

Universidade Federal de Juiz de Fora
Faculdade de Engenharia
Programa de Pós Graduação em Engenharia Elétrica

Guilherme Marins Maciel

**Arquitetura Híbrida para a Previsão de Afluências em Reservatórios Baseada
no Modelo SMAP e Deep Learning**

Juiz de Fora

2022

Guilherme Marins Maciel

**Arquitetura Híbrida para a Previsão de Afluências em Reservatórios Baseada
no Modelo SMAP e Deep Learning**

Tese apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas de Energia Elétrica.

Orientador: André Luís Marques Marcato, D. Sc.

Coorientador: Ivo Chaves da Silva Junior, D. Sc.

Juiz de Fora

2022

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Maciel, Guilherme Marins.

Arquitetura Híbrida para a Previsão de Afluências em Reservatórios Baseada no Modelo SMAP e Deep Learning / Guilherme Marins Maciel. -- 2022.

87 p. : il.

Orientador: André Luís Marques Marcato

Coorientador: Ivo Chaves da Silva Júnior

Tese (doutorado) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2022.

1. Deep Learning. 2. SMAP. 3. Previsão de vazão. 4. Modelo Híbrido. I. Marcato, André Luís Marques , orient. II. Silva Júnior, Ivo Chaves da , coorient. III. Título.

Guilherme Marins Maciel

Arquitetura Híbrida para a Previsão de Afluências em Reservatórios Baseada no Modelo SMAP e Deep Learning

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas de Energia Elétrica

Aprovada em 03 de maio de 2022.

BANCA EXAMINADORA

Prof. Dr. André Luís Marques Marcato - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Ivo Chaves da Silva Junior - Coorientador

Universidade Federal de Juiz de Fora

Prof. Dr. Edmarcio Antonio Belati

Universidade Federal do ABC

Prof. Dr. Reinaldo Castro Souza

Pontifícia Universidade Católica do Rio de Janeiro

Prof. Dr. João Alberto Passos Filho

Universidade Federal de Juiz de Fora

Prof. Dr. Leonardo Willer de Oliveira

Universidade Federal de Juiz de Fora

Juiz de Fora, 19/04/2022.



Documento assinado eletronicamente por **Andre Luis Marques Marcato, Professor(a)**, em 03/05/2022, às 21:31, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Ivo Chaves da Silva Junior, Professor(a)**, em 03/05/2022, às 21:43, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Leonardo Willer de Oliveira, Professor(a)**, em 03/05/2022, às 21:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Joao Alberto Passos Filho, Professor(a)**, em 04/05/2022, às 10:20, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Edmarcio Antonio Belati, Usuário Externo**, em 06/05/2022, às 16:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **REINALDO CASTRO SOUZA, Usuário Externo**, em 31/05/2022, às 10:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **0747275** e o código CRC **B8807C13**.

RESUMO

Modelos hidrológicos são ferramentas essenciais para prever ou gerar cenários da disponibilidade de recursos hídricos, usados para planejar a operação de sistemas elétricos de curto a longo prazo. No entanto, há um empasse ao escolher um determinado modelo. Modelos complexos, como modelos distribuídos, podem fornecer bons resultados por procedimentos analíticos, otimizações robustas e dados sofisticados. Já modelos mais simples, como os concentrados, oferecem resultados razoáveis com abordagens de ajuste muito mais fáceis. Para melhorar a qualidade de modelos mais simples, este trabalho propõe o acoplamento do modelo hidrológico Soil Moisture Accounting Procedure (SMAP) com uma arquitetura Deep Learning baseada em camadas Conv3D-LSTM. Na metodologia proposta, o SMAP é primeiramente otimizado para obter parâmetros gerais da bacia hidrográfica. Este modelo recebe um ajuste local, baseado no algoritmo Twiddle, usado como entrada para a arquitetura Conv3D-LSTM. Este modelo de estimativa de caixa cinza pode gerar resultados rápidos e precisos. Estudos, visando prever a vazão natural com sete dias de antecedência, foram realizados em duas grandes hidrelétricas brasileiras para validar o método. Os resultados obtidos pela arquitetura híbrida são melhores que os obtidos com as técnicas desacopladas.

Palavras-chave: SMAP, Deep Learning, Previsão de Vazão, Modelo Híbrido, Conv3D, LSTM

ABSTRACT

Hydrological models are essential tools to generate scenarios or forecast water resources availability, which are used to plan the short-term electrical systems operation. However, there is a trade-off when choosing a given model. Complex models may provide good results depending on very complicated analytical and optimization procedures beyond sophisticated data, whereas simpler models offer reasonable results with much more amenable tuning approaches. To improve the quality of simpler models this work proposes the coupling of the Soil Moisture Accounting Procedure (SMAP) hydrological model with a Deep Learning architecture based on Conv3D-LSTM. In the proposed methodology, the SMAP is first optimized to obtain general parameters of the hydrographic basin. This optimized model's output is used as input to the Conv3D-LSTM estimator to provide the final results. This gray estimator model can generate fast and accurate results. Studies with the goal of forecast the natural flow seven days ahead are carried out for two large Brazilian hydroelectric plants to validate the method. The results obtained by the architecture are better than those obtained with decoupled techniques.

Key-words: SMAP, Deep Learning, Runoff Forecasting, Hybrid Model, Conv3D, LSTM

LISTA DE ILUSTRAÇÕES

Figura 1 – Matrizes Elétricas Mundial e Brasileira	12
Figura 2 – Dilema do Operador	13
Figura 3 – Ilustração do espectro cinzento em modelos.	17
Figura 4 – Modelo hidrológico concentrado vs distribuído.	18
Figura 5 – Modelo hidrológico determinístico vs estocástico.	19
Figura 6 – Possibilidades de acoplamento SMAP-DNN	26
Figura 7 – Fluxograma da metodologia híbrida proposta.	27
Figura 8 – Diagrama do modelo SMAP/ONS.	28
Figura 9 – Representação da calibração do SMAP-ONS	35
Figura 10 – Fluxograma do processamento dos dados de chuva	37
Figura 11 – Serviço do INMET para fornecimento de observações meteorológicas	38
Figura 12 – Obtenção da evapotranspiração sobre a área de escoamento.	39
Figura 13 – Sistema da Agência Nacional de Águas para o fornecimento de vazões no SIN.	40
Figura 14 – Ajuste local do Modelo SMAP	43
Figura 15 – Comportamento das restrições da otimização local no SMAP e no modelo Híbrido	45
Figura 16 – Exemplo de RNA com camadas densas.	47
Figura 17 – Ilustração dos processos convolucionais.	48
Figura 18 – Exemplo da aplicação dos filtros na camada convolucional 2D	49
Figura 19 – Parâmetros Stride da Convolução	49
Figura 20 – Ilustração de uma camada LSTM	50
Figura 21 – Parâmetros da unidade LSTM	51
Figura 22 – Esquema da arquitetura desenvolvida	53
Figura 23 – Estrutura da entrada de precipitação.	54
Figura 24 – Cross-Validation	55
Figura 25 – Arquitetura de Software	59
Figura 26 – Localização das regiões de estudo e áreas de drenagem das bacias hidrográficas: UHE Peixe Angical e UHE Mascarenhas.	65
Figura 27 – Comparação entre o escoamento observado e calculado pelo modelo SMAP na etapa de calibração. (a) and (b) UHE Mascarenhas; (c) and (d) UHE Peixe Angical.	66
Figura 28 – Ajuste Local - UHE Mascarenhas	67
Figura 29 – Ajuste Local - UHE Peixe Angical.	68
Figura 30 – Tanques Ajustados - UHE Mascarenhas.	69
Figura 31 – Tanques Ajustados - UHE Peixe Angical	70
Figura 32 – Comparação entre SMAP, DNN and DNN-SMAP. (a) e (b) UHE Mas- carenhas; (c) e (d) UHE Peixe Angical.	71

Figura 33 – Simulação de Previsão de Escoamento para UHE Mascarenhas.	72
Figura 34 – Simulação de Previsão de Escoamento para UHE Peixe Angical.	72
Figura 35 – Gráfico de dispersão de dados de escoamento observados e previstos. (a) UHE Mascarenhas; (b) UHE Peixe Angical.	73
Figura 36 – Linearização das curvas de vazão	75
Figura 37 – Histograma normalizado das tendências de variação da vazão observada na UHE Mascarenhas	76
Figura 38 – Histograma normalizado das tendências de variação da vazão observada na UHE Peixe Angical.	76
Figura 39 – Simulação de Previsão Estocástica para UHE Mascarenhas.	77
Figura 40 – Simulação de Previsão Estocástica para UHE Peixe Angical.	78
Figura 41 – Possibilidade de acoplamento SMAP-DNN Série e Paralelo.	81
Figura 42 – SMAP-GRID	82

