, …, Wn}:** são os valores utilizados para ponderar cada sinal de entrada da rede. Seus valores são aprendidos durante o treinamento e tem o objetivo de quantificar a relevância de cada sinal de entrada em relação a funcionalidade do neurônio;
c) **Combinador linear {Σ}:** tem a função de agregar todos os sinais de entrada que foram ponderados, a fim de produzir um valor de potencial de ativação;
d) **Limiar de ativação {θ}:** é uma variável que define o limiar apropriado para que o resultado produzido pelo combinador linear possa gerar um valor de disparo de ativação;
e) **Potencial de ativação {u}:** É o resultado obtido pela diferença entre o valor calculado no combinador linear e o limiar de ativação. Se este valor for positivo, então o neurônio produz um potencial excitatório; caso contrário, o potencial será inibitório;
f) **Função de ativação {g}:** é uma função que tem por objetivo limitar a saída do neurônio dentro de um intervalo de valores, a serem assumidos pela sua própria imagem funcional. "Uma rede neural sem função de ativação é essencialmente apenas um modelo de regressão

linear. A função de ativação faz a transformação não-linear nos dados de entrada, tornando-o capaz de aprender e executar tarefas mais complexas" (DSA, 2020);

g) **Sinal de saída {y}:** é o valor final de saída, produzido pelo neurônio em relação a um determinado conjunto de sinais de entrada, podendo ser utilizado por outros neurônios que estejam sequencialmente interligados.

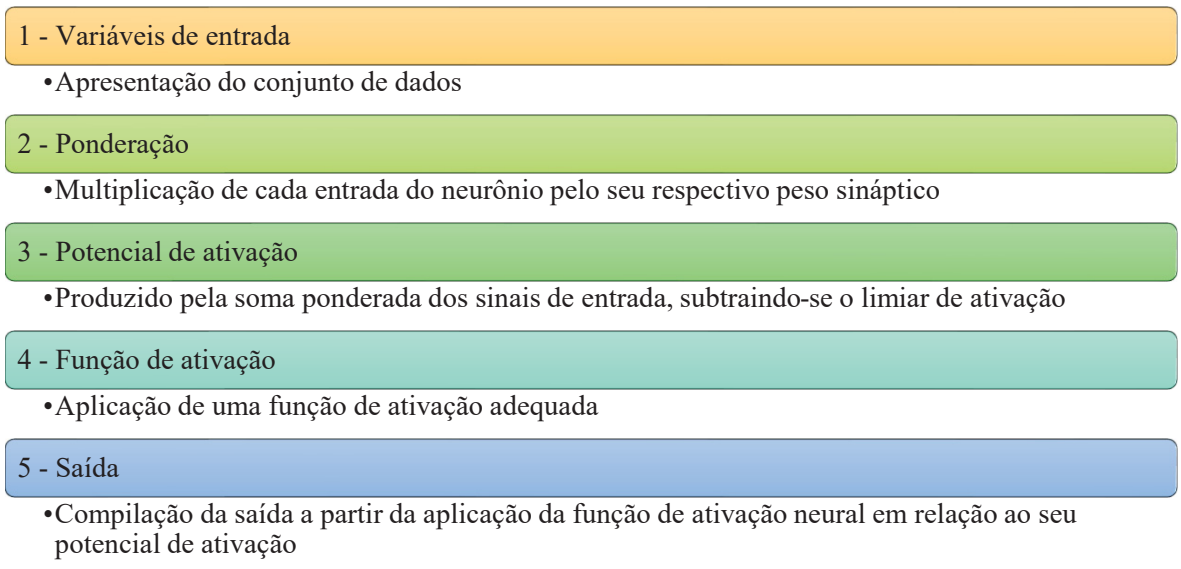
Em termos algébricos, o neurônio apresentado na FIGURA 4 é expresso por meio das equações 1 e 2.

$$u = \sum_{i=1}^{n} w_i * x_i - \theta \quad (1)$$

$$y = g(u) \quad (2)$$

Logo, pode-se resumir o funcionamento de um neurônio artificial por meio dos seguintes passos, como mostra a FIGURA 5.

FIGURA 5 - RESUMO DO FUNCIONAMENTO DE UM NEURÔNIO ARTIFICIAL



FONTE: Adaptado de SILVA et al. (2010)

2.3.2 Arquiteturas das RNA

A arquitetura de uma rede neural define o arranjo e o relacionamento entre os neurônios. Segundo SILVA et al. (2010), "*esses arranjos são essencialmente estruturados através do direcionamento das conexões sinápticas dos neurônios*".

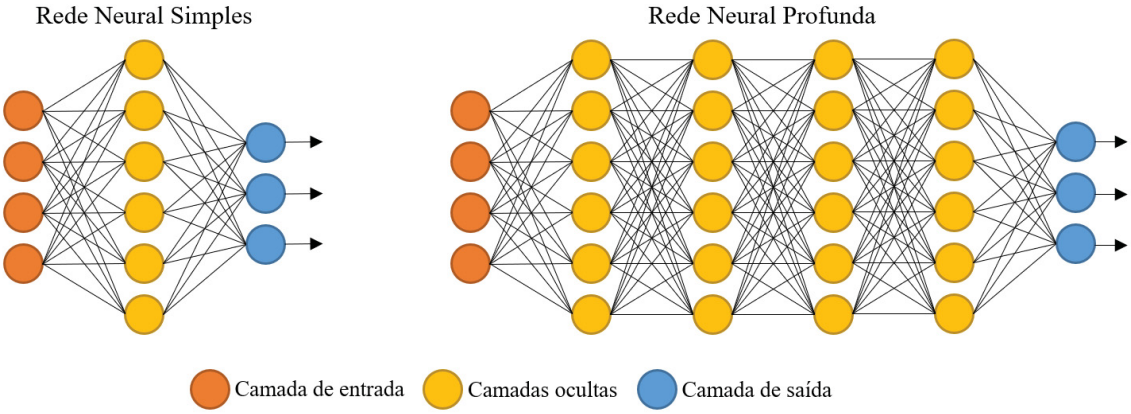
Uma RNA pode ser dividida em três partes, denominadas de camadas, as quais são nomeadas da seguinte forma:

a) **Camada de entrada:** responsável por receber os dados de entrada, sendo estes usualmente normalizados. Está normalização implica numa melhor precisão numérica frente às operações matemáticas realizadas pela rede;

b) **Camadas ocultas:** compostas pelos neurônios que possuem a responsabilidade de extrair as características associadas ao processo ou sistema a ser inferido. Quase todo o processo interno da rede é realizado nessas camadas. Quanto mais camadas ocultas a rede possuir, mais profundo será o aprendizado (FIGURA 6);

c) **Camada de saída:** também é constituída de neurônios, sendo responsável pela produção e apresentação dos resultados da rede, os quais são advindos dos processamentos efetuados pelos neurônios das camadas anteriores.

FIGURA 6 - REDE NEURAL SIMPLES E REDE NEURAL PROFUNDA



FONTE: Traduzida de DSA (2020)

As principais arquiteturas de redes neurais artificiais considerando a disposição de seus neurônios, formas de interligação entre eles e a constituição de suas camadas, podem ser divididas

em: redes de camada simples, redes de camadas múltiplas, redes recorrentes e redes reticuladas (DSA, 2020).

2.3.3 Aprendizagem

De acordo com Sampaio (2021), o processo de aprendizagem (treino) de uma rede neural pode ser dividido em dois tipos, o aprendizado supervisionado e o não supervisionado. Para melhor entendimento, as características de treinamento de uma RNA foram agrupadas no QUADRO 2.

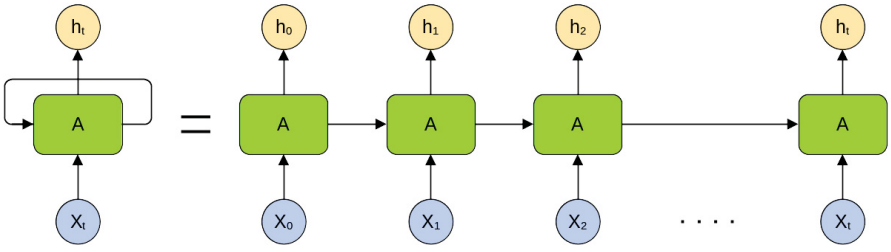
QUADRO 2 - MÉTODOS DE APRENDIZAGEM DE UMA RNA

Algoritmo supervisionado	**Algoritmo não supervisionado**
É necessário ter uma resposta pré-estabelecida	Útil para extrair informações que não seriam identificadas com análises tradicionais de gráficos e relatórios
Exemplo: Desenvolver algoritmo para análise de cliente de banco e verifica-se: ▪ Pode pegar empréstimo (SIM ou NÃO) ▪ O banco pode conceder aumento no limite de créditos (SIM ou NÃO)	Exemplo: Aplicar o algoritmo para identificar: ▪ Perfis de clientes semelhantes ▪ Filmes semelhantes com base na classificação e desenvolver um sistema de recomendação de séries e filmes

FONTE: Adaptado de Sampaio (2021)

2.4 REDES NEURAIS RECORRENTES

FIGURA 7 - ARQUITETURA SIMPLIFICADA DAS REDES NEURAIS RECORRENTES



FONTE: DSA (2020)

Uma Rede Neural Recorrente (RNR) pode ser imaginada como múltiplas cópias da mesma rede, ou seja, em formato de *loop* (FIGURA 7), na qual a mensagem de treino de cada *loop* é repassada ao seu sucessor. Na literatura, as RNRs são amplamente empregadas em séries

temporais, estando intimamente relacionadas a sequências e listas. Entretanto, boa parte do sucesso das RNRs se deve a uma de suas variações, a LSTM (DSA, 2020; SAMPAIO, 2021).

2.4.1 Arquitetura da LSTM

A arquitetura *Long Short Term Memory* – LSTM, em português memória de longo prazo, é indicada para séries temporais longas, nas quais as redes recorrentes comuns possuem dificuldades em encontrar padrões.

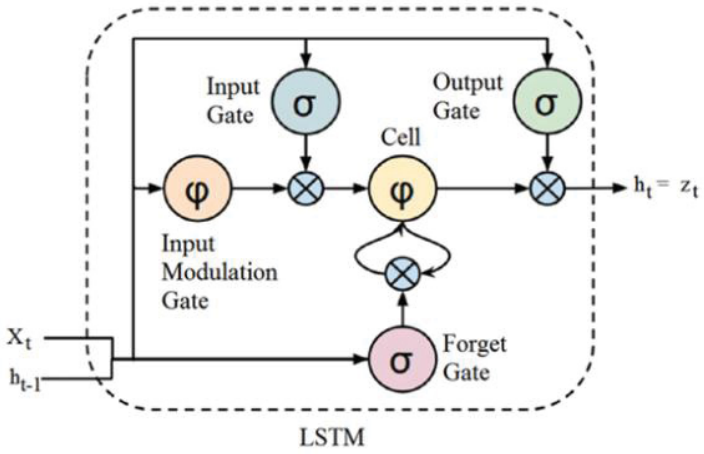
Essencialmente, para fazer predições em uma série temporal, passa-se os valores de Y tanto como entrada quanto de saída. A entrada deve ter uma série de valores anteriores para predizer um valor futuro, em outros termos, é possível passar um intervalo com conjunto de valores anteriores ao que se quer estimar.

A LSTM é uma arquitetura de rede neural recorrente que "lembra" valores em intervalos arbitrários, adequada para classificar, processar e prever séries temporais com intervalos de tempo de duração desconhecida.

A insensibilidade relativa ao número de intervalos de entrada dá uma vantagem à LSTM em relação as RNRs tradicionais. Assim, ao prever a sequência após 1000 intervalos em vez de 10, o modelo RNR esqueceu o ponto de partida até então. Mas, um modelo LSTM é capaz de "lembrar" por conta de sua estrutura de células (DSA, 2020).

A LSTM possui uma estrutura em cadeia que contém quatro redes neurais e diferentes blocos de memória chamados "células". A FIGURA 8 ilustra essa estrutura. As informações são retidas pelas células e as manipulações são feitas pelos portões (*gates*).

FIGURA 8 - ARQUITETURA DA LSTM



FONTE: DSA (2020)

- ***Forget Gate:*** Controla as informações que não são mais úteis ao estado da célula, ou seja, as informações que devem ser esquecidas;
- ***Input Gate:*** Controla adição de informações úteis ao estado da célula, ou seja, as informações que devem ser lembradas;
- ***Output Gate:*** Extrair as informações úteis do estado da célula atual para serem apresentadas como uma saída.

2.5 REDES CONVOLUCIONAIS

A idealização das redes convolucionais remonta aos anos 70. Porém, foi a partir do artigo "*Gradient-based learning Applied to document recognition*" de Yann LeCun et al., 1998, que a terminologia de redes convolucionais foi estabelecida.

> "A inspiração neural [biológica] em modelos como redes convolucionais é muito tênue. É por isso que eu os chamo de 'redes convolucionais' e não 'redes neurais convolucionais', e por isso os nós eu chamo de 'unidades' e não 'neurônios'" (LeCun et al. 1998, apud DSA, 2020).

Ainda assim, as redes convolucionais usam muitos dos mesmos conceitos que as redes neurais utilizam, como: retro propagação, gradiente descendente, funções de ativação não lineares, entre outros.

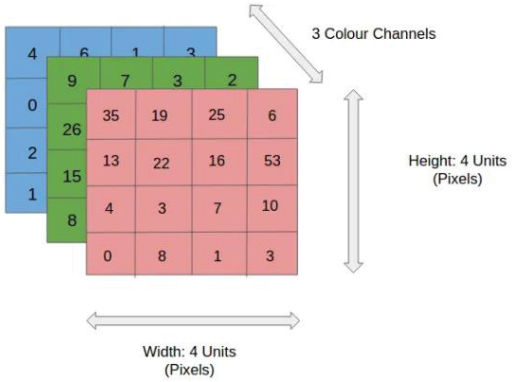
Uma rede neural convolucional – RNC (ConvNet / *Convolutional Neural Network* - CNN) consiste em um algoritmo de aprendizado profundo com aplicabilidade em imagens, ou seja, imagens são dadas como entrada e a rede atribui importância (pesos e vieses) a vários aspectos e características, sendo capaz de encontrar padrões e classificações de objetos nessas imagens.

2.5.1 Características das RNC

Diferente de outras arquiteturas de RNA, nas convolucionais o número de parâmetros da rede não está vinculado ao tamanho da entrada, isto resolve o problema de super-parametrização (DAS, 2020), em outras palavras, nas RNC a quantidade de parâmetros é definida pelo usuário. Tantos os modelos sub-parametrizados quanto os super-parametrizados produzem resultados de baixa qualidade. Enquanto os sub-parametrizados geram resultados pouco acurados, os super-parametrizados perdem a capacidade do modelo em generalizar o fenômeno estudado capturando e assimilando os ruídos presentes nos dados.

As RNC são capazes de interpretar imagens de forma mais intuitiva, pois trabalham com matrizes 3D (largura, altura e profundidade) conforme ilustrado pela FIGURA 9, ou seja, com imagens multicamadas.

FIGURA 9 - DADOS DE ENTRADA DAS RNC



FONTE: XAVIER (2019)

Do mesmo modo, os neurônios (também organizados em 3D) possuem um campo de visão limitado (tamanho do kernel). As principais operações de uma RNC são:

I. Convolução: transforma a imagem em características;
II. *Pooling*: processo de sub-amostragem.

A RNC transforma hierarquicamente a imagem em características de cada vez mais alto nível, e então realiza a inferência. Para isto, as camadas da RNC alternam entre a transformação e a sub-amostragem, e quanto mais profunda a rede, mais mapas de características a representação vai ter (LARANJEIRA, 2021).

Camadas de uma RNC:

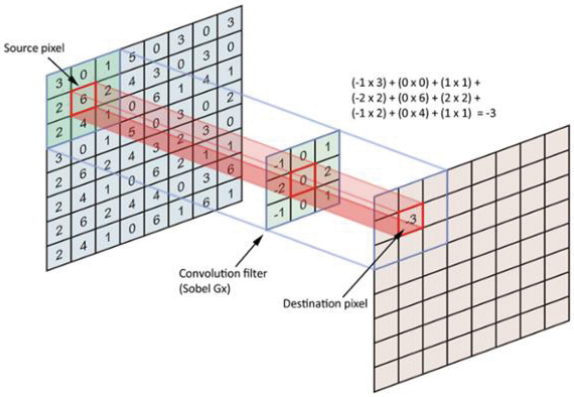
I. Convolucional: camadas responsáveis pela convolução;
II. *Pooling*: camadas responsáveis pela sub-amostragem;
III. *Batch normalization*: melhora o aprendizado da rede;
IV. Totalmente conectada: camadas de inferência.

Desta forma, as primeiras camadas de uma RNC buscam características de baixo nível (genéricas), não apresentam informações de alta carga semântica dos dados em questão. Por outro lado, as últimas camadas responsabilizam-se pela identificação de características de alto nível, e quanto mais profunda for a rede, mais semânticas serão essas características.

2.5.2 Filtro convolucional

No contexto de processamento digital de imagens, o *kernel* é um filtro convolucional. Este consiste em uma matriz n-dimensional que será operada com o dado através de uma convolução. Na FIGURA 10 é ilustrado um exemplo de convolução, na qual é possível ver o papel do *kernel* na análise da vizinhança de um pixel.

FIGURA 10 - EXEMPLO DE CONVOLUÇÃO



FONTE: XAVIER (2019)

Presume-se que a convolução mede a semelhança entre os dois sinais. Para tanto, precisa-se propor um *kernel* que simula o padrão procurado. Todavia, enfatiza-se que a correlação vai medir este padrão procurado, operando funções após inverter o *kernel*.

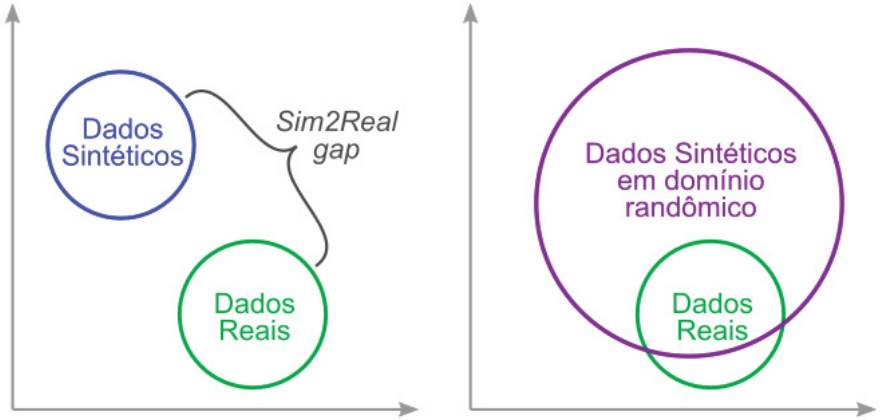
2.6 DADOS SINTÉTICOS

A aprendizagem com dados sintéticos é uma forma cada vez mais popular de robustecer (encorpar, fortalecer) os modelos de redes neurais profundas, que demandam grande volume de dados. Isto torna o modelo treinado mais robusto a variações nos dados, ou seja, aumenta a capacidade da rede em generalizar.

Atualmente, é fácil de visualizar a aplicação de dados sintéticos para o reconhecimento de padrões/objetos em imagens, pois se enfatiza a aleatoriedade do modelo digital em diferentes posições, orientações e padrões de fundos, buscando-se forçar a rede a aprender somente o essencial em um domínio randômico.

No entanto, existe uma lacuna entre os dados produzidos por simulação e os dados reais, conhecida como "*sim2real gap*". Essa lacuna pode ser preenchida através de uma técnica conhecida como randomização de domínio (Weng, 2019; Tobin et al., 2017 apud PONTE, 2020). Assume-se que o domínio randômico é o conjunto de todas as possibilidades possíveis para representar um fenômeno, e que grande parte do conjunto dos dados reais está contido no domínio randômico, conforme ilustra a FIGURA 11.

FIGURA 11 - SIM2REAL GAP



FONTE: Traduzida de PONTE (2020).

De acordo com Weng (2019 apud PONTE, 2020), o preenchimento da *sim2real gap* tem por objetivo tornar os dados sintéticos mais próximos da realidade, e para tanto, existem algumas abordagens, sendo:

a) Identificação do sistema:
    - Consiste em construir um modelo matemático para um sistema físico;
b) Adaptação de domínio (DA):
    - Refere-se a um conjunto de técnicas de transferência de aprendizagem desenvolvidas para atualizar a distribuição dos dados simulados com a distribuição real por meio de um mapeamento ou regularização imposta pelo modelo.
c) Randomização de domínio (DR):
    - Cria-se ambientes simulados com propriedades aleatórias e treinar um modelo que funciona em todos eles;
    - É possível que esse modelo possa se adaptar ao ambiente do domínio real, uma vez que, espera-se que o sistema real seja uma amostra daquela rica distribuição de variações de treinamento.

O DA e o DR não são supervisionados. Comparado ao DA que requer uma quantidade adequada de amostras de dados reais para capturar a distribuição, o DR pode precisar de poucos ou nenhum dado real.

2.7 CURVE NUMBER

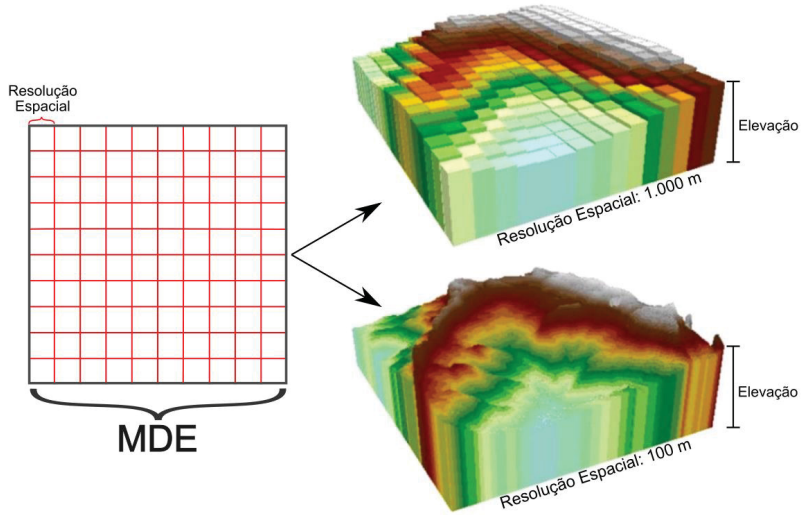
"*O método Curve Number (CN) desenvolvido pelo Soil Conservation Servive (SCS) é um método simples, muito difundido e eficiente para determinar o volume aproximado de escoamento superficial de um evento de chuva em uma região*" (ANA, 2018).

Soares et al. (2017) apresentam o parâmetro CN como um valor tabelado que descreve a combinação do tipo de solo, umidade antecedente e o uso e cobertura na bacia. Em síntese, o CN é uma forma de classificar uma região pelo potencial de retenção de águas pluviais.

2.8 MODELOS DIGITAIS DE ELEVAÇÃO

"*Os modelos digitais de elevação (MDE) são arquivos que contêm registros altimétricos estruturados em linhas e colunas georreferenciadas, como uma imagem com um valor de elevação em cada pixel*" (VALERIANO, 2008, p. 21). Segundo o mesmo autor, a utilização direta da elevação como fator analítico tem grande importância em estudos que envolvem hidrologia superficial, inundações e processos envolvendo movimentos gravitacionais. Na FIGURA 12 é possível visualizar a relação entre a resolução espacial (dimensão do pixel no terreno) e a representação do relevo.

FIGURA 12 - MODELO DIGITAL DE ELEVAÇÃO



FONTE: Modificado de SAMPSON; CASALI (2020).

De acordo com Valeriano (2008), o projeto TOPODATA do Instituto Nacional de Pesquisas Espaciais (INPE) oferece dados geomorfométricos para todo o território brasileiro derivados de dados do SRTM (*Shuttle Radar Topographic Mission*). Esses dados foram refinados da resolução espacial original de 3 arco-segundos (~90 metros) para 1 arco-segundo (~30 metros) por krigagem e estão estruturados em quadrículas compatíveis com a articulação 1:250.000, portanto, em folhas de 1° de latitude por 1°30' de longitude salvas em arquivos no formato GEOTiff e disponíveis no site do INPE.

2.9 SIPREC

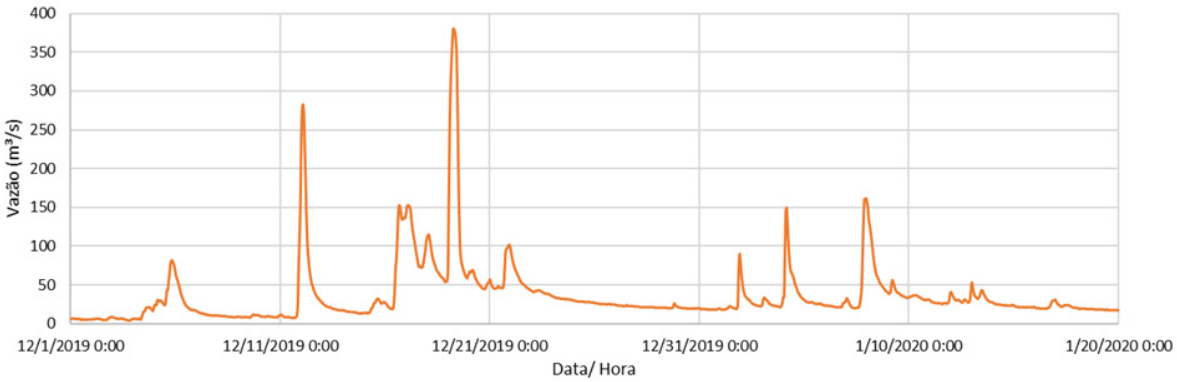
A estimativa de precipitação quantitativa de alta resolução por radar e satélite combinada com pluviômetros é um dos guias mais importantes para previsões hidrológicas. Enquanto os pluviômetros fornecem medições precisas em um ponto, o sensoriamento remoto ajuda a recuperar o padrão espacial da precipitação. Assim, o SIPREC consiste em um algoritmo multi-sensor empregado na obtenção de uma grade georreferenciada de precipitação (dados raster), utilizando um mosaico de radar, estimativa de satélite e uma rede de pluviômetros (CALVETTI et al., 2016).

A resolução temporal e espacial do SIPREC é de 1 hora e 1 km2, respectivamente.

2.10 SÉRIES HIDROLÓGICAS

Em geral, as variáveis hidrológicas (cota fluvial, vazão, etc.) são registradas por meio das chamadas séries hidrológicas, que constituem as observações organizadas no modo sequencial de sua ocorrência no tempo (RENNÓ et al, 2017). Na FIGURA 13 é apresentado um exemplo de série hidrológica, mais especificamente uma série de vazão (Q).

FIGURA 13 - EXEMPLO DE SÉRIE HIDROLÓGICA DE VAZÃO



FONTE: O autor (2022).

Segundo Rennó et al. (2017), as séries hidrológicas podem apresentar erros, sendo eles:

- Pontuais ou isolados: falhas na leitura ou no arquivamento dos dados;
- Sistemáticos: mudança do local da medição ou perda de calibração do aparelho.

3 MATERIAIS E MÉTODOS

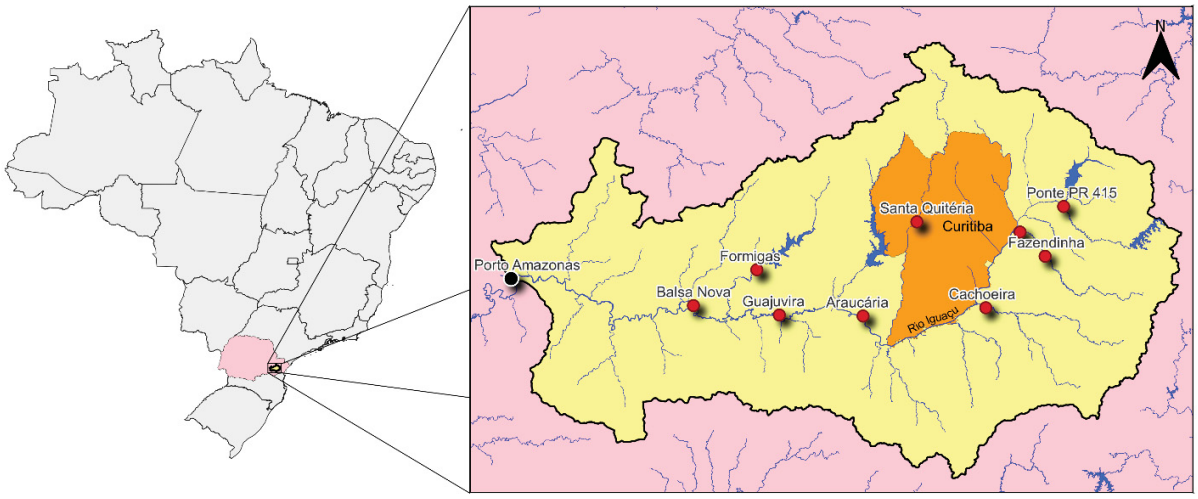
3.1 MATERIAIS

Os materiais a serem utilizados nesta pesquisa são descritos a seguir:

3.1.1 Área de estudo

Para estudo e validação desta pesquisa, foi selecionada a área de contribuição da Estação Hidrológica de Porto Amazonas – PR, a qual é apresentada na FIGURA 14. Uma área de contribuição consiste em uma superfície de terreno que direciona topograficamente o escoamento de água para um determinado ponto.

FIGURA 14 - ÁREA DE CONTRIBUIÇÃO DA ESTAÇÃO HIDROLÓGICA DE PORTO AMAZONAS – PR



FONTE: O autor (2022).

Descrevendo as informações contidas na figura acima, em amarelo está representada a abrangência geográfica da área de contribuição da Estação Hidrológica de Porto Amazonas; em laranja o munícipio de Curitiba e em pontos vermelhos/preto as estações hidrológicas utilizadas nesta pesquisa. Parte desta área contempla a bacia hidrográfica do Alto Iguaçu, sendo considerada uma das mais significativas bacias do estado do Paraná, segundo a Secretaria Estadual do Meio Ambiente.