LISTA DE TABELAS

Tabela 1 – Trabalhos de otimização do SMAP	22
Tabela 2 – Parâmetros calibráveis do modelo SMAP-ONS.	32
Tabela 3 – Vazões fornecidas pela ANA/ONS.	41
Tabela 4 – Limites das Métricas	57
Tabela 5 – Principais Pacotes utilizados	58
Tabela 6 – Arquitetura de cada bacia	65
Tabela 7 – Resultados dos Parâmetros Globais do SMAP.	66
Tabela 8 – Desempenho calibração global do SMAP.	67
Tabela 9 – Desempenho da calibração local do SMAP.	67
Tabela 10 – Resultados Estatísticos das Previsões.	73
Tabela 11 – Matriz Confusão UHE Mascarenhas	74
Tabela 12 – Matriz Confusão UHE Peixe	75
Tabela 13 – Matriz Confusão UHE Mascarenhas - Chuva sem adição de erros	78
Tabela 14 – Matriz Confusão UHE Mascarenhas - Chuva subestimada.	78
Tabela 15 – Matriz Confusão UHE Mascarenhas - Chuva superestimada	78
Tabela 16 – Matriz Confusão UHE Peixe - Chuva sem adição de erros	79
Tabela 17 – Matriz Confusão UHE Peixe - Chuva subestimada	79
Tabela 18 – Matriz Confusão UHE Peixe - Chuva superestimada	79

LISTA DE ABREVIATURAS E SIGLAS

ANA	<i>Agência Nacional de Águas</i>
BAT	<i>Bat Algorithm</i>
CPTEC	<i>Centro de Previsão de Tempo e Estudos Climáticos</i>
DL	<i>Deep Learning</i>
DE	<i>Differential Evolution</i>
DNN	<i>Deep Neural Network</i>
GA	<i>Genetic Algorithm</i>
GWO	<i>Grey Wolf Optimizer</i>
INPE	<i>Instituto Nacional de Pesquisas Espaciais</i>
MAE	<i>Mean Absolute Error</i>
MOPSO	<i>Multi-Objective Particle Swarm Optimization</i>
MSE	<i>Mean Squared Error</i>
NSE	<i>Nash–sutcliffe Model Efficiency Coefficient</i>
LSTM	<i>Long short-term memory</i>
ONS	<i>Operador Nacional do Sistema Elétrico</i>
PSO	<i>Particle Swarm Optimization</i>
RNA	<i>Rede Neural Artificial</i>
SMAP	<i>Soil Moisture Accounting Procedure</i>
UHE	<i>Usina Hidrelétrica</i>

SUMÁRIO

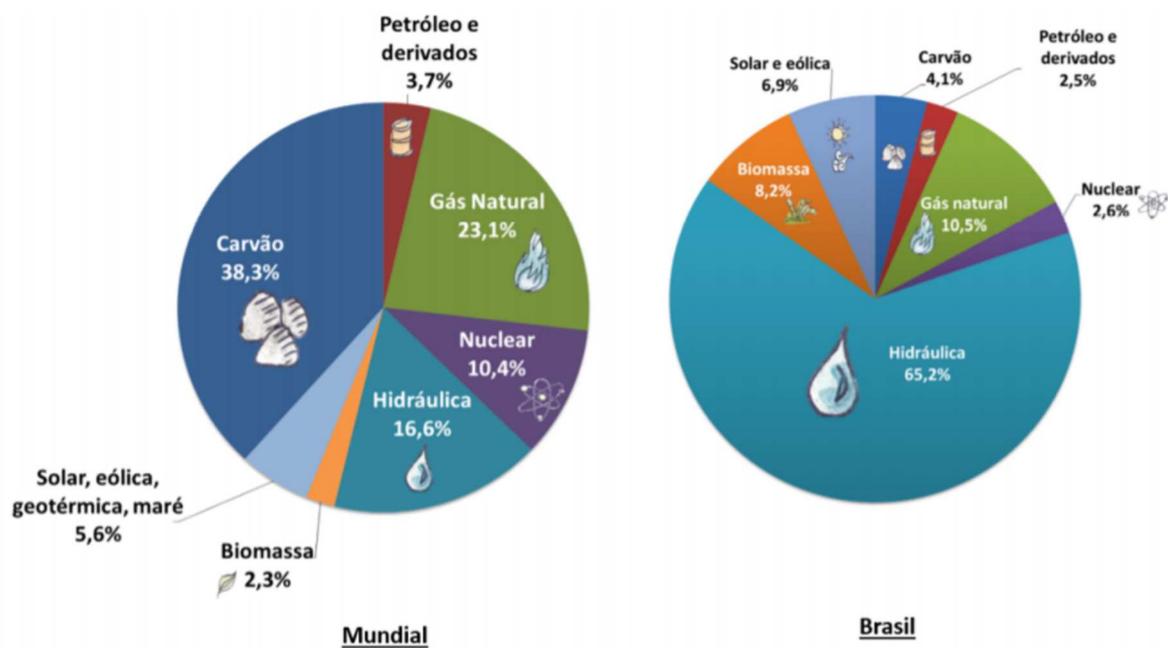
1	INTRODUÇÃO	12
1.1	MOTIVAÇÃO	12
1.2	OBJETIVOS E CONTRIBUIÇÕES	14
1.3	PUBLICAÇÃO DIRETA	15
1.4	ESTRUTURA DO DOCUMENTO	15
2	REVISÃO DA LITERATURA	16
2.1	MODELOS HIDROLÓGICOS CONCEITUAS	17
2.2	MODELOS BASEADOS EM INTELIGÊNCIA COMPUTACIONAL	22
2.3	MODELOS ACOPLADOS	24
3	METODOLOGIAS	26
3.1	MODELO SMAP/ONS	27
3.2	AQUISIÇÃO E CONDICIONAMENTO DE DADOS	34
3.3	PROPOSTA DE OTIMIZAÇÃO DO MODELO SMAP	41
3.4	MODELO HÍBRIDO SMAP-DNN	46
3.5	ARQUITETURA DE SOFTWARE	58
3.6	Resumo	63
4	RESULTADOS	64
4.1	DESCRIÇÃO DOS ESTUDOS DE CASO	64
4.2	CALIBRAÇÃO GLOBAL DO MODELO SMAP	65
4.3	OTIMIZAÇÃO LOCAL DO SMAP	67
4.4	SMAP-DNN	68
4.5	Resumo	79
5	CONCLUSÕES	80
5.1	TRABALHOS FUTUROS	81
	REFERÊNCIAS	83

1 INTRODUÇÃO

1.1 MOTIVAÇÃO

A geração de energia elétrica, através de usinas hidrelétricas (UHEs), é uma das principais formas de geração de energia aplicadas no mundo. Esta fonte é a mais utilizada em países que possuem um potencial hídrico considerável, devido ao seu baixo custo operacional quando comparada à hidrocarbonetos. O Brasil possui uma matriz elétrica composta predominantemente por hidrelétricas. Dessa forma, faz-se necessária a realização de estudos para garantir um melhor aproveitamento desta fonte e reduzir a dependência da geração de energia por termelétricas, pois estas apresentam custos operacionais mais elevados. A Figura 1 apresenta a matriz elétrica mundial e a brasileira de 2020, evidenciando a relevância da geração hidráulica no setor elétrico.

Figura 1 – Matrizes Elétricas Mundial e Brasileira



Fonte: (EPE, 2020).

No Brasil, o sistema elétrico é coordenado pelo Operador Nacional do Sistema Elétrico (ONS), cujo objetivo é a otimização e o planejamento da geração e transmissão de energia, considerando diferentes horizontes de tempo. O mesmo faz o uso do modelo de chuva-vazão SMAP (*Soil Moisture Accounting Procedure*) para o planejamento de curto e longo prazo na maior parte das bacias brasileiras. O SMAP é extremamente importante no cenário nacional, interferindo diretamente no preço da energia em todo País.

O ONS realiza estudos da disponibilidade de recursos hídricos para o planejamento operacional das UHEs. Assim, é relevante prever a quantidade de água que estará disponível e, conseqüentemente, a energia que poderá ser gerada. Um dos principais aspectos que deve

ser considerado no planejamento é analisar as afluições que aumentam os níveis de água dos reservatórios. Períodos de seca levam a um aumento da geração pelas termelétricas devido à menor disponibilidade de geração pelas hidrelétricas.