De acordo com Secretaria de Desenvolvimento Sustentável e do Turismo, a bacia do Alto Iguaçu engloba a Região Metropolitana de Curitiba, situando-se no Primeiro Planalto é caracterizada geograficamente por:

- Ao Norte: apresenta altas declividades, baixa fertilidade do solo e grande potencial geológico para minerais não metálicos;
- A Leste, tem-se a Serra do Mar e as nascentes do Rio Iguaçu, com relevo plano de solos hidromórficos sujeitos a inundações, na parte rural ocupados por agricultura de hortigranjeiros.

Na área de estudo localizam-se cerca de 80 estações fluviométricas, cujos dados estão disponíveis em diferentes bancos de dados, como: do SIMEPAR, ANA, IAT e SNIRH. Entretanto, diante desse grande número de estações encontrou-se três dificuldades, sendo: (1) a primeira caracterizada pela diferença na escala temporal de leitura dos dados, ou seja, apenas algumas estações apresentam dados horários; (2) a segunda é marcada por possuírem apenas dados de cotas fluviais e não de vazão. Posto isso, apenas a estação de Porto Amazonas dispõe de dados de cota e vazão, sendo esta portanto a principal estação a ser utilizada no processo de verificação da qualidade das previsões geradas; (3) diz respeito à existência de lacunas e dados espúrios na série temporal de algumas estações.

Desta forma, visando-se escolher as melhores estações diante das dificuldades citadas no parágrafo anterior, selecionou-se as seguintes estações (QUADRO N) e agrupando-as segundo a completude da série temporal:

QUADRO 3 - ESTAÇÕES SELECIONADAS

Estação	**Completude (%) para o intervalo 01/08/2018 a 31/05/2021**	**Classificação**	**Banco de dados**
Araucária	99.82 %	Qualidade Boa	IAT
Balsa Nova	99.64 %	Qualidade Boa	IAT
Fazendinha	99.94 %	Qualidade Boa	IAT
Guajuvira	99.93 %	Qualidade Boa	IAT
Ponte PR-415	99.11 %	Qualidade Boa	IAT
Porto Amazonas	99.99 %	Qualidade Boa	SIMEPAR
Cachoeira	85.03 %	Qualidade Média	IAT
Comp. C. Água Limpa	86.20 %	Qualidade Média	IAT
Formigas	88.03 %	Qualidade Média	IAT
Santa Quitéria	83.94 %	Qualidade Média	IAT

FONTE: O autor (2022).

3.1.2 Recursos de *hardware* e *software*

- 1 computador. S.O. Linux Ubuntu 20.04, processador AMD Fx(tm)-8320 eight-core, 14GB de memória RAM, 2,4TB de HD, placa de vídeo Radeon HD 8760 2GB. Preparação e pré-processamento dos dados de entrada.
- 1 notebook. S.O. Windows 10, processador Intel Core i7, 8GB de memória RAM, 893GB de HD, placa de vídeo NVIDIA GTX 1050 Ti. Pesquisa e elaboração da dissertação.
- Visual Studio Code: editor de código-fonte *open source*, como python, java, css, php, entre outros;
- QGIS: software livre com código-fonte aberto, multiplataforma de SIG que permite a visualização, edição e análise de dados georreferenciados;
- Google Colab Pro+: é um ambiente de notebooks *Jupyte*r que não requer configuração e é executado na nuvem. Permite a criação e execução de códigos em Python no navegador web.
- Linguagem de programação Python 3;
- Pacote *Microsoft Office*: editor de planilha, apresentações e de texto (proprietário).

3.1.3 Dados de entrada

Os dados de entrada compreendem os dados utilizados para o treinamento de uma RNA. Os dados utilizados nesta pesquisa são oriundos das bases oficiais de órgãos de Cartografia de Referência e órgãos responsáveis pela coleta e disponibilização de dados em suas respectivas linhas de pesquisas.

QUADRO 4 - DADOS DE ENTRADA

Dado	**Escala original**	**Fonte:**
Altimetria	Grade de 30 metros	TOPODATA / INPE
Curve Number (CN)	1 / 250.000	Agência Nacional de Águas – ANA
Cota fluvial / Vazão	Horária (15 min.)	Instituto Água e Terra – IAT / SIMEPAR
Chuva	Grade de ~1.2 km	SIPREC / SIMEPAR

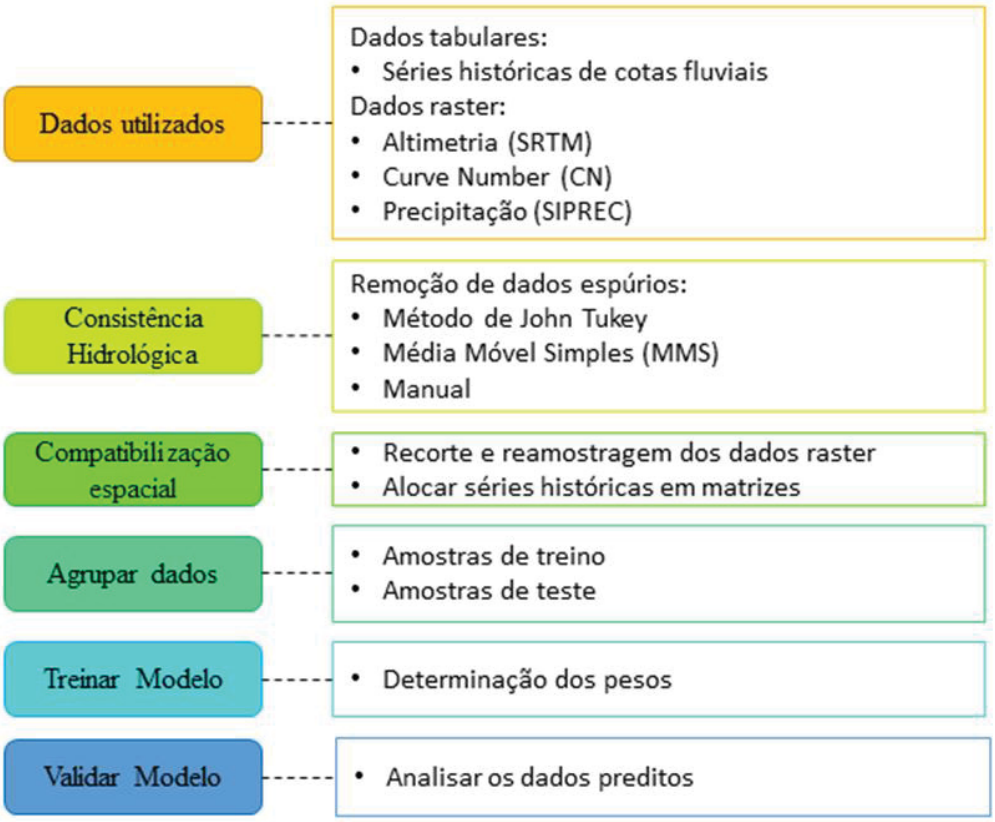
FONTE: O autor (2022).

Por envolver essencialmente uma análise espacial e baseando-se na pesquisa de Hua et al. (2020), os dados de entrada (QUADRO 4) são: altimetria, cota fluvial, chuva e CN. A escolha pelo CN se deve ao fato dele incorporar em seu valor o uso do solo e tipo do solo, resultando na taxa de retenção de escoamento superficial. Como o valor da vazão é produto de um modelo matemático denominado de curva chave da secção transversal, o sensor mede o valor da cota fluvial e posteriormente aplica-se a transformação de seu valor em vazão. Portanto, pela disponibilidade de dados de cotas fluviais e não disponibilidade da curva chave de todas as estações hidrológicas, as cotas serão utilizadas no treinamento dos modelos, cujo resultado será a predição de cotas fluviais.

3.2 MÉTODO

As etapas do método proposto estão resumidas na FIGURA 15.

FIGURA 15 – ETAPAS

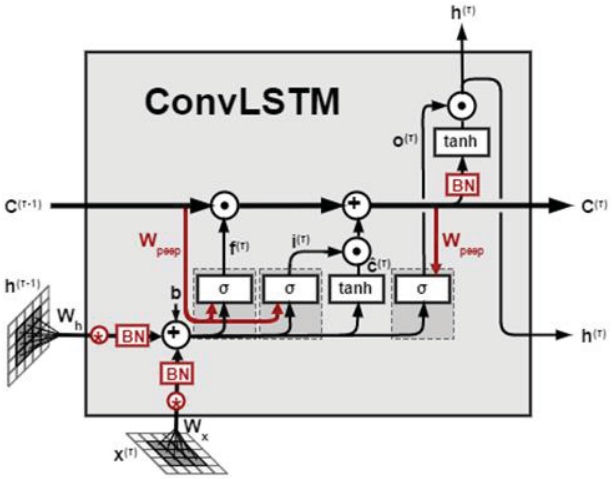


FONTE: O autor (2022).

3.2.1 Rede ConvLSTM

Uma rede com arquitetura ConvLSTM visa unir as características essenciais das LSTMs e das CNNs. De acordo com XAVIER (2019), "*a ConvLSTM é uma camada recorrente, assim como a LSTM, porém todas as operações internas de multiplicação de matrizes que aconteceria em uma LSTM são substituídas por convoluções*". Na FIGURA 16 é apresentado a composição de um neurônio ConvLSTM.

FIGURA 16 - NEURÔNIO DE UMA REDE CONVLSTM

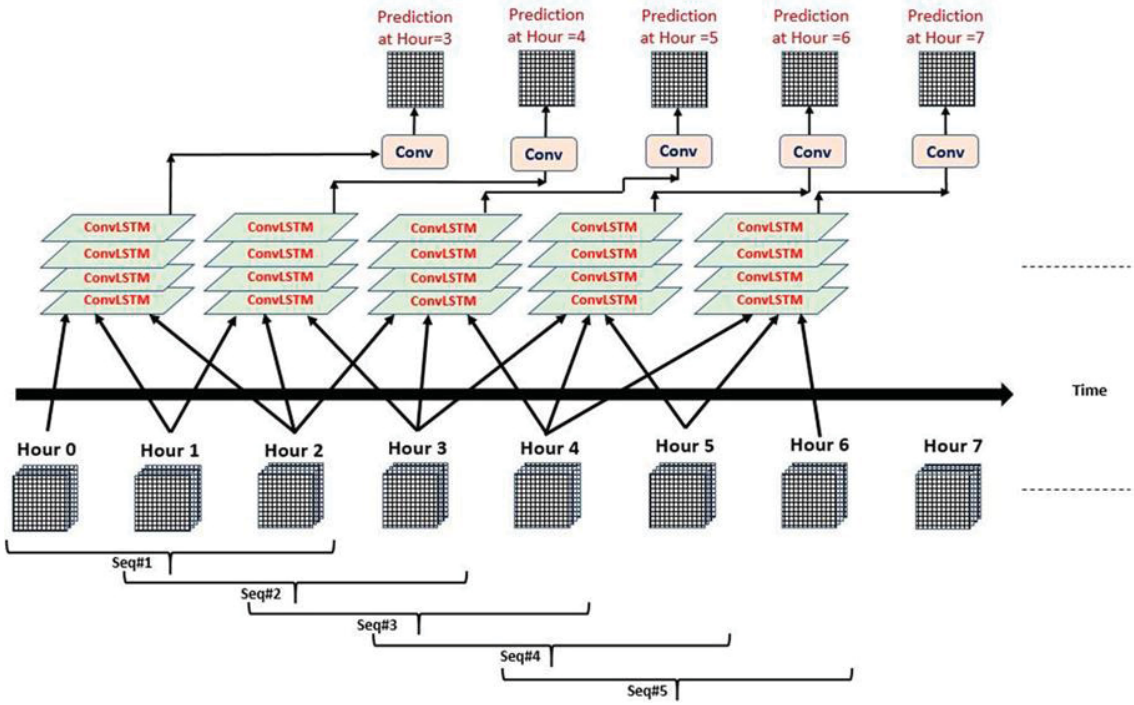


FONTE: XAVIER (2019)

XAVIER, (2019) ressalta que um modelo ConvLSTM é diferente de um modelo Convolução-LSTM. No modelo Convolução-LSTM a entrada passa primeiro por camadas de convolução, no qual o resultado desta convolução é transformado em um vetor 1D com as características obtidas, e estas repassadas a LSTM.

Na FIGURA 17 é esquematizado o princípio de predição realizada por uma ConvLSTM, em que um conjunto de dados de um determinado intervalo de tempo [t0,...,tn] é dado como entrada e determina-se o dado seguinte [tn+1].

FIGURA 17 - MÚLTIPLAS SEQUÊNCIAS DE ENTRADAS E SAÍDAS DE UMA CONVLSTM

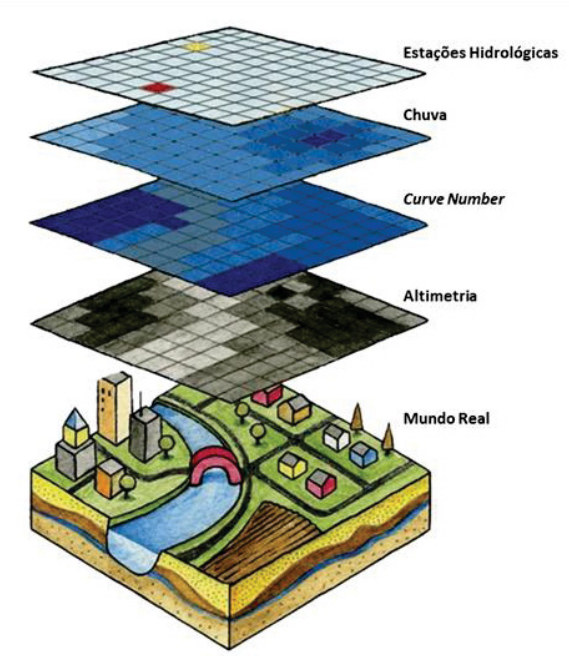


FONTE: ELSAADANI et al. (2021).

3.2.2 Processamento dos dados de entrada

Nesta etapa, os dados citados na seção 3.1.3 são processados de modo a serem compatíveis com o formato de entrada da ConvLSTM. Como os dados de entrada desta arquitetura de rede são matriciais, estes devem possuir a mesma escala espacial e o mesmo número de linhas e colunas, a fim de permitir que os dados sejam sobrepostos em forma de camadas, como mostra a FIGURA 18.

FIGURA 18 - DADOS UTILIZADOS



FONTE: Modificado de <https://www.pngwing.com/pt/free-png-dgneg>

Analisando a FIGURA 18, os dados de entrada são dispostos em camadas, em que cada camada representa um respectivo dado do mundo real. Todos estes dados são de entrada para o treinamento da rede, no entanto, somente os dados que expressam as estações hidrológicas são fornecidos para as saídas, visto que o objetivo desta pesquisa é obter os valores de cotas fluviais para as estações hidrológicas.

3.2.2.1 Dados altimétricos

Conforme descrito na seção 2.8, os dados altimétricos utilizados são fornecidos pelo INPE através do projeto TOPODATA. Utilizar-se-ão quatro folhas de MDE, sendo elas: (1) 24s51_, (2) 24s495, (3) 25s51_ e (4) 25s495, as quais precisarão passar pelo processo de mosaicagem, reamostragem espacial e posteriormente por recorte, obtendo-se um arquivo único e com as dimensões de interessa para a área de estudo.

A reamostragem espacial consiste na compatibilização da escala geométrica do TOPODATA (~30m) com a malha do SIPREC (~1km), assim, para não perder a informação

altimétrica dos canais de drenagem, o menor valor será utilizado como interpolador no processo de reamostragem.

3.2.2.2 Dados de potencial de retenção de águas pluviais

Devido à complexidade em classificar os solos, existindo inúmeras classificações, e ainda, considerar o aspecto hidrológico tema desta pesquisa, substituir o tipo de solo pelo valor CN visa ser mais conveniente e proveitoso no treinamento da ConvLSTM.

Como os dados de CN do Catálogo de Metadados da ANA estão disponibilizados em formato vetorial (shapefile), estes precisam passar por um processo de "rasterização", ou seja, necessitam ser transformados para dados em formato matricial (raster) para integrarem nos dados de entrada da ConvLSTM.

3.2.2.3 Dados de precipitação

Os dados de precipitação (SIPREC) são obtidos no banco de dados do SIMEPAR, sendo rasters que recobrem boa parte do hemisfério ocidental, estes necessitam serem recortados para o tamanho da área de interesse, diminuindo sua área de cobertura e consequente seu tamanho expressivo de armazenamento.

3.2.2.4 Séries hidrológicas

Para as estações escolhidas na delimitação da área de estudo, analisaram-se seus respectivos dados para o período de 01 de agosto de 2018 a 31 de maio de 2021 com escala temporal de 15 minutos para as leituras, constatando a existência de diferenças quantitativas e qualitativas nos dados, como, a presença de *outliers*, lacunas sem dados e deslocamentos (*offsets*).

Desta forma, primeiramente cada estação passou por um processo denominado consistência hidrológica, em que o primeiro passo consiste na detecção de *outliers*. Segundo Macêdo (2018), os *outliers* são dados anômalos que não seguem o padrão de comportamento de uma distribuição normal. A detecção dessas anomalias é distinta e está relacionada à "remoção de

ruído", ou seja, dados indesejados que não são de interesse (TENG et al., 1990 apud MACÊDO, 2018).

O método da média móvel simples (MMS) é uma das estratégias aplicadas para controlar a tendência em determinados dados da série temporal. É calculada pelo somatório de um conjunto de valores numéricos dividido pela quantidade de elementos somados. Este conjunto de valores consiste em uma fração (janela) do conjunto de dados total, a qual se desloca sequencialmente ao longo dos dados, e, assim, para cada janela é calculada uma média. A representação matemática da MMS é dada por:

$$MMS = \sum_{i=0}^{n} \frac{X_{n-i}}{n} \tag{4}$$

Onde:

- MMS: média móvel simples;
- n: tamanho do conjunto (janela);
- X_{n-i}: o valor do dado no instante $(n - i)$.

De acordo com Macêdo (2018), para uma determinada janela contínua de período de tempo, a média e o desvio padrão são calculados. Se a próxima entrada no conjunto de dados estiver entre a média ± 2*desvio padrão, é considerado normal, caso contrário é considerado uma anomalia.

Outra forma de identificar os dados outliers é utilizar o método de John Tukey[2], o qual considera para cálculo o primeiro (q1) e o terceiro (q3) quartil, os outliers serão os valores que atendam a condição: $valor < q1 - 1.5 * (q3 - q1)$ ou $valor > q3 + 1.5 * (q3 - q1)$.

Nem sempre os *outliers* representam erros de leitura, às vezes representam situações climáticas adversas, sendo, portanto, algo de difícil interpretação e remoção automática por algoritmos. Por isto, a remoção desses valores espúrios dá-se por:

1) Utilizar o método de John Turkey;

[2] Para mais detalhes, a implementação do método em linguagem python está disponível em: <https://kb.elipse.com.br/removendo-dados-discrepantes-outliers-com-a-linguagem-python/>

2) Utilizar o método de médias móveis, porém, dados com valor igual a média com ± 2*desvio padrão e diferença entre Xn e a média menor que 2 metros, passam por uma análise gráfica (visual/manual) para o respectivo período da média móvel.

Assim, procedeu-se a consistência hidrológica para os dados com escala temporal de 15 min. Em seguida, realizou-se a reamostragem de 15min para a escala horária 1hr, através de uma média simples entre os valores lidos no intervalo horário anterior ao instante T, dada pela equação (5).

$$D_h = (D_{15}^m + D_{30}^m + D_{45}^m + D_{60}^m)/4 \quad (5)$$

Onde:

- D_h: dado horário;
- $D_{15}^m, D_{30}^m, D_{45}^m, D_{60}^m$: dados com leituras a cada 15 minutos.

Dando sequência à manipulação destes dados, os dados das estações foram agrupados em matrizes espaciais considerando sua "qualidade" em conformidade com os respectivos experimentos a serem realizados. Para tanto, cada período (h) das leituras correspondem a uma camada de uma matriz de 3 dimensões, na qual a matriz tem valor zero em todas as células exceto nas células que correspondem a localização espacial de cada estação, atribuindo o valor da leitura da respectiva estação no período (h).

3.2.2.5 Dados sintéticos

Conforme descrito no item 2.6, a utilização de dados sintéticos no processo de treinamento de RNA tem por objetivo encorpar o modelo treinado, tornando-o mais robusto a variações nos dados e assim aumentar a capacidade do modelo em generalizar. Para tanto, idealizaram-se duas estruturas de dados sintéticos, sendo: (1) Mesclar dados SIPREC de períodos distintos na área de contribuição e nas bacias adjacentes; (2) Gerar a completude das séries hidrológicas das estações que possuem qualidade média (item 3.1.1).

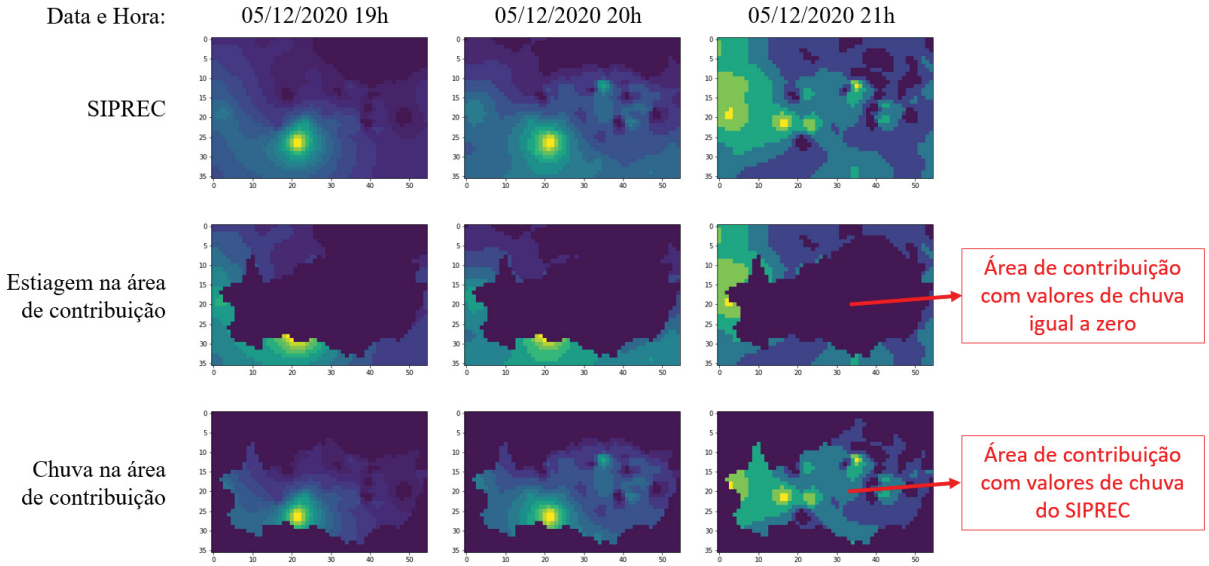
3.2.2.5.1 Mesclagem dos dados SIPREC

Os dados sintéticos aqui gerados visam destacar a relevância que os “pixels” representativos da área de contribuição possuem sobre as leituras das estações hidrológicas, mais precisamente, destacar que os “pixels” externos não devem influenciar as leituras. Para tanto, observaram-se as seguintes etapas:

1) Identificar nos dados SIPREC o maior período de estiagem na área de estudo;
2) Identificar nos dados SIPREC o maior período de precipitação na área de estudo;
3) Empregar uma máscara binária para filtrar os “pixels” internos e externos à área de contribuição;
4) Agrupar os dados das estações e do SIPREC para os respectivos períodos acima citados (1 e 2);
5) Mesclar dados SIPREC de estiagem na área de contribuição com dados de chuva no exterior desta, e utilizar as leituras das estações no período de estiagem;
6) Contrário ao passo anterior, neste a mesclagem é realizada com dados SIPREC de chuva na área de contribuição com dados de estiagem no exterior desta, e utilizar as leituras das estações no período de chuva.

Para melhor entendimento, na FIGURA 19 é apresentada uma amostra dos dados sintéticos de chuva, nos quais a primeira linha de imagens consiste no dado SIPREC descrito no item 3.2.3.3. Na segunda linha é observado a existência de chuva apenas na área externa, enquanto na terceira linha é mostrado a chuva somente na área de interna.

FIGURA 19 - Dados Sintéticos de Chuva



Nos eixos são exibidos o nº de pixels, cuja resolução espacial é 2 km

Fonte: O autor (2022)

3.2.2.5.2 Completude das estações médias

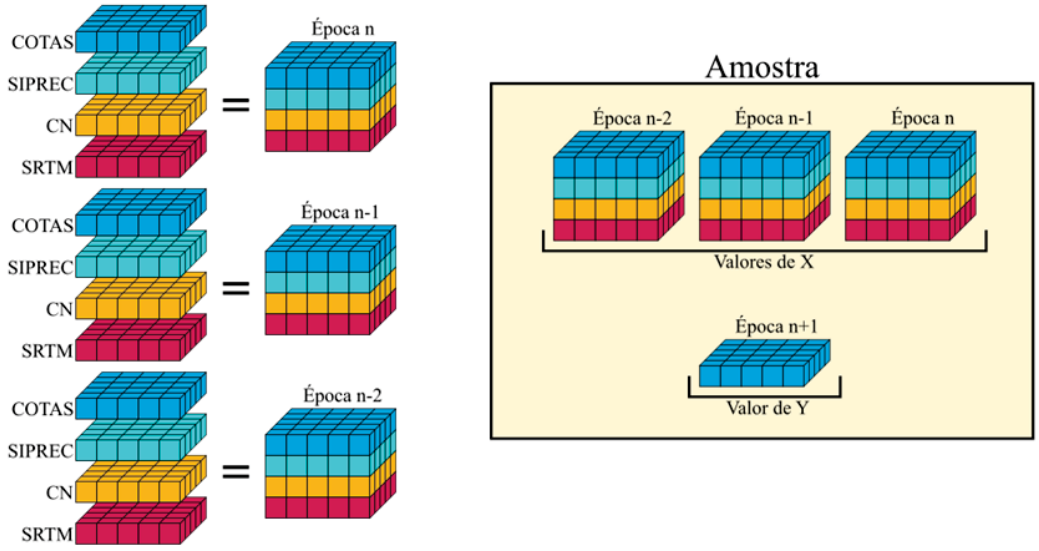
Devido à existência de lacunas na série hidrológica de cada estação, treinou-se um modelo LSTM para cada estação média, tendo como entrada as leituras das estações boas e da respectiva estação média, cujo valor de saída é o valor da leitura da estação média. Para tanto, utilizou-se no treinamento somente épocas sem lacunas de dados. Após o treinamento do modelo, passa-se toda a série temporal por um algoritmo, no qual é verificado de forma sequencial se o valor da leitura da estação média é nulo para aquele instante, se sim, este valor é substituído pelo valor calculado pela rede.

3.2.3 Treinamento do modelo ConvLSTM

No treinamento do modelo ConvLSTM utilizou-se a biblioteca *Tensorflow* integrada à biblioteca Keras, em linguagem Python 3. A plataforma em que o código foi executado é o Google Colab, o qual permite o processamento em nuvem com alto desempenho de *hardware* (GPU) próprio para o treinamento de RNC.

Os dados de treino correspondem ao conjunto de amostras utilizadas para treinar uma RNC, é a partir destes dados que a rede (modelo) irá “aprender” a caracterizar o fenômeno estudado, buscando o padrão existente nos dados de entrada (X) que geram a saída (Y), como exemplificado na FIGURA 20.

FIGURA 20 - ESTRUTURA DOS DADOS DE ENTRADA (X) E SAÍDA (Y) DA REDE CONVLSTM



FONTE: O autor (2022).

Explicando a FIGURA 23, a matriz valor Y caracterizada neste exemplo pela época n+1 consiste na saída do modelo, ou seja, consiste no resultado obtido a partir do conjunto de matrizes representadas pelos valores de X. O número de matrizes que compõem os valores de X é definido pelo parâmetro *time_steps* igual a três, portanto, os valores de entrada X constituem uma matriz de quatro dimensões, de tamanho [*time_steps*=3, nº de camadas = 4, nº de linhas = 5, nº de colunas = 5]. Desta forma, os dados de treinamento são constituídos por um conjunto de amostras (conjunto de valores de X).

3.2.4 Validação dos resultados

Também chamado de verificação de previsões, consiste no processo de avaliar o desempenho de um sistema de previsão. De acordo com Fan et al. (2020), “*a avaliação da*

qualidade de um modelo consiste em diagnosticar qual é o grau de correspondência entre as previsões e uma referência, tipicamente dada por observações."

As métricas estatísticas utilizadas para quantificar a qualidade das previsões são: (1) Erro Médio Absoluto (EMA) e (2) Coeficiente de *Nash-Sutcliffe* (CNS).

3.2.4.1 Erro Médio Absoluto (EMA)

Fan et al. (2020) descrevem o erro médio absoluto (EMA) como a diferença absoluta entre previsões e observações correspondentes para cada um dos horizontes de previsão, cuja formulação matemática é representada pela equação (6).

$$EMA_h = \frac{1}{N}\sum_{t=1}^{N}\left|P_{h,t} - O_{h,t}\right| \tag{6}$$

Onde:

- EMA_h: erro médio absoluto do horizonte h;
- N: número total de previsões;
- h: horizonte de previsão avaliado;
- t: instante de tempo;
- $P_{h,t}$: valor previsto no instante t;
- $O_{h,t}$: valor observado no instante t.

3.2.4.2 Coeficiente de *Nash-Sutcliffe* (CNS)

O CNS traduz a eficiência de realizar previsões mais acertadas nas cheias, ou seja, quando o aproveitamento se encontra com vazões bastante elevadas (Rocha et al., 2007). De acordo com Milléo (2020), o CNS corresponde a um índice adimensional que varia de -∞ a 1, sendo interpretado da seguinte forma:

a) CNS = 1: combinação perfeita dos dados estimados com os dados observados;
b) CNS = 0: as previsões do modelo são tão precisas quanto a média dos valores observados;

c) CNS < 0: a média dos valores observados é melhor preditora que o modelo.