Assim, os estudos de recursos hídricos proporcionam aos órgãos competentes lidar com o “Dilema do Operador”(vide Figura 2). O termo se refere ao problema da decisão na alocação de geração hídrica/termoelétrica para suprir a demanda energética do país. Portanto, as decisões tomadas no presente devem considerar o futuro da precipitação nas áreas de interesse. Dessa forma, não haverá desperdício ou falta de recursos hídricos, diminuindo o risco de racionamento e tornando a geração de energia menos custosa.

Figura 2 – Dilema do Operador



Fonte: (MOROMISATO, 2012).

Uma boa previsão de afluição e um bom planejamento podem proporcionar: (i) maior eficiência de geração; (ii) minimizar vertimento; (iii) otimização do despacho das máquinas da planta; (iv) segurança da estrutura da barragem. Com base nos dados meteorológicos, as vazões afluentes nos rios podem ser obtidas por meio de modelos que caracterizam a bacia em que a usina está inserida. Estes modelos são conhecidos como chuva-vazão.

Nos últimos anos, a Universidade Federal de Juiz de Fora vem atuando na melhora do modelo SMAP através de Projetos de Pesquisa e Desenvolvimento (P&D). O antigo programa foi substituído por aplicativos de inteligência computacional baseado em algoritmos evolutivos bioinspirados. Houve uma grande melhora de performance, resultando em artigos e dissertações (SILVA, 2019).

Dada a importância do modelo SMAP e o progresso da UFJF no tema, o presente

trabalho buscou uma alternativa de aperfeiçoar o modelo para previsões de aflúências diárias, fazendo parte de um projeto P&D. A estratégia principal foi investigar inovações no acoplamento do SMAP com ferramentas *Deep Learning*, amplamente utilizadas em problema de previsões de séries temporais. A Seção 2 irá apresentar diversos trabalhos na literatura que contextualizam o tema, desde o uso de técnicas de aprendizado de máquina, modelos hidrológicos matemáticos clássicos e modelos híbridos.

1.2 OBJETIVOS E CONTRIBUIÇÕES

Modelos Hidrológicos são ferramentas essenciais na estimativa dinâmica das bacias hidrográficas. O Setor Elétrico Brasileiro (SEB), que possui matriz predominantemente hidráulica, tem seu planejamento diretamente dependente da disponibilidade de recursos hídricos nos grandes reservatórios. Assim, modelos chuva-vazão buscam, a partir de cenários anteriores e/ou previsões de precipitação, obter valores de aflúência.

Com a definição do problema, as contribuições vão desde métodos inovadores de aplicação dos dados meteorológicos no processo de otimização dos modelos, até a busca no aumento da assertividade de previsões através de acoplamento do SMAP com arquiteturas *Deep Learning*.

A pesquisa possui uma grande replicabilidade para demais áreas de escoamento de grandes UHE's. Desenvolvida na plataforma Python, utiliza de fontes de dados gratuitas do INPE (Instituto Nacional de Pesquisas Espaciais), ANA (Agência Nacional de Águas) e INMET (Instituto Nacional de Meteorologia), utilizados para obtenção da chuva e evapotranspiração. Os softwares são relativamente simples, não necessitando de hardware avançado. Diferente do padrão adotado pelo ONS para aplicação do SMAP, a pesquisa utiliza de dados de precipitação no formato *grid*, fornecidos pelo produto MERGE (CPTEC/INPE). Assim, foi contornado o problema de algumas regiões apresentarem poucas estações pluviométricas, sendo que há, muitas vezes, falhas de dados relevantes nas mesmas. Fora isso, os modelos de previsão meteorológica também constituem dados no formato *grid*, sendo mais fácil adaptar modelos treinados para a operação. O trabalho visa o uso do modelo de previsão BAM (*Brazilian Atmospheric Model*) que fornece até 11 dias de previsões. Espera-se que as previsões de vazão natural, neste horizonte, possibilitem tomadas importantes de decisões relacionadas ao mercado de energia e a manutenção/operação de turbinas e vertedouros.

Resumidamente as contribuições são:

- Foi descartada a tradicional aquisição de dados de precipitação observada por estações. Ao invés disso, o produto MERGE (CPTEC/INPE) fornece dados em *grid* de forma homogênea, contornando o problema dos dados falhos e esparsados da ANA para algumas estações. Fora que torna os modelos compatíveis às previsões meteorológicas

em *grid*. Um software de aquisição automática e uma metodologia de filtragem de contorno foram desenvolvidos em *python* para tratamento dos dados do INPE.

- Uma abordagem para aquisição automática dos dados de temperatura de estações do INMET e cálculo da evapotranspiração média com ponderamento espacial foi desenvolvida.
- Diferente da abordagem do ONS, foi proposta uma calibração local do SMAP que permite pequenas variações de parâmetro de bacia. A partir de parâmetros globais, permitem-se pequenos ajustes para acomodar curva de vazão das últimas 2 semanas. Fora isso, também é ajustado o valor inicial dos reservatórios e chuva de entrada. Para essa otimização foram utilizadas e propostas modificações no algoritmo de busca local *Twiddle*. Estes ajustes suaves e contínuos conduzem rodadas de calibração em um padrão semelhante, facilitando o treinamento da ferramenta de *Deep Learning* que opera acoplada ao SMAP.
- Principal contribuição do trabalho: um modelo híbrido acopla o SMAP ao *deep learning* por meio das variáveis de estado intermediárias (valor armazenado nos reservatórios). Uma arquitetura baseada em Convoluções 3D e camada recursiva *LSTM* foi proposta para processar dados e *grid* de previsões de chuva de forma espaço-temporal. Essa precipitação, com as entradas referentes ao SMAP, fornecem uma curva de saída de vazão natural prevista.

1.3 PUBLICAÇÃO DIRETA

- GM Maciel, VA Cabral, ALM Marcato, ICS Júnior, LDM Honório. Daily Water Flow Forecasting via Coupling Between SMAP and Deep Learning. *IEEE Access*, v. 8, p. 204660-204675, 2020.

1.4 ESTRUTURA DO DOCUMENTO

O trabalho foi organizado da seguinte forma: O capítulo 1 apresenta a introdução e contribuições. O capítulo 2 contém a revisão da literatura, descrevendo os trabalhos correlatos. O capítulo 3 apresenta as metodologias, exibe os principais fundamentos teóricos utilizados de forma resumida, além de todo o desenvolvimento e a proposta detalhada da tese. O capítulo 4 contém os resultados, análises e discussões. Por último, o Capítulo 5 conclui e apresenta possíveis continuações.

2 REVISÃO DA LITERATURA

Este capítulo apresenta trabalhos correlatos ao tema estudado. A pesquisa buscou artigos e teses com os temas: (i) modelos hidrológicos clássicos com ênfase no SMAP, (ii) modelos de chuva-vazão baseados em inteligência computacional e (iii) modelos chuva-vazão híbridos, que utilizam duas ou mais técnicas acopladas, com ênfase em exemplos matemáticos assistidos por alguma ferramenta de *machine learning*.

Independente da área de conhecimento, o uso de modelos busca uma representação de um sistema, de forma que possa ser operado e analisado. Segundo Nelles (2020), os modelos de sistemas podem ser divididos em três categorias:

- Caixa branca: São modelos totalmente derivados de princípios físicos, biológicos, econômicos, etc. Todas as equações e os parâmetros são determinados por análise e modelagem teórica. As vantagens desta categoria são: (i) boa extrapolação, com facilidade de especular, estimar ou chegar a uma conclusão com base em fatos e em observações conhecidas; (ii) alta confiabilidade, entendimento e escalabilidade. As principais desvantagens são: (i) a necessidade de uma maior perícia sobre o processo; (ii) maior demanda de tempo para a modelagem; (iii) limitação superior, pelo fato do modelo estar preso a uma matemática que se aproxima do processo real.
- Caixa preta: os parâmetros do modelo não têm relação direta com os princípios fundamentais. São baseados exclusivamente em dados de medição. Para a construção de modelos caixa preta, nenhum ou muito pouco conhecimento prévio é explorado. As principais vantagens são: (i) a não necessidade de conhecimento prévio do processo básico que rege o modelo real; (ii) menor demanda temporal para sua análise e confecção. As desvantagens são: (i) o baixo entendimento intrínseco da modelagem; (ii) baixa escalabilidade e extrapolação; (iii) restrição à acurácia dos dados observados.
- Caixa cinza: É a combinação das características dos modelos caixa branca e caixa preta, almejando extrair vantagens para uma solução que supere o desempenho da aplicação isolada destas vertentes. Busca acoplar um equacionamento baseado em princípios teóricos fundamentais do processo e, em simultâneo, abstrair certas partes utilizando dados observacionais. Assim, internamente, todas as peculiaridades do sistema não são totalmente conhecidas como os modelos caixa branca.