O valor CNS é calculado pela equação (7).

$$CNS_h = 1 - \frac{\sum_{t=1}^{N}\left(P_{h,t} - O_{h,t}\right)^2}{\sum_{t=1}^{N}\left(O_{h,t} - \overline{O_{h,t}}\right)^2} \quad (7)$$

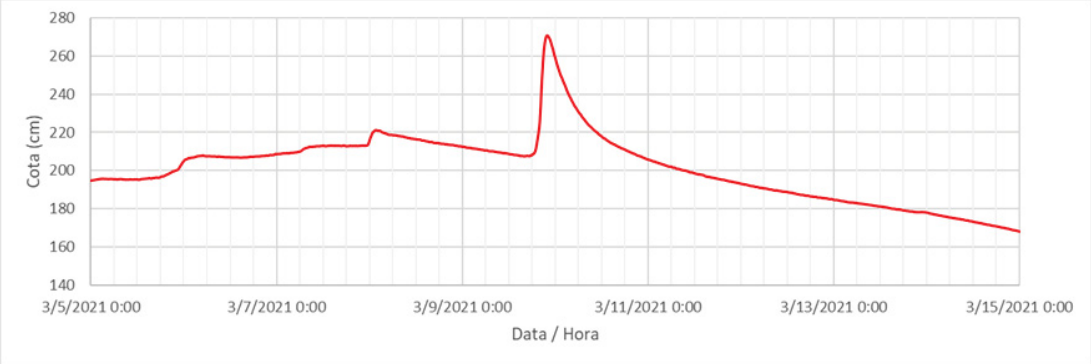
Onde:

- CNS_h: Coeficiente de *Nash-Sutcliffe* do horizonte h;
- N: número total de previsões;
- h: horizonte de previsão avaliado;
- t: instante de tempo;
- $P_{h,t}$: valor previsto no instante t;
- $O_{h,t}$: valor observado no instante t;
- $\overline{O_{h,t}}$: média dos valores observados no horizonte h.

3.2.5 Intervalo de validação

Com o intuito de avaliar o potencial de aplicação do modelo treinado, utilizar-se-á um intervalo contido no conjunto de dados de testes, que apresente um crescimento repentino nas leituras e que não tenha sido utilizado no treinamento do modelo. Logo, o valor de leitura de crescimento repentino teve seu ápice em 09 de março de 2021 às 22hrs. Na FIGURA 21 é apresentado parte da série hidrológica da estação de Porto Amazonas que contém o pico de vazão de interesse.

FIGURA 21 - TRECHO A SER UTILIZADO NA VALIDAÇÃO PRÁTICA DO MODELO



FONTE: O autor (2022).

O processo de avaliação procederá da seguinte maneira:

a) Definir a quantidade de horas anterior ao momento de interesse T;
b) Calcular o intervalo de datas a ser utilizado;
c) Utilizar o modelo para prever de forma progressiva os valores para o intervalo definido em a).

Por exemplo, para um modelo com *time_steps* de valor dez, a quantidade de horas (h) definidas em a) for seis, em b) o intervalo de datas calculado será dado por (T-h-*time_steps*) → [(09/03/2021 22h) – 6h – 10h], logo, o intervalo será de 09/03/2021 6h a 09/03/2021 22h.

No intervalo de 09/03 6h a 09/03 15h será a primeira entrada da rede para predizer o dado 09/03 16h, este valor predito é acrescentado ao novo intervalo de entrada 09/03 7h a 09/03 16h para prever o valor de 09/03 17h, e, assim, progressivamente até atingir o momento 09/09/2021 22h.

4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS

É importante citar que cada experimento foi conduzido de forma separada um do outro, ou seja, cada modelo teve treinamento exclusivo e não foram compartilhados os dados de treinamento entre os experimentos. Por exemplo, quando mencionado que se acrescentaram dados sintéticos de chuva, esses dados foram acrescentados nas amostras de entrada de um novo modelo.

4.1 CONVLSTM

Este experimento consiste no treinamento de uma rede ConvLSTM estruturada conforme o QUADRO 5, sendo esta estrutura aplicadas a todos os modelos para efeitos de comparação. Assim, utilizou-se os dados:

- 1 modelo digital de terreno (SRTM);
- 1 raster do potencial de retenção de águas pluviais (CN);
- 24.840 dados de chuva (SIPREC);
- 24.840 leituras de cota fluvial para cada estação hidrológica “boa”.

QUADRO 5 - ESTRUTURA DA REDE ("*Model*")

	Layers	***Filters***	***Kernel***	***Padding***	***Activation***	***Return sequences***
1	ConvLSTM2D	64	5 x 5	*same*	Relu	*True*
2	*BatchNormalization*	-	-	-	-	-
3	ConvLSTM2D	64	3 x 3	*same*	Relu	*True*
4	*BatchNormalization*	-	-	-	-	-
5	ConvLSTM2D	32	3 x 3	*same*	Relu	*True*
6	*BatchNormalization*	-	-	-	-	-
7	ConvLSTM2D	32	3 x 3	*same*	Relu	*True*
8	*BatchNormalization*	-	-	-	-	-
9	ConvLSTM2D	1	3 x 3	*same*	Relu	*False*
10	*BatchNormalization*	-	-	-	-	-
11	Conv3D	1	4 x 1 x 1	*same*	Relu	-
Time_steps = 14 Input shape: 14 x 4 x 36 x 55 Loss: Distância Euclidiana (função customizada) Optimizer: adamax						

FONTE: O autor (2022).

Das 24.840 leituras, selecionaram-se 90% (iniciais) para compor os dados de treinamento e 10% (finais) para os dados de teste e validação da qualidade do treinamento. Logo, o tamanho destes dados é exibido no QUADRO 6:

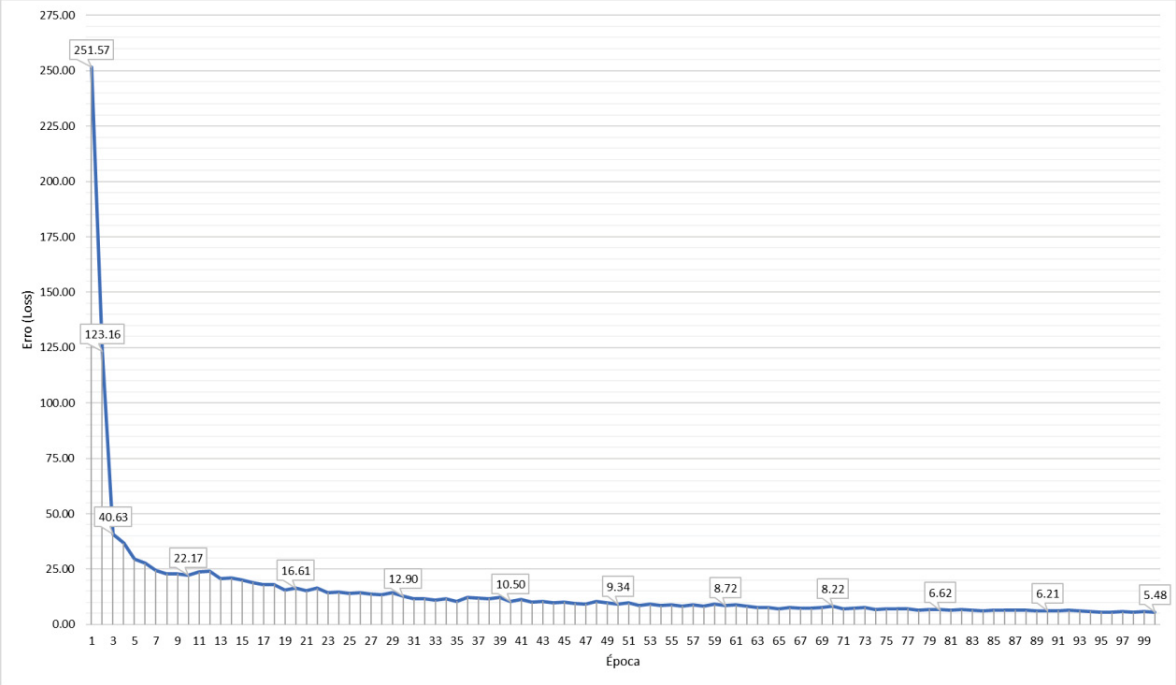
QUADRO 6 - TAMANHO DAS MATRIZES UTILIZADAS NO TREINAMENTO

Tipo	**Amostras**	***Time_steps***	**Camadas**	**Linhas**	**Colunas**
Treino X	22.354	14	4	36	55
Treino Y	22.354	-	1	36	55
Teste X	2.458	14	4	36	55
Teste Y	2.458	-	1	36	55

FONTE: O autor (2022).

A evolução do treinamento é apresentada na FIGURA 22. O decremento do valor do erro (*loss*) ao longo do tempo indica uma boa a qualidade do treinamento. Para este experimento, o cálculo do erro é realizado por uma função energia definida pela distância euclidiana, caracterizada pela raiz quadrada da soma do quadrado dos resíduos, na qual os resíduos são calculados pela diferença entre o dado predito (n+1) de X e seu valor correspondente em Y. Cabe ressaltar que os resultados deste experimento foram obtidos com 400 épocas de treino, visto que cada época de treinamento é muito onerosa, não foi possível ter épocas maiores em tempo hábil para a entrega deste trabalho, talvez mais épocas fornecessem resultados melhores.

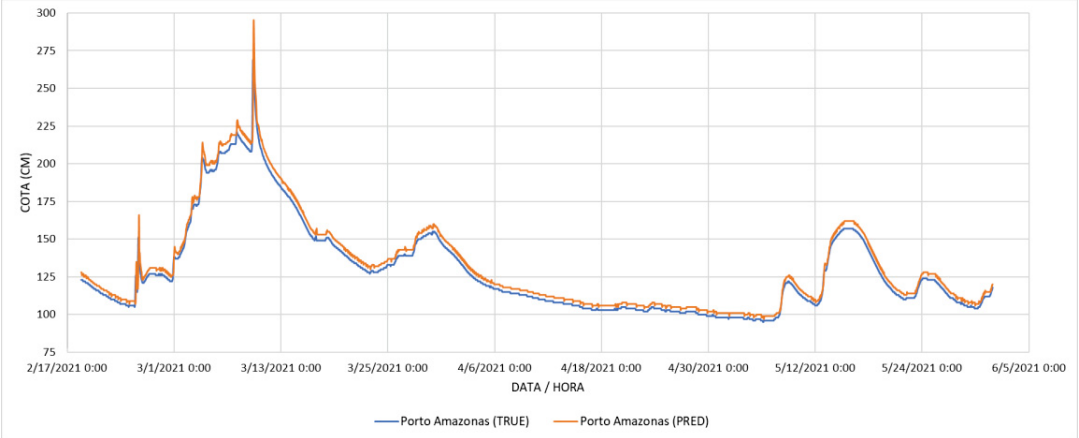
FIGURA 22 - DIMINUIÇÃO DO ERRO NO TREINAMENTO DO MODELO CONVLSTM



FONTE: O autor (2022).

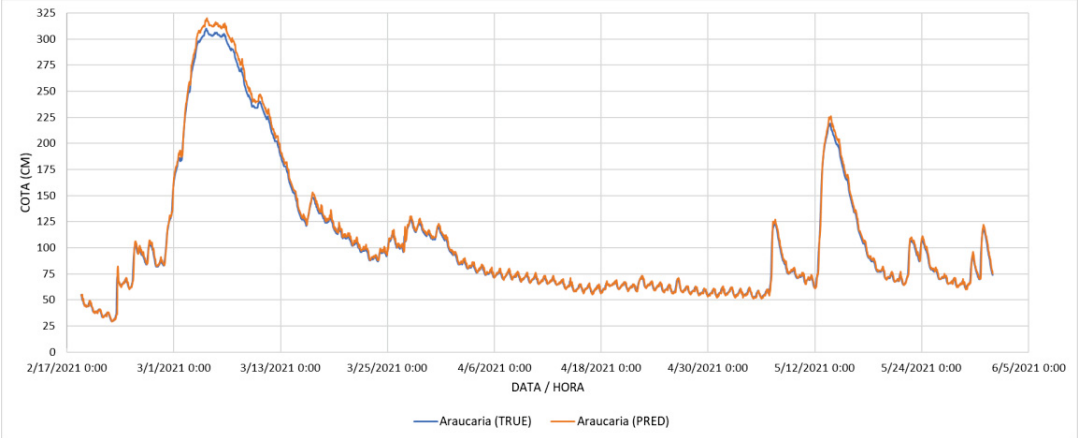
A validação do treinamento é dada pela utilização dos dados de teste. Estes valores não são utilizados pela rede para aprender, mas sim para visualizar se os dados previstos (n+1) são compatíveis com seus respectivos valores verdadeiros. Plotar em gráfico os dados previstos e os dados observados possibilitam uma análise visual do comportamento da rede, assim, nas FIGURAS 23 a 28 são mostrados esses valores para cada estação hidrológica.

FIGURA 23 - CONVLSTM: VALORES DO TESTE – PORTO AMAZONAS



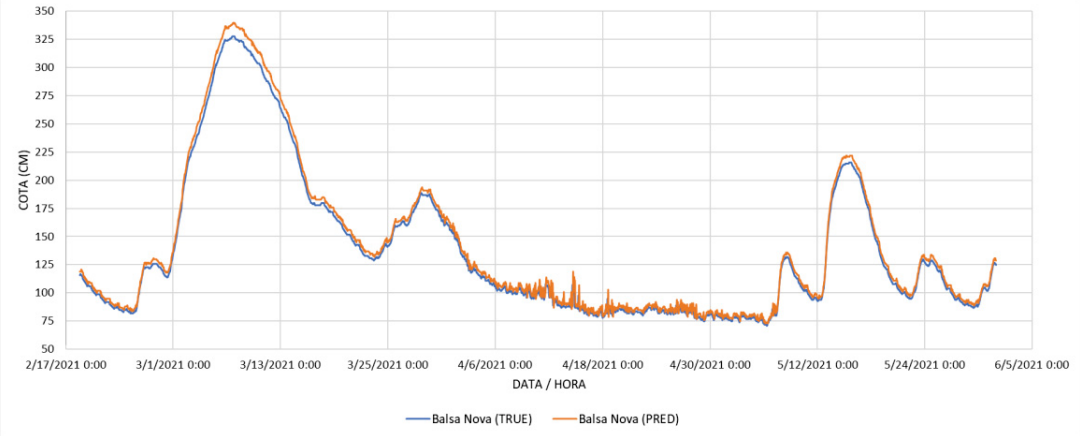
FONTE: O autor (2022).

FIGURA 24 - CONVLSTM: VALORES DO TESTE – ARAUCÁRIA



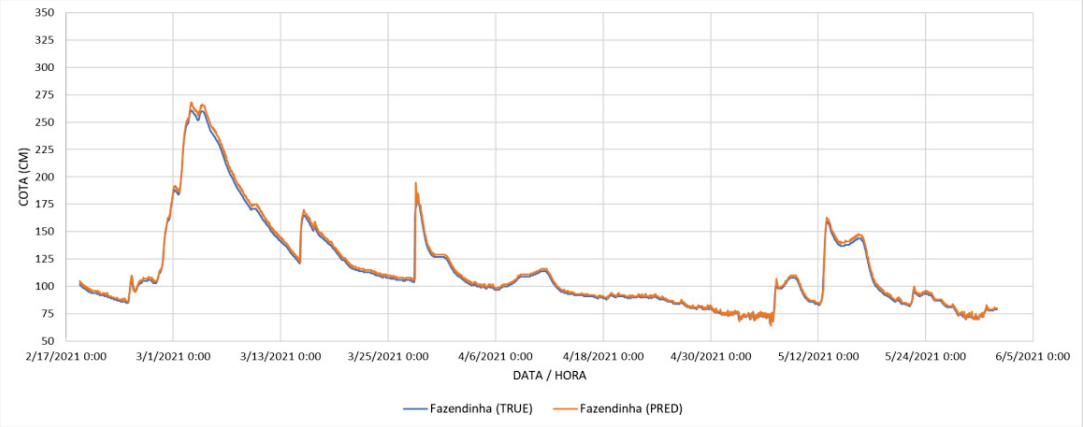
FONTE: O autor (2022).

FIGURA 25 - CONVSLTM: VALORES DO TESTE - BALSA NOVA



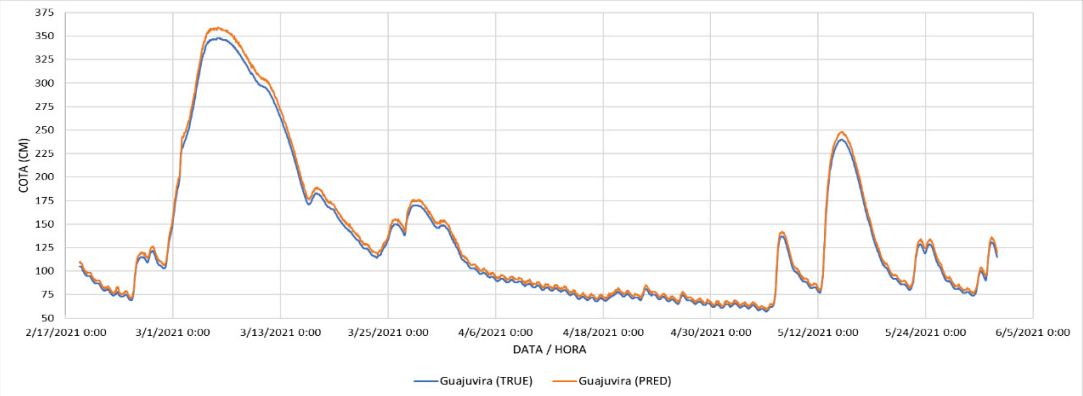
FONTE: O autor (2022).

FIGURA 26 - CONVLSTM: VALORES DO TESTE – FAZENDINHA



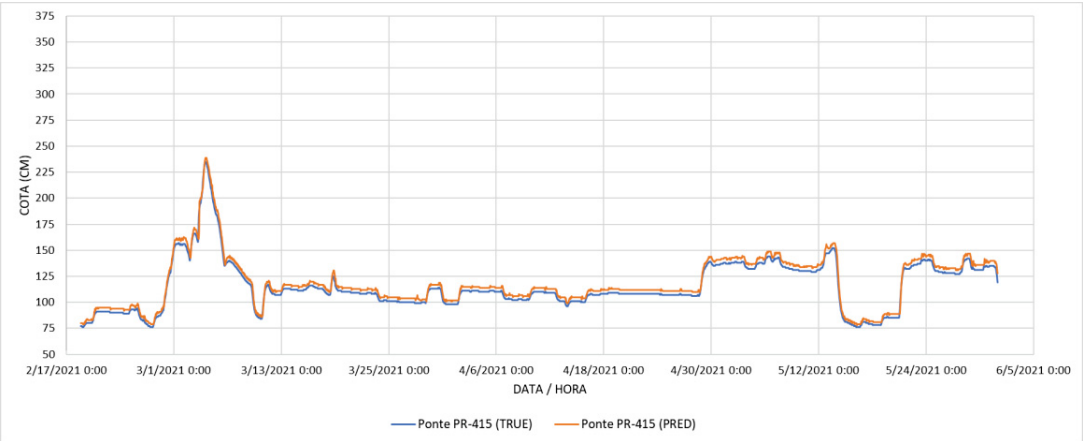
FONTE: O autor (2022).

FIGURA 27 - CONVLSTM: VALORES DO TESTE – GUAJUVIRA



FONTE: O autor (2022).

FIGURA 28 - CONVLSTM: VALORES DO TESTE - PONTE PR-415



FONTE: O autor (2022).

Visualmente, a rede teve um bom desempenho no treinamento. Mas quanto a capacidade de predizer em situações prática? Por isto, também é preciso dispor de informações quantitativas, também chamadas de métrica neste trabalho.

4.1.1 Validação ConvLSTM

Conforme procedimento descrito no item 3.2.4, o processo de validação consiste no cálculo da métricas para um horizonte de previsão, neste processo se objetiva avaliar a qualidade do modelo em fazer predições. Neste processo, os valores preditos (n+1) são inseridos na amostra de entrada para calcular o valor (n+2), e assim por diante.

No QUADRO 7 é apresentado os resultados obtidos, no qual T corresponde ao momento em que a leitura de cota fluvial atinge seu ápice dentro do intervalo escolhido para validação.

QUADRO 7 - VALORES PREDITOS POR CONVLSTM

Momento	**P. Amazonas**		**Araucária**		**Balsa Nova**		**Fazendinha**		**Guajuvira**		**Ponte PR 415**	
	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito
T - 3h	210	211.63	236	243.38	311	325.25	171	177.25	311	319.25	108	106.88
T - 2h	219	220.25	235	257.75	312	348.50	171	191.50	310	342.00	104	102.81
T - 1h	247	239.75	236	285.00	312	401.25	171	222.00	310	397.75	101	100.25
T	269	278.25	236	345.50	312	473.25	171	283.00	309	480.75	98	95.88
T + 1h	269	332.25	236	413.00	311	539.50	171	351.50	309	544.00	96	93.13
T + 2h	262	398.50	236	472.00	310	594.00	171	388.25	308	598.00	93	88.31
T + 3h	255	484.00	235	542.00	310	641.00	171	440.25	307	653.00	92	89.38
T + 4h	249	575.00	235	592.00	309	672.00	171	486.50	306	676.00	90	95.13
T + 5h	244	646.50	234	620.50	308	670.50	170	471.00	306	638.00	90	101.94

FONTE: O autor (2022).

Percebe-se que há diferença na qualidade de predição entre as estações hidrológicas, como para as estações de Porto Amazonas e Ponte PR 415 o modelo apresentou boa qualidade de predição, e para as demais houve grande diferença entre os valores reais e preditos. Isso pode ser melhor visualizado analisando as métricas apresentadas no QUADRO 8.

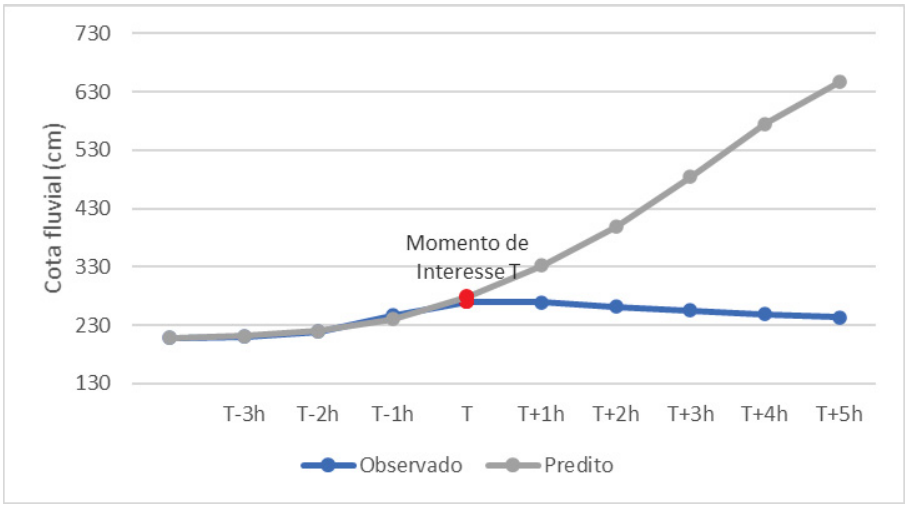
QUADRO 8 - QUALIDADE DE PREDIÇÃO DO MODELO CONVLSTM

Previsão	P. Amazonas		Araucária		Balsa Nova		Fazendinha		Guajuvira		Ponte PR 415	
	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS
1 Hora	1.63	1.00	7.38	-175.2	14.25	< -100	6.25	< -100	8.25	-9.42	-1.13	0.99
2 Hora	1.44	1.00	15.06	< -100	25.38	< -100	13.38	< -100	20.13	< -100	1.16	0.98
3 Hora	3.38	0.97	26.38	< -100	46.67	< -100	25.92	< -100	42.67	< -100	1.02	0.98
4 Hora	4.84	0.95	47.16	< -100	75.31	< -100	47.44	< -100	74.94	< -100	1.30	0.96
5 Hora	16.53	-0.33	73.13	< -100	106.0	< -100	74.05	< -100	107.0	< -100	1.61	0.92
6 Hora	36.52	-5.80	100.3	< -100	135.6	< -100	97.92	< -100	137.5	< -100	2.13	0.82
7 Hora	64.02	-21.06	129.8	< -100	163.5	< -100	122.4	< -100	167.3	< -100	2.20	0.81
8 Hora	96.77	-52.17	158.2	< -100	188.5	< -100	146.5	< -100	192.6	< -100	2.56	0.75
9 Hora	130.74	-99.35	183.6	< -100	207.8	< -100	163.7	< -100	208.1	< -100	3.60	0.35

FONTE: O autor (2022).

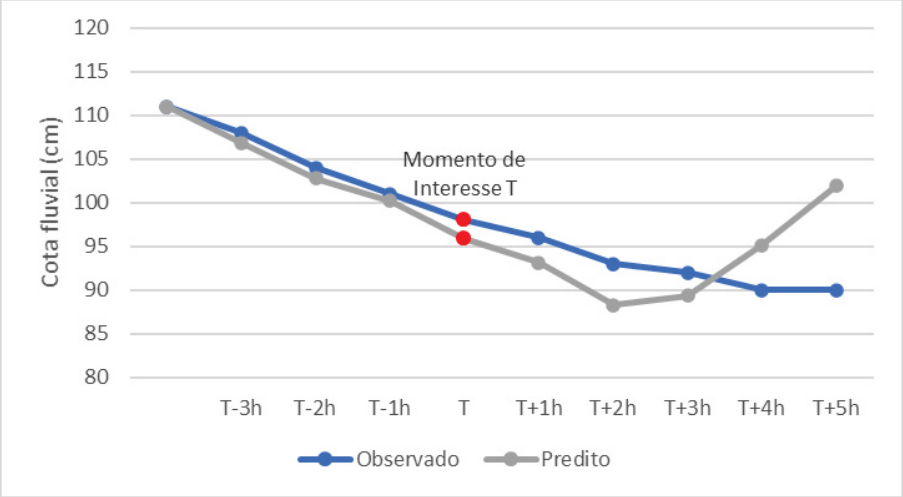
Já era esperado o decaimento na qualidade de predição à medida que o horizonte de predição se estende para o futuro. Entretanto, desejava-se que todas as estações tivessem um comportamento semelhante entre si. Nas FIGURAS 29 e 30 é possível visualizar o comportamento de predição para as duas estações que apresentaram melhor resultado.

FIGURA 29 - PORTO AMAZONAS – CONVLSTM



FONTE: O autor (2022).

FIGURA 30 - PONTE PR 415 – CONVLSTM



FONTE: O autor (2022).

Segundo a definição do CNS, quando seu valor é inferior a zero significa que a média dos valores observados são melhores preditores que o modelo, e portanto, as estações de Araucária, Balsa Nova, Fazendinha e Guajuvira encontram-se nesta situação.

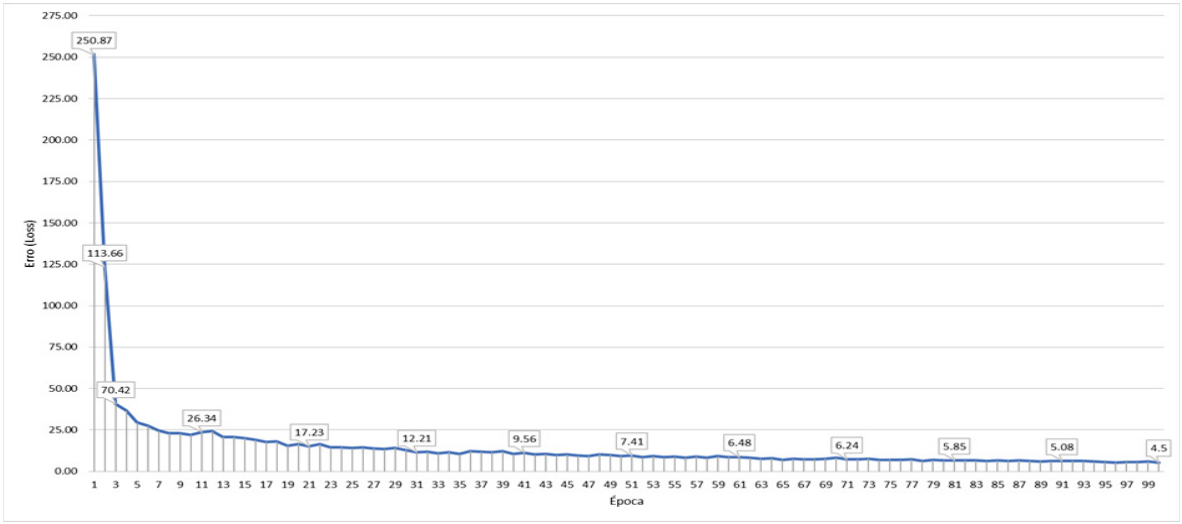
4.2 CONVLSTM COM DADOS SINTÉTICOS DE CHUVA

Seguindo a mesma estrutura do modelo utilizado no experimento 4.1, acrescentaram-se amostras com dados gerados no item 3.2.2.5.1, logo, as amostras de entrada são compostas por:

- 1 modelo digital de terreno (SRTM);
- 1 raster de potencial de retenção de águas pluviais (CN);
- 24.840 dados de chuva (SIPREC);
- 24.840 leituras de cota fluvial de seis estações hidrológicas, aquelas classificadas como boas;
- 290 dados sintéticos de chuva (SIPREC) no maior período de estiagem na área de estudo e 290 dados de leituras de cota fluvial para este período;
- 96 dados sintéticos de chuva (SIPREC) no maior período chuvoso na área de estudo e 96 dados de leituras de cota fluvial para este período.

A evolução do treinamento do modelo deste experimento é exibida na FIGURA 31, sendo expressas 100 épocas de treino, contudo os resultados deste experimento foram obtidos com 350 épocas de treino.

FIGURA 31 - DIMINUIÇÃO DO ERRO NO TREINAMENTO DO MODELO CONVLSTM COM DADOS SINTÉTICOS DE CHUVA



FONTE: O autor (2022).

4.2.1 Validação ConvLSTM com dados sintéticos de chuva

Os resultados de predição do Modelo ConvLSTM com dados sintéticos de chuva estão exibidos no QUADRO 9.