Alguns autores preferem não abordar a classificação dos modelos em apenas três categorias. Karplus (1977) propõe avaliar os sistemas através de um espectro cinzento, que vai desde um modelo caixa branca baseado em apenas estudos e equações teóricas sobre o problema, até um modelo puramente caixa preta, onde há apenas um processamento

puro de dados observados. A medida que o modelo aumenta sua abstração em relação ao mundo real, utilizando maior informação auxiliar e dados observados, ele se aproxima do extremo preto. A Figura 3 exemplifica essa abordagem.

Figura 3 – Ilustração do espectro cinzento em modelos.



Fonte: elaborada pelo autor.

No contexto hidrológico, há diversas abordagens de modelos chuva-vazão em diferentes regiões do espectro cinzento, desde modelos hidrológicos puramente teóricos baseado em estudos geológicos, até modelos regressivos baseados puramente em séries de vazões observadas. Nas últimas décadas, diversos estudos foram desenvolvidos visando prever o comportamento de sistemas hidrológicos. Em estudos mais atuais, as técnicas de aprendizado de máquina são as mais utilizadas, principalmente em trabalhos que têm o objetivo de prever a vazão de rios (FLEMING; GOODBODY, 2019; YASEEN *et al.*, 2019).

Na literatura há inúmeras outras abordagens para o problema, utilizando modelos hidrológicos clássicos, técnicas de aprendizado de máquina e acoplamentos entre os mesmos. As próximas seções apresentam trabalhos correlatos, divididos pela sua caracterização do espectro cinzento. Modelos clássicos tendem a apresentar características caixa branca, enquanto modelos de inteligência computacional, como redes neurais artificiais e regressões, tendem a se comportar como caixas pretas. Já os modelos híbridos, podem apresentar características destes dois extremos, se enquadrando em modelos caixa cinza ou centralizados no espectro cinzento.

2.1 MODELOS HIDROLÓGICOS CONCEITUAS

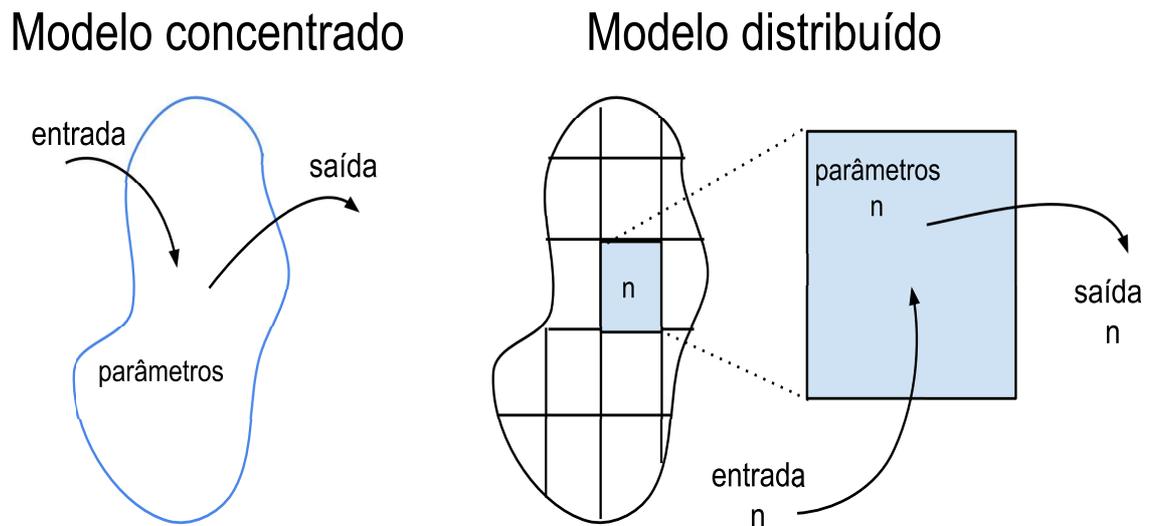
Modelos hidrológicos buscam representar o ciclo da água em bacias hidrográficas. Os mesmos são chamados conceituais quando há um conjunto de equações explícitas que descrevem todo o processo físico (ALMEIDA; SERRA, 2017). Conforme apresentado nas Figuras 4 e 5, os mesmos podem ser classificados:

- Concentrados: toda a área da bacia é representada de maneira única. As variáveis de entrada e saída são valores unitários representativos. São considerados simplificados

e de maior facilidade de calibração.

- Distribuídos: A área da bacia é subdividida em seções, sendo que cada uma recebe valores independentes de entrada sendo parametrizadas independentemente. Assim há uma distribuição espacial das variáveis e dos parâmetros. Posteriormente as seções são acopladas para reconhecer a vazão em um ponto de interesse. São complexos, exigindo alto esforço computacional.
- Determinísticos: Modelos que, estando em um mesmo estado, produzem a mesma resposta para um mesmo conjunto de entradas.
- Estocásticos: O processo modela uma distribuição de probabilidade através de aleatoriedades no conjunto de variáveis e entradas. Portanto, a saída pode ser um conjunto de possíveis vazões, ou uma região probabilística onde as mesmas se encontram.

Figura 4 – Modelo hidrológico concentrado vs distribuído.

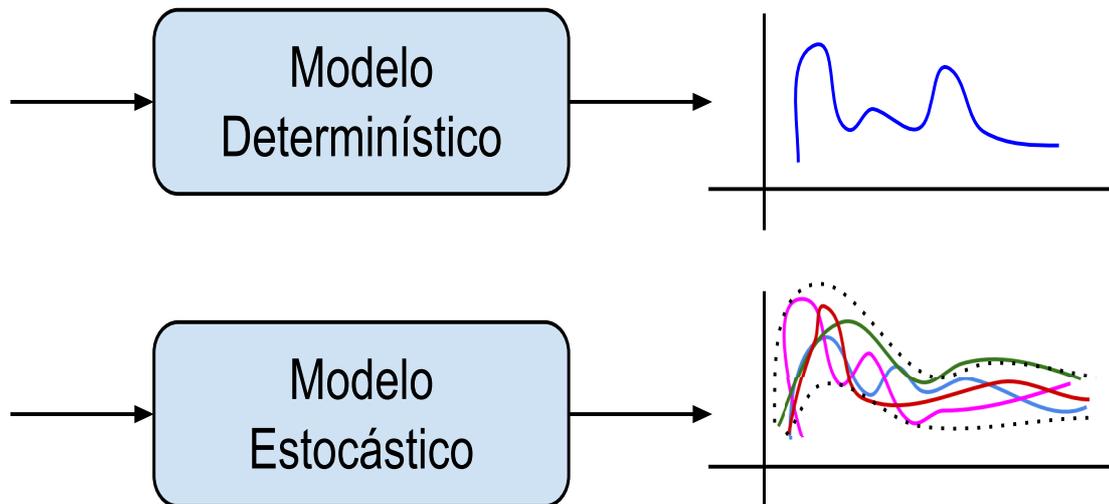


Fonte: elaborada pelo autor.

Os modelos conceituais são considerados paramétricos, já que as áreas são representadas de acordo com parâmetros geológicos, que interferem na dinâmica do processo chuva-vazão. Diferentes categorias de vegetação, relevo e solos apresentam comportamentos distintos ao receberem uma precipitação. Nas últimas décadas, diversos trabalhos utilizaram modelos conceituais em inúmeras bacias, focando em sua aplicação e na otimização de seus parâmetros.

Em (CHANG *et al.*, 2019), a disponibilidade de água foi prevista pelo modelo hidrológico *Variable Infiltration Capacity (VIC)*, o mesmo foi aplicado para estimar a

Figura 5 – Modelo hidrológico determinístico vs estocástico.



Fonte: elaborada pelo autor.

altura do Rio Mekong na Ásia. Em Bhuiyan *et al.* (2017), em um modelo conceitual foi aplicada a previsão de vazão. No entanto, a abordagem dos autores enfatiza regiões subárticas, considerando o derretimento da neve da temporada de inverno, e aplica uma análise de sensibilidade da inicialização dos parâmetros de umidade do solo. Os autores utilizaram o modelo *Hydrologic Modeling System* (HEC-HMS), que simula o processo chuva-vazão em bacias.

Em Wöhling e Burbery (2020), foi proposto um modelo matemático baseado em reservatórios lineares para traduzir o comportamento dos fluxos em uma bacia em função da precipitação e evapotranspiração. Os parâmetros do modelo foram estimados com base em funções de densidade de probabilidade usando um método de Monte Carlo de cadeia de Markov.

Em Wijayarathne e Coulibaly (2020) os seguintes modelos são comparados para a previsão de vazão: *Sacramento Soil Moisture Accounting* (SAC-SMA), *modèle du Génie Rural à 4 paramètres Journalier* (GR4J), *McMaster University Hydrologiska Byråns Vattenbalansavdelning* (MAC-HBV), *Hydrologic Engineering Center Hydrologic Modeling System* (HEC-HMS) e o *University of Waterloo Flood Forecasting System* (WATFLOOD). O método *Dynamically Dimensioned Search* (DDS) foi usado para a calibração de todos os métodos.

Os autores de Meng *et al.* (2017), Samuel *et al.* (2019) aplicam um modelo de chuva-vazão para determinar o estado do solo para a previsão de vazão a curto prazo com estratégias que aplicam o *Ensemble Kalman filter* (*EnKF*) para assimilar e rastrear os dados do solo de forma dinâmica. A aplicação deste método pode ser usada em outros

problemas de previsão dos aspectos hidrológicos, tais como qualidade da água (LOOS *et al.*, 2020). Os autores de Patil e Ramsankaran (2017) aplicam o *EnKF* para assimilar dados para a previsão do fluxo do rio usando o programa *Soil and Water Assessment Tool (SWAT)*, um completo simulador de bacias com modelagem distribuída.

Em Rezaie-Balf *et al.* (2019), o coeficiente de *Mallow* foi usado com os modelos de dados de pré-processamento *M5Tree* e *MARS* para previsão de vazão de rios. Em Ghorbani *et al.* (2018), uma técnica baseada na teoria do caos foi aplicada em conjunto com uma variação de algoritmo genético, com o mesmo objetivo.

O modelo SMAP é uma das ferramentas utilizadas pelo ONS para criar cenários de fluxos para as bacias brasileiras. Este modelo determinístico e centralizado influencia diretamente no planejamento da operação do sistema elétrico e na dinâmica do mercado energético brasileiro. Na última década surgiram pesquisas visando sua aplicação em bacias brasileiras, aspirando melhorias no processo de otimização de seus parâmetros e entradas. Atualmente, o ONS trabalha com otimização do SMAP utilizando o algoritmo evolutivo *BAT*, pré-processando a precipitação de entrada provinda de estações.

Em algumas regiões do Brasil, existe baixa qualidade de registros pluviométricos, com poucas estações e falhas de dados. Uma solução encontrada por Silva *et al.* (2019) combinou dados de chuva observada de satélite, através do produto MERGE-GPM (CP-TEC/INPE). O trabalho aplicou os dados de chuva combinada como entrada no SMAP nas UHEs de Camargos, Funil, Furnas, Porto Colômbia, Euclides da Cunha, Marimondo e Água Vermelha. O desempenho do modelo foi aprimorado em quatro casos, sendo que o restante manteve o mesmo comportamento.

Em estudos brasileiros, o SMAP é também aplicado com objetivos fora do setor elétrico. Algumas regiões sofrem grandes problemas com as cheias dos rios, colocando em risco a vida e os bens materiais da população local. Assim, o trabalho Bou *et al.* (2015) utilizou a versão mensal do SMAP em uma sub-bacia do Rio Uruguai, na fronteira entre os estados de Santa Catarina e Rio Grande do Sul. Alimentando o modelo com dados de previsão meteorológica do Climate Forecast System V2 (CFSv2) o SMAP identificou a tendência de enchentes na maioria dos casos. O trabalho utilizou dados de estações fluviométricas no período de 1986 a 2005 para calibrar e validar o SMAP, e de 2005 a 2010 para realizar as simulações com o modelo de previsão.

Em Silva *et al.* (2019), os autores aplicaram o modelo de previsão mensal de vazões de água para a UHE Água Vermelha. O modelo é aplicado em Paiva *et al.* (2020), para realizar previsões de vazões mensais na bacia de Três Marias (Brasil). O modelo SMAP pode ser usado em previsões de curto prazo em sua versão diária como, por exemplo, alimentado com o modelo de previsão *Weather Research and Forecasting (WRF)* para prever enchentes no Rio de Janeiro (BARCELLOS; CATALDI, 2020).

Devido a sua característica não-linear e multimodal, pesquisas recentes vêm aplicando diferentes meta-heurísticas para calibração do modelo SMAP. Além do BAT, utilizado oficialmente pelo ONS, diversos outros algoritmos evolutivos são objetos destes estudos, variando o método e a função objetivo (FOB). No trabalho de Saraiva (2010) o modelo SMAP foi aplicado para a região do Paraíba e a calibração dos parâmetros foi realizada por meio dos algoritmos *Dynamically Dimensioned Search* (DDS) e *Shuffled Complex Evolution* (SCE). Em Marcon (2019), foi realizado um estudo do modelo SMAP em uma sub-bacia do rio Carreiro, Rio Grande do Sul. Utilizando o solver do Excel, calibraram-se versões diária e mensal do SMAP, utilizando 3 anos de dados históricos. Na validação foi utilizado 10 anos de histórico e diversas métricas de avaliação. Os modelos, em suas duas versões, apresentaram resultados satisfatórios para o coeficiente *RNSlog*, tanto na calibração quanto na validação.

Em Caponi *et al.* (2011) o Algoritmo Genético - AG - foi aplicado para calibrar os parâmetros do SMAP da bacia do rio São Francisco em sua versão mensal. O AG foi identificado como uma ferramenta robusta com capacidade de obter calibrações em alto grau de aderência entre as curvas de vazão observadas e calculadas. Em Rampinelli e Costa (2015), o modelo SMAP foi aplicado em uma região do semiárido do Nordeste brasileiro. A calibração do modelo foi realizada usando AG, *Particle Swarm Optimization* - PSO e *Nelder Mead algorithm* (NM). As soluções encontradas pelos algoritmos evolutivos eram pós-processadas pelo NM.

Outra meta-heurística utilizada em pesquisas nacionais é o *Multi-Objective Particle Swarm Optimization* (MOPSO) (PEREIRA, 2017; NASCIMENTO *et al.*, 2009). Esta versão do PSO foi modificada para melhor desempenho em problemas multi-objetivo, quando há mais de um critério de avaliação da solução a ser considerado simultaneamente. Pereira (2017) aplicou o MOPSO em quatro sub-bacias do estado do Ceará, com duas diferentes FOBs baseadas *Nash-Sutcliffe model efficiency coefficient* (NSE). Assim o MOPSO calibrou seis parâmetros do SMAP ponderando as FOBS no problema de multi-objetivo. De maneira similar, Nascimento *et al.* (2009) realizou um estudo com quatro FOBs baseados em NSE. O MOPSO calibrou dois parâmetros do SMAP para 21 áreas do Ceará, sendo cada uma referente à área de drenagem de uma estação fluviométrica.

Campos (2021) compara o uso de três meta-heurísticas para calibração operacional do SMAP-ONS: as técnicas bioinspiradas analisadas foram o *Grey Wolf Optimizer* (GWO); o *Bat Algorithm* (BA); e uma proposta acoplada dos dois *Hybrid Grey Wolf-Bat Algorithm* (GWOBA). Foram realizadas 1000 simulações para cada técnica em cada região de estudo, sendo elas as sub-bacias das usinas de Carumba, Emborcação e Nova Ponte.

Como pode ser visualizado, diversos estudos brasileiros contemplam a problemática da otimização do modelo SMAP. A Tabela 1 resume os trabalhos sobre o tema, abordados nessa revisão bibliográfica.

Tabela 1 – Trabalhos de otimização do SMAP

Trabalho	Região	Otimização
Nascimento, 2009	21 estações em rios do Ceará	MOPSO
Schardong, 2009	rio Piriqui	GA PSO DE
Saraiva, 2010	rio Paraopeba	DDS SCE
Capponi, 2011	3 sub-bacias rio São Francisco	GA
Rampineli, 2015	rio Ipanema	GA PSO NM
Pereira, 2017	4 bacias do Ceará	MOPSO
Marco, 2019	rio Carreiro	Solver Excel
Leone, 2021	usinas de Carumba, Emborcação e Nova Ponte	BA GWO GWOBA

2.2 MODELOS BASEADOS EM INTELIGÊNCIA COMPUTACIONAL

Os modelos baseados em inteligência computacional buscam modelar o comportamento chuva-vazão abstraindo o processo físico. Não há interesse de parametrizar a bacia, apenas encontrar soluções computacionais para correlacionar as entradas com a saída desejada, como às séries de precipitações e vazões. Assim, são utilizados, por exemplo: técnicas de aprendizado de máquina e regressões. A vantagem desta abordagem está na não necessidade de estudos/conhecimentos geológicos da bacia, apenas um banco de dados observados.