QUADRO 9 - VALORES PREDITOS POR CONVLSTM COM DADOS SINTÉTICOS DE CHUVA

Momento	P. Amazonas		Araucária		Balsa Nova		Fazendinha		Guajuvira		Ponte PR 415	
	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito
T - 3h	210	210.63	236	247.13	311	338.25	171	174.63	311	330.50	108	110.31
T - 2h	219	214.25	235	263.25	312	363.00	171	181.13	310	355.75	104	109.63
T - 1h	247	219.00	236	289.00	312	385.75	171	191.50	310	379.25	101	109.25
T	269	224.87	236	325.50	312	409.50	171	206.13	309	406.25	98	108.50
T + 1h	269	232.25	236	360.75	311	434.25	171	226.88	309	430.75	96	107.94
T + 2h	262	241.63	236	392.00	310	455.25	171	248.50	308	448.50	93	107.69
T + 3h	255	253.25	235	419.00	310	471.25	171	266.75	307	465.50	92	107.50
T + 4h	249	268.25	235	439.00	309	483.25	171	281.75	306	479.50	90	107.25
T + 5h	244	288.00	234	460.00	308	494.00	170	295.75	306	490.00	90	107.00

FONTE: O autor (2022).

Assim, como no experimento anterior, neste duas estações tiveram resultados melhores que as demais, sendo, a estação de Porto Amazonas e a Ponte PR 415. As métricas da qualidade deste experimento são mostradas no QUADRO 10.

QUADRO 10 - QUALIDADE DE PREDIÇÃO DO MODELO CONVLSTM COM DADOS SINTÉTICOS DE CHUVA

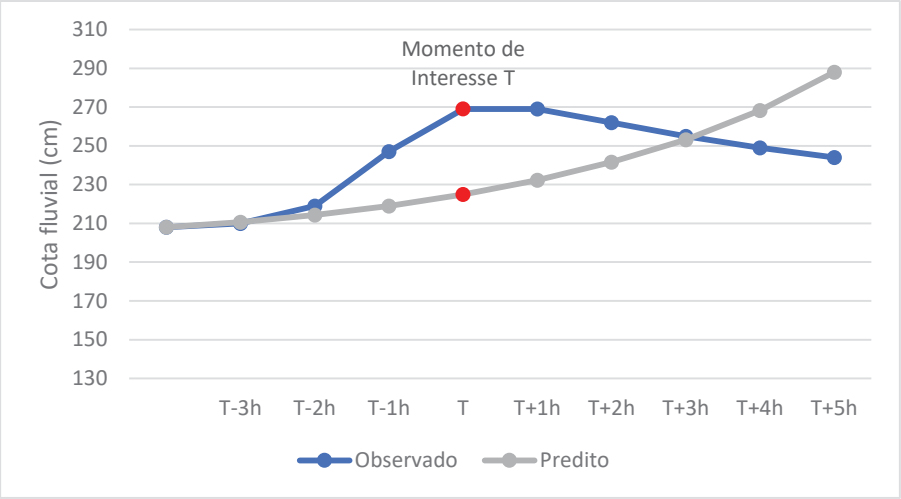
Previsão	P. Amazonas		Araucária		Balsa Nova		Fazendinha		Guajuvira		Ponte PR 415	
	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS
1 Hora	0.63	1.00	11.13	< -100	27.25	< -100	3.63	< -100	19.50	-57.22	2.31	0.96
2 Hora	2.69	0.99	19.69	< -100	39.13	< -100	6.88	< -100	32.63	< -100	3.97	0.79
3 Hora	11.13	0.63	30.79	< -100	50.67	< -100	11.42	< -100	44.83	< -100	5.40	0.45
4 Hora	19.38	-0.04	45.47	< -100	62.38	< -100	17.34	< -100	57.94	< -100	6.67	-0.12
5 Hora	22.85	-0.31	61.33	< -100	74.55	< -100	25.05	< -100	70.70	< -100	7.73	-0.85
6 Hora	22.44	-0.35	77.11	< -100	86.33	< -100	33.79	< -100	82.33	< -100	8.89	-1.76
7 Hora	19.48	-0.33	92.38	< -100	97.04	< -100	42.64	< -100	93.21	< -100	9.83	-2.51
8 Hora	19.45	-0.43	106.3	< -100	106.7	< -100	51.16	< -100	103.3	< -100	10.76	-2.98
9 Hora	22.18	-1.00	119.6	< -100	115.5	< -100	59.44	< -100	112.2	< -100	11.45	-3.28

FONTE: O autor (2022).

O CNS tem seu valor ideal igual a 1.00 e quanto mais próximo de 1.00 for o valor do CNS calculado para o modelo, melhor será sua capacidade de predição. Boa parte da comunidade de Hidrologia diz que um modelo é bom em predizer quando o valor do CNS é maior que 0.60, deste modo, este modelo apresentou boa qualidade para predições de até três horas futuras para as duas estações (Porto Amazonas e Ponte PR 415).

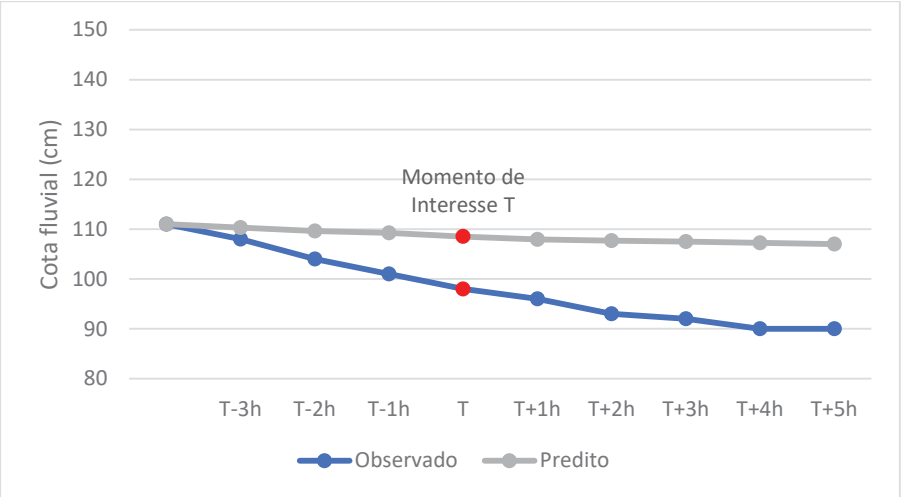
Nas FIGURAS 32 e 33 são apresentados graficamente os resultados deste experimento.

FIGURA 32 - PORTO AMAZONAS - CONVLSTM COM DADOS SINTÉTICOS DE CHUVA



FONTE: O autor (2022).

FIGURA 33 - PONTE PR 415 - CONVLSTM COM DADOS SINTÉTICOS DE CHUVA



FONTE: O autor (2022).

Analisando a FIGURA 32, percebe-se uma diminuição na amplitude dos valores preditos quando comparados ao experimento anterior. Já, na figura 33 nota-se que o modelo apresentou resultado inferior na qualidade de predição para a estação Ponte PR 415 em comparação com os resultados mostrados na FIGURA 30.

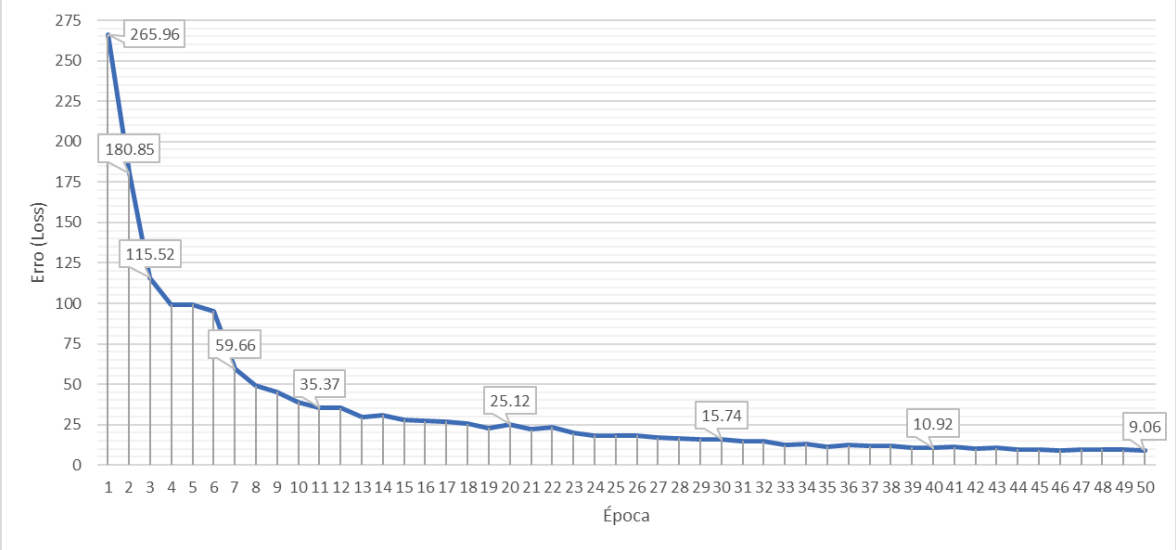
4.3 CONVLSTM COM DADOS SINTÉTICOS DE COTAS FLUVIAIS

Para o treinamento deste modelo, acrescentou-se os dados das estações médias calculadas no item 3.2.2.4, logo, as amostras de entrada são compostas por:

- 1 modelo digital de terreno (SRTM);
- 1 raster de potencial de retenção de águas pluviais (CN);
- 24.840 dados de chuva (SIPREC);
- 24.840 leituras de cotas fluviais para dez estações hidrológicas, ou seja, todas as estações que compõem o QUADRO 3.

Semelhante aos outros experimentos, o processo de treinamento (FIGURA 34) teve um comportamento bom e de rápida convergência, entretanto, percebeu-se que quanto maior o número de épocas melhor é o refinamento do modelo na capacidade de predição.

FIGURA 34 - DIMINUIÇÃO DO ERRO NO TREINAMENTO DO MODELO CONVLSTM COM DADOS SINTÉTICOS DE COTAS FLUVIAIS



FONTE: O autor (2022).

4.3.1 Validação ConvLSTM com dados sintéticos de cotas fluviais

No QUADRO 11 estão expressos os resultados deste experimento.

QUADRO 11 - VALORES PREDITOS POR CONVLSTM + DADOS SINTÉTICOS DE COTAS FLUVIAIS

Momento	P. Amazonas		Araucária		Balsa Nova		Fazendinha		Guajuvira		Ponte PR 415	
	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito	Real	Predito
T - 3h	210	225.25	236	278.00	311	381.75	171	197.75	311	360.25	108	121.38
T - 2h	219	225.25	235	274.25	312	381.75	171	195.25	310	355.75	104	120.13
T - 1h	247	225.50	236	269.75	312	382.25	171	196.88	310	356.25	101	115.44
T	269	262.25	236	383.25	312	527.50	171	286.00	309	508.25	98	128.25
T + 1h	269	262.25	236	374.00	311	528.00	171	277.75	309	496.00	96	127.94
T + 2h	262	253.38	236	332.75	310	517.50	171	258.50	308	461.75	93	119.19
T + 3h	255	344.00	235	547.00	310	666.50	171	466.25	307	658.50	92	124.13
T + 4h	249	340.75	235	543.00	309	660.00	171	447.25	306	640.00	90	129.50
T + 5h	244	294.50	234	443.75	308	666.00	170	392.50	306	614.00	90	124.13

FONTE: O autor (2022).

Analisando-se as métricas apresentadas no QUADRO 12, notou-se que a estação de Porto Amazonas apresentou resultados superiores a 0.6 para o CNS em um horizonte de 6h, o que é um valor bom segundo a maioria dos hidrólogos.

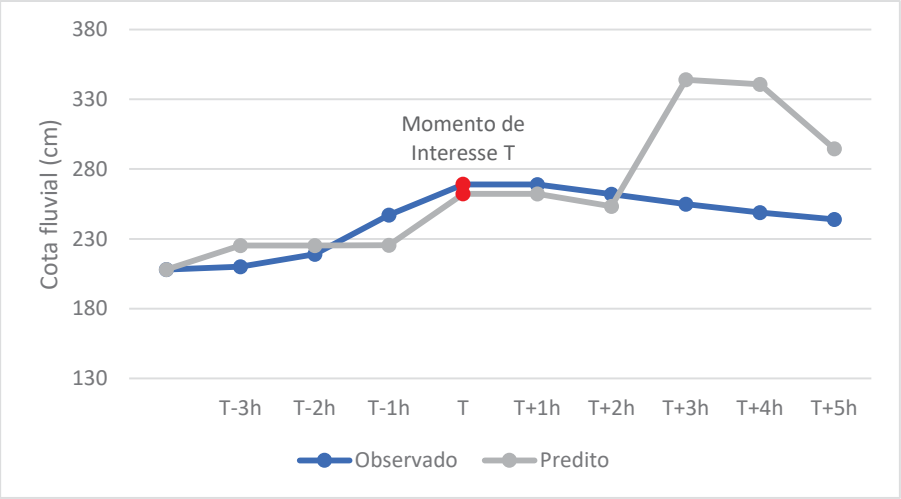
QUADRO 12 - QUALIDADE DE PREDIÇÃO DO MODELO CONVLSTM + DADOS SINTÉTICOS DE COTAS FLUVIAIS

Previsão	P. Amazonas		Araucária		Balsa Nova		Fazendinha		Guajuvira		Ponte PR 415	
	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS	EMA	CNS
1 Hora	15.25	0.83	42.00	< -100	70.75	< -100	26.75	< -100	49.25	< -100	13.38	-0.45
2 Hora	10.75	0.87	40.63	< -100	70.25	< -100	25.50	< -100	47.50	< -100	14.75	-1.52
3 Hora	14.33	0.66	38.33	< -100	70.25	< -100	25.63	< -100	47.08	< -100	14.65	-2.39
4 Hora	12.44	0.71	65.56	< -100	106.6	< -100	47.97	< -100	85.13	< -100	18.55	-7.13
5 Hora	11.30	0.74	80.05	< -100	128.6	< -100	59.73	< -100	105.5	< -100	21.23	-12.4
6 Hora	10.85	0.73	82.83	< -100	141.8	< -100	64.35	< -100	113.5	< -100	22.05	-14.7
7 Hora	22.02	-1.59	115.6	< -100	172.5	< -100	97.34	< -100	147.5	< -100	23.49	-17.5
8 Hora	30.73	-4.05	139.6	< -100	194.8	< -100	119.7	< -100	170.8	< -100	25.49	-19.9
9 Hora	32.93	-4,78	147.4	< -100	212.9	< -100	131.1	< -100	186.1	< -100	26.45	-20.5

FONTE: O autor (2022).

Tal resultado bom pode ser visualizado na FIGURA 35, no qual além do momento de interesse T, o modelo conseguiu duas cotas preditas com valores próximos aos observados.

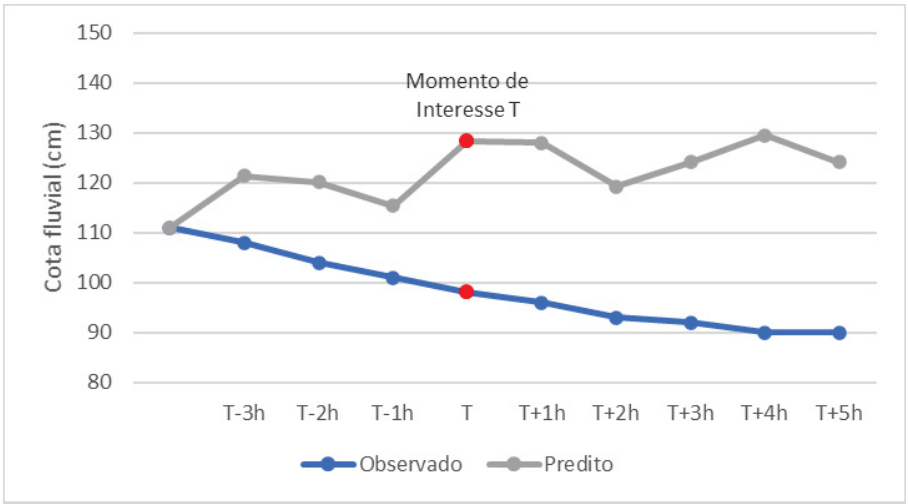
FIGURA 35 - PORTO AMAZONAS - CONVLSTM + DADOS SINTÉTICOS DE COTAS FLUVIAIS



FONTE: O autor (2022).

Entretanto, para a estação Ponte PR 415 os valores preditos tiveram um comportamento inverso aos valores observados, conforme é ilustrado na FIGURA 36.

FIGURA 36 - PONTE PR 415 - CONVLSTM + DADOS SINTÉTICOS DE COTAS FLUVIAIS



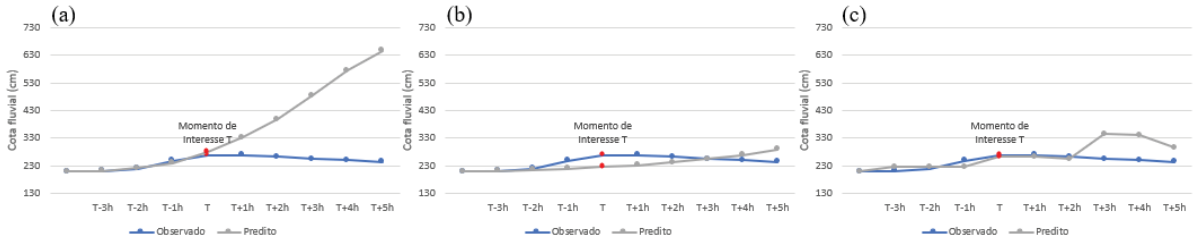
FONTE: O autor (2022).

4.4 COMPARAÇÃO DOS RESULTADOS PARA A ESTAÇÃO DE PORTO AMAZONAS

Como a estação hidrológica de Porto Amazonas é a estação de maior interesse neste estudo e apresentou resultados passíveis de serem comparados entre si, esta foi utilizada como referência

de análise. Na FIGURA 34 é realizada a comparação dos resultados dos três experimentos deste trabalho.

FIGURA 37 - COMPARAÇÃO DOS GRÁFICOS DE PREDIÇÃO DA ESTAÇÃO PORTO AMAZONAS



FONTE: O autor (2022).
LEGENDA: a) Modelo ConvLSTM.
b) Modelo ConvLSTM com dados sintéticos de chuva.
c) Modelo ConvLSTM com dados sintéticos de cotas

Nota-se que a inserção de dados sintéticos no treinamento possibilitou uma melhor qualidade de predição. Em a) a predição possui uma qualidade muito boa até o momento T, porém a partir deste ponto ela começa a apresentar valores cada vez mais altos para a cota, chegando a um erro de 4 metros no momento T+5h. Já em b) o acréscimo de dados sintéticos de chuva nas amostras de treinamento freou o aumento significativo dos valores de cotas se comparado a a), tendo no momento T+5h uma diferença de 44 cm entre o valor predito e o valor real. Em c) percebe-se uma melhor predição, principalmente até o momento T+2h, que corresponde a um horizonte de predição para 6 hrs.

5 CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

5.1 CONSIDERAÇÕES FINAIS

Os experimentos mostraram um caminho promissor e não definitivo para a metodologia de treino proposta. A estação hidrológica de Porto Amazonas foi utilizada como referência por ser o principal objeto de interesse neste estudo, uma vez que por estar na foz da área de estudo, nela estão representadas todas as interdependências das outras estações localizadas a montante.

Assim, percebeu-se que o incremento de dados sintéticos de chuva e principalmente os de cotas fluviais favoreceu uma melhor capacidade de predição do modelo em um horizonte de até 6h

para a estação de referência. Isso se deve ao fato de ter mais informações intermediárias de leituras no interior da bacia hidrográfica, bem como mais amostras de treino.

Contudo, devido à onerosidade de treinamento dos modelos, eles foram treinados com aproximadamente 350 a 400 épocas e sabe-se que algumas redes com estruturas convolucionais precisam de milhares de épocas de treino. Portanto, se aumentar o número de épocas para os experimentos deste trabalho, pode ser que a qualidade de predição melhore em relação a atual.

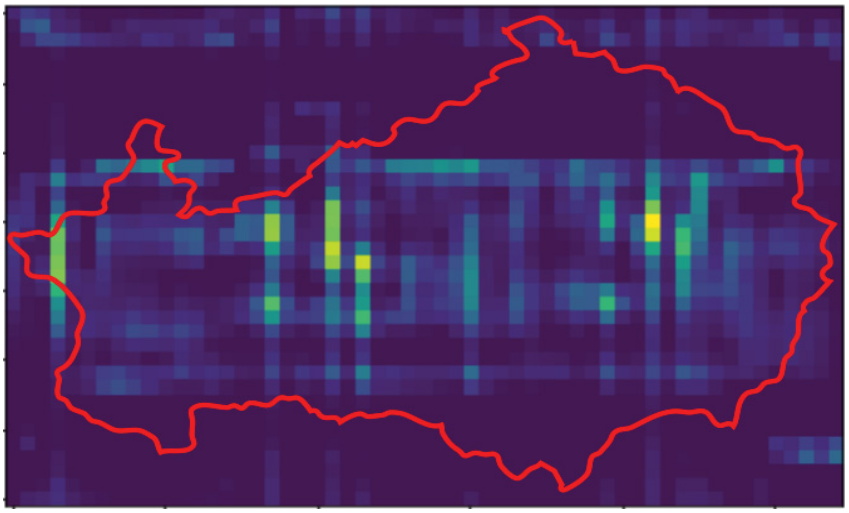
A utilização de uma função energia foi primordial para atingir os resultados obtidos, visto que as funções de erro padrão (*loss*) da biblioteca keras avaliam a imagem como um todo e a função energia utiliza a distância euclidiana apenas nos pixels com representatividade espacial das estações hidrológicas.

5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Como trabalhos futuros recomenda-se:

- Na FIGURA 39, percebe-se que uma saída da rede dispõe de valores maiores que zero não somente na localização das estações, mas sim espalhados por toda a área. Em vermelho está a delimitação da área de estudo deste trabalho e os pixels possuem valor mais alto quanto mais próximo da cor amarela se encontram. Portanto, estudar a saída que a ConvLSTM oferece, pode proporcionar a obtenção de leituras hidrológicas intermediárias.

FIGURA 38 - DADO DE SAÍDA DO EXPERIMENTO 4.4



FONTE: O autor (2022).

- Realização exaustiva de testes com treinamento mais longos, além de diferentes estruturas e funções energia (outros tipos de distâncias, além da Euclidiana);
- Avaliar a qualidade dos dados de chuva oriundos do SIPREC, uma vez que, o SIPREC subestima a rede pluviométrica;
- Acrescentar nos dados de entrada as estações da rede pluviométrica;

REFERÊNCIAS

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO. **Curve Number da Base Hidrográfica Ottocodificada**. Em: Catálogo de Metadados da ANA. Brasília, 2018. Disponível em: < https://metadados.snirh.gov.br/geonetwork/srv/por/catalog.search#/metadata /d1c36d85-a9d5-4f6a-85f 7-71c2dc801a67 >. Acesso em 05 de março de 2021

ALMEIDA, L.; SERRA, J. C. V. Modelos hidrológicos, tipos e aplicações mais utilizadas. **FAE**, Curitiba, v. 20, n. 1, p. 129 - 137, jan./jun. 2017.

BERKHAHN, S.; FUCHS, L.; NEUWEILER, I. An ensemble neural network model for real-time prediction of urban floods. **Journal of Hydrology**, v. 575, p. 743-754, 2019.

BLOG DE GEOGRAFIA. Não paginado. Disponível em: <https://suburbanodigital.blogspot.com/2020/07/o-mapageo-tem-mapas-mentais-de-geografia.html>. Acesso em: 15 fev. 2021.

CALVETTI, L.; BENETI C.; NEUNDORF, R. et al. Quantitative Precipitation Estimation Integrated by Poisson's Equation Using Radar Mosaic, Satellite, and Rain Gauge Network. **Journal of Hydrologic Engineering**, v. 22, n. 5, mai. 2016.

DSA, DATA SCIENCE ACADEMY. **Deep Learning Book**. Disponível em: <http://deeplearningbook.com.br>. Acesso em: 16 out. 2020.

ELSAADANI, M; HABIB, E; ABDELHAMEED, A. M; BAYOUMI, M. Assessment of a Spatiotemporal Deep Learning Approach for Soil Moisture Prediction and Filling the Gaps in Between Soil Moisture Observations. **Frontiers in Artificial Intelligence**. Mar. 2021.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda, 2011.

FAN, F. M.; COLLISCHONN, W.; PAIVA, R. C. D. DE; QUEDI, E.; GAMA, C. H.; SIQUEIRA, V. A. **Previsão de vazões usando modelos hidrológicos**. v. 1.0. Instituto de Pesquisas Hidráulicas, IPH. Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental, UFRGS, Porto Alegre. 2020. 235 f.

HERATH, M. V. V.; CHADALAWADA, J.; BABOVIC, V. Hydrologically Informed Machine Learning for Rainfall-Runoff Modelling: Towards Distributed Modelling. **Hydrology and Earth System Sciences**. Department of Civil and Environmental Engineering, National University of Singapore. Out. 2020. p. 42.

HUA, MOREA; SAILESH, SAMANTA. Multi-criteria decision approach to identify flood vulnerability zones using geospatial technology in the Kemp-Welch Catchment, Central Province, Papua New Guinea. **Applied geomatics**. Società Italiana di Fotogrammetria e Topografia (SIFET), v. 12, n. 4, 2020, p. 427-440, abr. 2020.

KARPATNE, A.; ATLURI, G.; FAGHMOUS, J. H.; STEINBACH, M.; BANERJEE, A.; GANGULY, A.; SHEKHAR, S.; SAMATOVA, N.; KUMAR, V. Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data. **IEEE Transactions on Knowledge and Data Engineering**. v. 29, n. 10, p. 2318-2331, nov. 2017.

LARANJEIRA, CAMILA. Redes Neurais Convolucionais: Deep Learning com PyTorch. **Plataforma ALURA de cursos de tecnologia**. 2021.

MACÊDO, CRISLÂNIO. **Um tratamento para outliers e dados ausentes em séries temporais multivariadas em Redes de Sensores sem Fio**. Disponível em: < https://crislanio-ufc.medium.com/um-tratamento-para-outliers-e-dados-ausentes-em-s%C3%A9ries-temporais -multivariadas-em-redes-de-7e2a83878e69>. Acesso em: 18 de out. de 2021.

MARINHO FILHO, G. M.; ANDRADE, R. DA S.; ZUKOWSKI JUNIOR, J. C.; MAGALHÃES FILHO, L. N. L. Modelos Hidrológicos: conceitos e aplicabilidades. **Revista de Ciências Ambientais**, Canoas, v.6, n.2, p. 35 a 47, 2012.

MILLÉO, CARLA. **Emprego de redes neurais artificiais na previsão climática de temperatura e precipitação no estado do Paraná**. Dissertação (Mestrado em Engenharia Ambiental) – Setor de Tecnologia, Universidade Federal do Paraná, Curitiba, 2020.

MORAES, J.; SCHULER, A.; GROPPO, J.; MILDE, L.; MARTINELLI, L.; GUANDIQUE, M.; VICTORIA, R. Propriedades Físicas dos Solos na Parametrização de um Modelo Hidrológico. **Revista Brasileira de Recursos Hídricos**. v. 8, n. 1, p. 61-70, mar. 2003.

PAGANI, R. N.; KOVALESKI, J. L.; RESENDE, L.M. **Methodi Ordinatio**: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication. Scientometrics, Hungary, n. 105, p. 2019-2135, set. 2015.

POLETO, CRISTIANO. **Bacias Hidrográficas - Estudos Aplicados**. GFM Gráfica & Editora, 2019. 220 p.

PONTE, H. **Training AI with CGI**. Towards data science. 2019. Disponível em: <https://towardsdatascience.com/training-ai-with-cgi-b2fb3ca43929>. Acesso em: 16 out. 2020.

PROJETO MAPBIOMAS – Coleção 5 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil, acessado em 21 mar. 2021, através do link: https://mapbiomas.org/colecoes-mapbiomas-1?cama_set_language=pt-BR

RENNÓ, C. D.; BORMA, L. de S. **Processos Hidrológicos**: Métodos estatísticos aplicados à Hidrologia. Instituto Nacional de Pesquisas Espaciais. Disponível em: <http://www.dpi.inpe.br/~camilo/prochidr/pdf/04estat.pdf>. Acesso em: 02 de junho de 2021.

SAMPAIO, CÁSSIA. Deep Learning com keras. **Plataforma ALURA de cursos de tecnologia**. 2021.

SILVA, I. N. DA; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais: para engenharia e ciências aplicadas**. São Paulo: Artliber, 2010.

SILVEIRA, C. T. da. **Análise de bacias hidrográficas**. Notas de Aula da disciplina Elementos de Geomorfologia. Curso de Engenharia Cartográfica e de Agrimensura. Universidade Federal do Paraná. Curitiba, 2017.

SOARES, M.R.G.J.; FIORI, C. O.; SILVEIRA, C. T.; KAVISKI, E. Eficiência do método *Curve Number* de retenção de águas pluviais. **Mercator**, Fortaleza, v. 16, e16001, 2017. DOI. 10.415. Disponível em: <https://doi.org/10.4215/rm2017.e1600>. Acesso em: 5 jan. 2022.

VALERIANO, MARCIO DE MORISSON. **TOPODATA: Guia para utilização de dados geomorfológicos locais**. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2008.

WENG, L. **Domain Randomization for Sim2Real Transfer**. Lil'Log. 2019. Disponível em: <https://lilianweng.github.io/lil-log/2019/05/05/domain-randomization. html>. Acesso em: 17 out. 2020.

XAVIER, A. **An introduction to ConvLSTM**. Neuronio.AI. Disponível em: <https://medium.com/ neuronio/an-introduction-to-convlstm-55c9025563a7>. Acesso em: 19 dez. 2020.

XIANG, Z.; YAN, J.; DEMIR, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. **Water Resources Research**. v. 56. 2020.

YASEEN, Z. M.; EL-SHAFIE, A.; JAAFAR, O.; AFAN, H. A.; ASYL, K. N. Artificial intelligence based models for stream-flow forecasting: 2000-2015. **Journal of Hydrology**. V. 530, P. 829-844, 2015.