Em (ZHU *et al.*, 2020) uma *Feed Forward Neural Network* (FFNN) foi aplicada e comparada com uma arquitetura *Long Short-Term Memory* (LSTM), que também foi utilizada nesta tese. Seu objetivo era previsão de níveis de água, aplicado para lagos poloneses, que está diretamente relacionada com a vazão afluente. No trabalho (ZHOU *et al.*, 2020a), dos mesmos autores, uma rede neural *Multilayer Perceptron* (MLP) foi aplicada para o mesmo fim, mas para um rio navegável. Em (SHANG *et al.*, 2019), foram utilizadas redes neurais para previsão de nível em uma UHE chinesa. O modelo permitiu

a previsão direta, precisa e em tempo real das mudanças nos níveis de água a jusante, usando dados de operação da hidrelétrica. Fora isso, este estudo também contribui para garantir a segurança das estruturas da barragem.

Além do nível da água, estudos para prever a precipitação também são muito comuns. Em Belayneh *et al.* (2014) redes neurais artificiais (RNAs) e *Support Vector Machine* (SVM) foram aplicados, com a transformação wavelet para o pré-processamento dos dados de entrada. Estes modelos buscaram prever o comportamento das chuvas até 6 meses no futuro. Rasouli *et al.* (2012) também aplicou técnicas de Aprendizado de Máquina para prever o fluxo do rio com até 7 dias de antecedência. As técnicas utilizadas foram: *Support Vector Regression* (SVR), *Bayesian Neural Network* (BNN), *Gaussian Process* (GP) e *Multiple Linear Regression* (MLR), todas comparadas com relação à eficiência de previsão.

Yaseen *et al.* (2019) apresenta uma extensa revisão da literatura e demonstra a aplicabilidade das técnicas de DL para prever o fluxo de água em rios e outros aspectos hidrológicos. Além disso, eles aplicam o SVR para comparar com um modelo de *Extreme Learning Machine* (ELM) na previsão do fluxo de água em um rio da Malásia.

O fluxo de água disponível nos rios é altamente dependente dos aspectos climáticos e deve ser considerado durante o processo de previsão. Rokaya *et al.* (2020) mostram que o congelamento e descongelamento dos rios impactam as afluições observadas nas regiões frias. Para a previsão considerando esses aspectos, eles aplicaram redes neurais artificiais (RNA) e mostraram que essa ferramenta também é amplamente utilizada quando se considera os efeitos sazonais do congelamento de rios.

A aplicação de uma *Deep Belief Network* (DBN) com *Variational Mode Decomposition* (VMD) foi a estratégia aplicada em (XIE *et al.*, 2019) para prever o fluxo da bacia do rio Han, na China. Diferente do *Deep Learning* tradicional, estas redes são pré-treinadas camada por camada, constituindo uma boa solução inicial de fuga de mínimos locais. Para a calibração do DBN, foi utilizada a otimização do enxame de partículas aprimorada (IPSO) e a previsão foi feita com 1, 3, 5 e 7 dias de antecedência. A mesma região foi objeto de estudo em (ZUO *et al.*, 2020), onde o LSTM foi aplicado com o VMD.

Belvederesi *et al.* (2020) aplica o *adaptive neuro-fuzzy inference system* (ANFIS) para prever o fluxo em um rio canadense, que mistura o alto poder de reconhecimento de padrão das redes neurais com a robustez dos sistemas *Fuzzy*. Em Chen *et al.* (2019), os autores aplicam uma Regressão de *Convolution Regression based on Machine Learning* (CRML) para prever o fluxo de água de um rio localizado em uma bacia hidrológica chinesa.

2.3 MODELOS ACOPLADOS

O acoplamento de técnicas tem sido utilizado em diversos estudos visando desenvolver modelos de previsão de variáveis hidrológicas. A ideia desse acoplamento é obter um melhor desempenho de previsão, em comparação, ao aplicar as técnicas separadamente. Acoplamentos híbridos vêm sendo estudados há décadas em várias áreas da Engenharia, destacando-se em modelos de sistemas de controle. A aplicação em sistemas hidrológicos é relativamente recente, com poucos trabalhos na literatura.

Em Xu *et al.* (2018), a previsibilidade de secas em diferentes regiões da China foi investigada através de uma série de modelos estatísticos, dinâmicos e híbridos. Quando modelos estatísticos e dinâmicos foram acoplados, alcançaram-se os melhores resultados. Já Zhou *et al.* (2020b) desenvolveu uma técnica de previsão de 4 dias da afluência do reservatório da UHE Três Gargantas, China. Neste estudo, o *Unscented Kalman Filter* (UKF) foi aplicado combinado com duas técnicas de DL: BPNN e uma rede neural recorrente. Ao pós-processar a saída dos modelos DL com o UKF, houve uma melhoria no erro e inseriu uma análise estocástica à saída.

As técnicas de aprendizado de máquina podem ser utilizadas de maneira a melhorar resultados de modelos hidrológicos, realizando um pós-processamento. Pode-se, por exemplo, aplicar algoritmos de regressão de quantis em aprendizado de máquina para o pós-processamento probabilístico da vazão calculada pelo modelo hidrológico GR4J (PAPACHARALAMPOUS *et al.*, 2019).

Em Yin *et al.* (2016) o modelo hidrológico *Soil Moisture Accounting and Routine* (SMAR) foi comparado, em relação à precisão da previsão de vazão do rio Heihe na China, às seguintes estratégias de Deep Learning: *back-propagation neural network* (BPNN), *general regression neural network* (GRNN) e *rotated GRNN model* (RGRNN). O trabalho provou que ao inserir a saída do modelo SMAR como entrada das técnicas de Deep Learning, obtiveram-se melhores resultados.

Kwon *et al.* (2020) acoplaram um modelo hidrológico conceitual em série com uma ferramenta de aprendizado de máquina. A pesquisa propõe um modelo híbrido através de um pós-processamento das variáveis de estado intermediárias do modelo hidrológico, semelhante ao SMAP, denominado Tank Model, através do SVM. A estratégia abordada vai ao encontro da proposta desta tese, mas com métodos distintos.

Song (2020) utilizou redes neurais convolucionais 2D para prever vazões na Coreia do Sul. Para isso, ele criou entradas baseadas em imagens hidrológicas da região. Estas imagens acoplam diversas informações geológicas da região, como evapotranspiração e precipitação em cada ponto do mapa. Além disso, um dos canais dessa imagem é resultado do modelo SCS *Curve Number Method* aplicado em cada *grid*, uma técnica analítica que calcula vazões em uma determinada área baseada em curvas que relacionam a precipitação

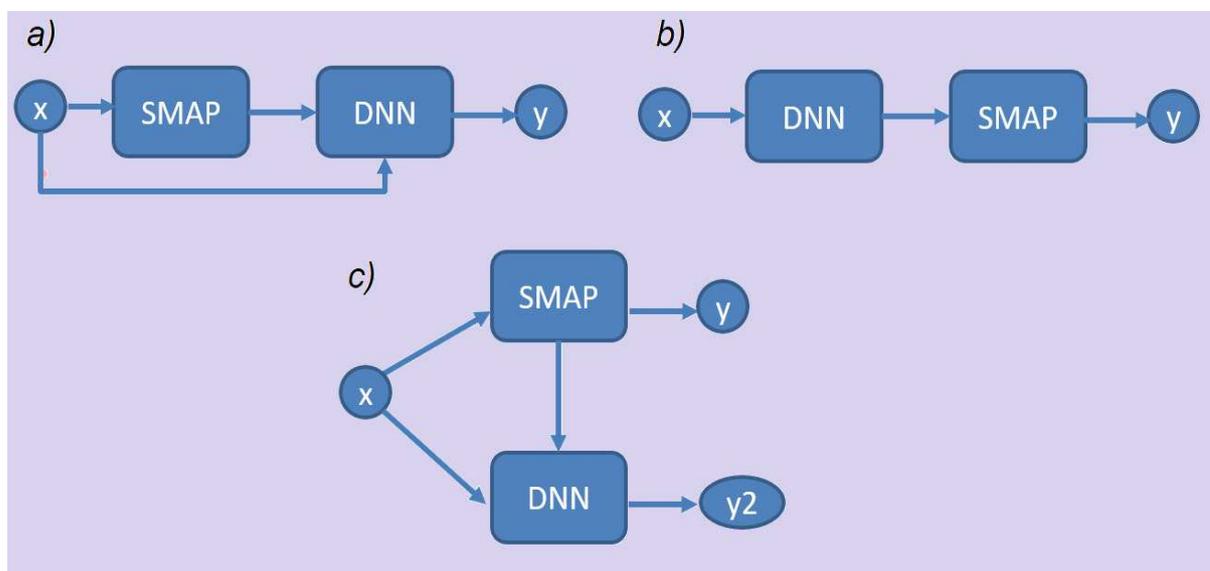
com o volume de água acumulado. Assim, com este acoplamento híbrido série, a rede fornece como saída a vazão futura em diversos pontos da imagem que contém estações fluviométricas.

3 METODOLOGIAS

A metodologia explora o acoplamento de redes neurais profundas como modelo SMAP para previsões de vazões naturais afluentes em reservatórios. Esta fusão obtém resultados melhores aos encontrados às técnicas isoladas.

Conforme apresentado nos trabalhos correlatos (Capítulo 2), existem diversas maneiras de acoplamento entre modelos hidrológicos com ferramentas de *machine learning*. Na fase preliminar deste trabalho, algumas possibilidades de fusão entre o modelo SMAP e uma rede neural foram avaliadas. A Figura 6, apresenta três possibilidades:

Figura 6 – Possibilidades de acoplamento SMAP-DNN



Fonte: elaborada pelo autor.

- **a) Pós-Processamento Série:** Neste tipo de acoplamento, o modelo SMAP é operado em primeira instância. A saída de vazão do SMAP é pós-processada por uma Rede Neural Profunda, buscando uma segunda vazão mais acurada. Para isso, a DNN receberia, por exemplo, as mesmas entradas do SMAP adicionada de sua saída (vazão pré calculada).
- **b) Pré-Processamento Série:** Esta arquitetura é semelhante à otimização realizada pelo ONS, que será apresentado na Seção 3.1.1. As entradas do modelo SMAP são pré-processadas por uma inteligência computacional, buscando obter entradas mais acuradas ao modelo. Por exemplo, séries de precipitação observadas em diversos pontos da bacia podem gerar um valor otimizado para alimentar a média de precipitação de entrada do modelo.
- **c) Processamento Paralelo:** O modelo hidrológico opera de independente. Ao invés de fornecer sua saída para a rede neural, o SMAP complementa as entradas

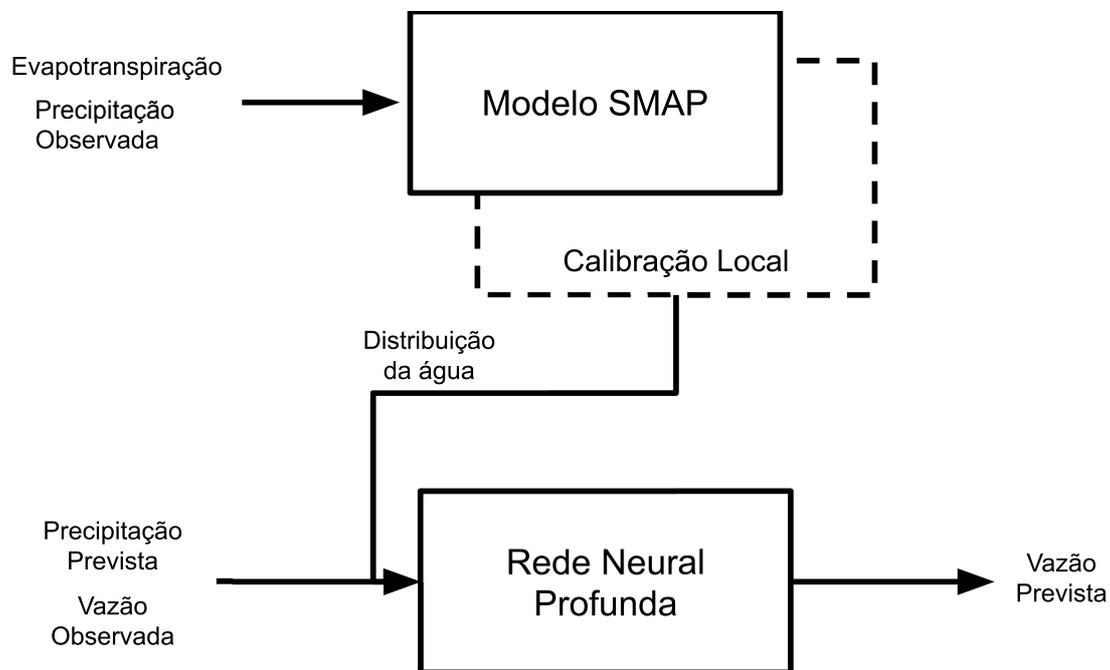
de rede neural com informações intrínsecas do modelo, através de suas variáveis de estado. Assim, a saída do original do SMAP não é relevante.

Após uma avaliação preliminar, a terceira opção mostrou-se a mais promissora, escolhida como objeto de pesquisa deste trabalho. O modelo hidrológico, executado em paralelo, fornece a atual distribuição da água do solo (variáveis intermediárias do modelo) após um ajuste de correção local. A RNA recebe esta informação com a vazão observada e com previsões de chuva.

Assim, há uma diminuição do espaço de busca e maior discriminação as entradas, tornando o processo de treinamento da RNA mais robusto. Alimentar uma máquina de aprendizado somente com as vazões anteriores está mais suscetível a mínimos locais e redundâncias. Por exemplo, não se sabe a parcela d'água proveniente dos lençóis freáticos e do escoamento direto. Fora isso, não se sabe o quanto o solo está saturado e conseguirá absorver das próximas precipitações.

Em suma, o acoplamento SMAP-DNN é descrito no diagrama da Figura 9, e os detalhes serão explicados ao longo deste capítulo. Antes de abordar as contribuições do trabalho, a seção 3.1 apresentará o modelo SMAP proposto pelo ONS.

Figura 7 – Fluxograma da metodologia híbrida proposta.



Fonte: elaborada pelo autor.

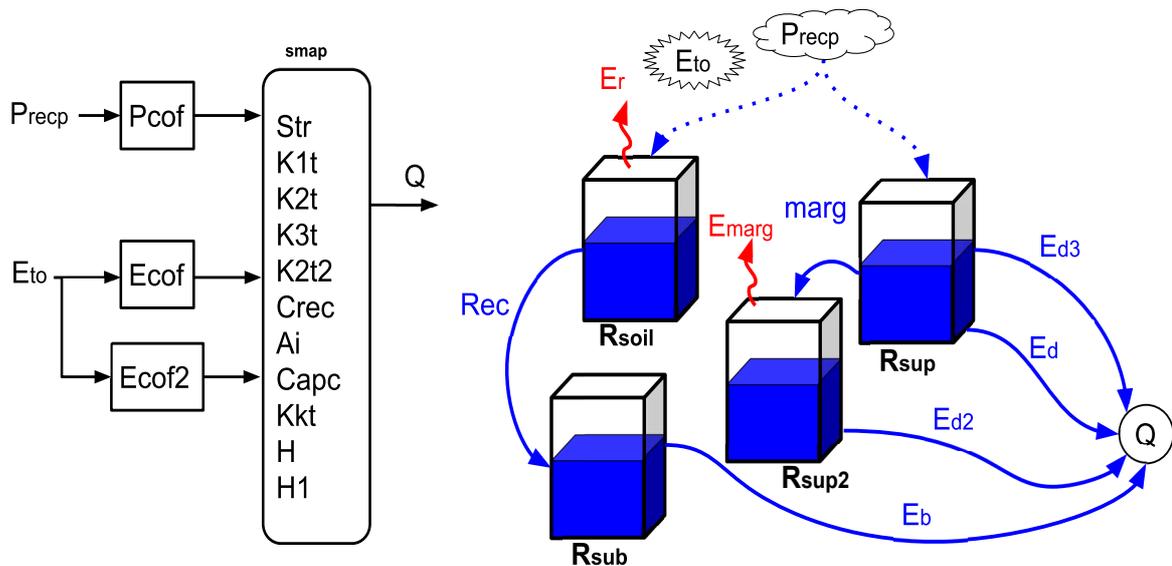
3.1 MODELO SMAP/ONS

O modelo SMAP é um modelo chuva-vazão proposto originalmente por Lopes *et al.* (1982) e que calcula a vazão em uma bacia hidrográfica em função da precipitação e

evapotranspiração ocorrida na região. Trata-se de um modelo centralizado onde, apesar da grande extensão territorial das bacias, trabalha com valores singulares de entradas. Em sua formulação original, o modelo diário é composto por três reservatórios, modelando a distribuição de volume d'água em diferentes posições do solo. Na última década, o Operador Nacional do Sistema - ONS - aprimorou o modelo, inserindo um quarto tanque referente à água acumulada em planícies e reformulando algumas equações. Este novo modelo é utilizado amplamente em bacias brasileiras, contribuindo diretamente com o planejamento da operação do sistema elétrico brasileiro.

A Figura 8 descreve o comportamento do modelo SMAP-ONS (ONS, 2017), bem como os parâmetros e as principais variáveis de entrada e de saída. Como pode ser observado, a vazão calculada é dependente dos escoamentos de cada um dos quatro reservatórios do modelo. Uma das diferenças do modelo desenvolvido e do aplicado atualmente pelo ONS está relacionada à obtenção dos dados de precipitação e evapotranspiração. Dessa forma, a obtenção desses dados será descrita antes de ser apresentada a proposta da tese.

Figura 8 – Diagrama do modelo SMAP/ONS.



Fonte: elaborada pelo autor.

Conforme já descrito, o modelo SMAP requer um conjunto de dados de precipitação e evapotranspiração. A evapotranspiração é a perda de água para a atmosfera, impactando o nível dos reservatórios R_{soil} e R_{sup2} através das evapotranspirações potencial E_r e de planície E_{marg} , obtidos através das Equações 3.1 e 3.2, respectivamente. Os coeficientes E_{cof} e E_{cof2} são parâmetros calibráveis do modelo.

$$E_r = E_{to} \cdot E_{cof} \quad (3.1)$$

$$Emarg = Eto \cdot Ecof2 \quad (3.2)$$

As condições iniciais dos reservatórios devem ser definidas para ser possível determinar a dinâmica do sistema. Essas condições são definidas com base nos valores do escoamento básico $Eb(0)$, do teor de umidade $Tu(0)$ e do escoamento superficial $Sup(0)$, utilizados para serem obtidos os estados iniciais dos reservatórios $Rsolo$, $Rsub$ e $Rsup$, respectivamente, como mostram as seguintes equações. O reservatório $Rsup2$ é inicializado como 0 como característica padrão do modelo.

$$Rsoil(0) = \frac{Tu(0) \cdot 100}{Str} \quad (3.3)$$

$$Rsub(0) = \frac{Eb(0)}{(1 - kk) \cdot AD} \cdot 86.4 \quad (3.4)$$

$$Rsup(0) = \frac{Sup(0)}{(1 - k2) \cdot AD} \cdot 86.4 \quad (3.5)$$

$$Rsup_2(0) = 0 \quad (3.6)$$

Nessas equações, Str é um parâmetro calibrável referente a capacidade de saturação do solo, e AD é a área de drenagem da região. kk e $k2$ são constantes temporais calculadas a partir das Equações 3.7 e 3.8, respectivamente, onde os parâmetros Kkt e $K2t$ também são calibráveis.

$$kk = \left(\frac{1}{2}\right)^{1/Kkt} \quad (3.7)$$

$$k2 = \left(\frac{1}{2}\right)^{1/K2t} \quad (3.8)$$

Uma vez definidos todos os parâmetros descritos, inicia-se o processo iterativo de definição de vazões e estados do reservatório a cada dia considerado. O primeiro parâmetro a ser definido é o teor de umidade $Tu(t)$ em função do estado do reservatório $Rsoil$, como mostra a Equação 3.9.

$$Tu(t) = \frac{Rsoil(t)}{Str} \quad (3.9)$$

Em seguida, é obtido o escoamento superficial no dia t ($Es(t)$) que será zero, caso a precipitação observada no dia t seja inferior ou igual à abstração do solo Ai . Caso contrário, $Es(t)$ é obtido através da Equação 3.10.

$$Es(t) = \frac{[Precp(t) - Ai]^2}{Precp(t) - Ai + Str - Rsoil(t)} \quad (3.10)$$

A evapotranspiração real $Er(t)$ é igual à potencial $Er(t)$, caso a diferença $Precp(t) - Es(t)$ seja superior à evapotranspiração potencial $Er(t)$. Caso contrário, $Er(t)$ é calculada conforme mostra a Equação 3.11.

$$Er(t) = Precp(t) - Es(t) + [Er(t) - Precp(t) + Es(t)] \cdot Tu(t) \quad (3.11)$$

A recarga subterrânea $Rec(t)$ é um escoamento que modela a água sendo absorvida pelo solo deslocando-se para regiões subterrâneas, representadas pelo reservatório subterrâneo $Rsub$. $Rec(t)$ é zero, caso o estado do reservatório do solo $Rsoil(t)$ seja inferior ao produto entre as constantes do modelo capacidade de campo $Capc$ e saturação do solo Str . Caso seja superior ao produto, $Rec(t)$ é obtido como mostra a Equação 3.12, em que $Crec$ é uma constante de recarga que também deve ser calibrada.

$$Rec(t) = Crec \cdot Tu(t) \cdot (Rsoil(t) - Capc \cdot Str) \quad (3.12)$$

O excesso de água que fica armazenado nas margens é representado pelo reservatório $Rsup2$ sendo caracterizado como um extravasamento do reservatório $Rsup$. Para o cálculo desse extravasamento $Marg(t)$, deve-se verificar se $Rsup(t)$ é inferior à altura $H1$ de $Rsup$. Em caso positivo, não há extravasamento e $Marg(t)$ é definido em zero. Caso contrário, é calculado como mostra a Equação 3.13, em que $k1$ é uma constante dependente de $K1t$, como mostra a Equação 3.14, e $K1t$ e $H1$ são parâmetros calibráveis.

$$Marg(t) = (Rsub(t) - H1) \cdot (1 - k1) \quad (3.13)$$

$$k1 = \left(\frac{1}{2}\right)^{1/K1t} \quad (3.14)$$

Uma vez definidos os estados dos reservatórios e os parâmetros característicos da bacia no instante t , são calculados os escoamentos provenientes de cada reservatório que compõem a vazão calculada pelo modelo. O escoamento $Ed(t)$ está relacionado ao reservatório $Rsup$ e obtido através da Equação 3.15, em que $k2$ é uma constante dependente de $K2t$, como mostra a Equação 3.16 e $K2t$ e H são parâmetros calibráveis.

$$Ed(t) = \min [Rsup(t) - Marg(t), H] \cdot (1 - k2) \quad (3.15)$$

$$k2 = \left(\frac{1}{2}\right)^{1/K2t} \quad (3.16)$$

O escoamento $Ed3(t)$ representa um escoamento existente quando o reservatório de superfície $Rsup(t)$ é superior a H , representando um armazenamento alto, característico de grandes vazões. Assim, caso $Rsup(t)$ é inferior a H , não haverá escoamento e $Ed3(t)$ é definido como zero. Caso contrário, $Ed3(t)$ é obtido a partir da Equação 3.17, em que $k2t$ é dependente do parâmetro calibrável $K2t2$, como mostra a Equação 3.18.

$$Ed3(t) = [Rsup(t) - H - Marg(t)] \cdot (1 - k2t2) \quad (3.17)$$

$$k2t2 = \left(\frac{1}{2}\right)^{1/K2t2} \quad (3.18)$$

Os escoamentos $Ed2(t)$ e $Eb(t)$ dos reservatórios $Rsup2$ e $Rsub$, respectivamente, podem ser obtidos através das Equações 3.19 e 3.20, em que $k3$ depende do parâmetro calibrável $K3t$, conforme mostra a Equação 3.21.

$$Ed2(t) = Rsup2(t) \cdot (1 - k3) \quad (3.19)$$

$$Eb(t) = Rsub(t) \cdot (1 - k3) \quad (3.20)$$

$$k3 = \left(\frac{1}{2}\right)^{1/K3t} \quad (3.21)$$

Por fim, a vazão natural $Q(t)$ é calculada no instante t pelo modelo através da Equação 3.22, onde o termo 86,4 é utilizado como um fator de conversão.

$$Q(t) = \frac{Ed(t) + Ed2(t) + Ed3(t) + Eb(t)}{86,4} \cdot AD \quad (3.22)$$

Uma vez calculada a vazão, a última etapa é a obtenção dos estados dos reservatórios para a próxima iteração, conforme mostram as Equações 3.23 a 3.26. Estas formulações são as equações de estado do SMAP.

$$Rsoil(t + 1) = \min [Rsoil(t) + Precp(t) - Es(t) - Er(t) - Rec(t), Str] \quad (3.23)$$

$$Rsub(t + 1) = Rsub(t) + Rec(t) - Eb(t) \quad (3.24)$$

$$Rsup(t + 1) = Rsup(t) + Es(t) - Marg(t) - Ed(t) - Ed3(t) + \max(0, Rsolo(t) + Precp(t) - Es(t) - Er(t) - Rec(t) - Str) \quad (3.25)$$

$$R_{sup2}(t + 1) = \max(R_{sup2}(t) + Marg(t) - Ed2(t) - Emarg(t), 0) \quad (3.26)$$

O processo de calibração do modelo SMAP consiste em calibrar 14 parâmetros referentes às características da bacia. Além disso, por se tratar de um modelo integrativo, as condições iniciais são calibráveis por 3 parâmetros de inicialização. A Tabela 7 resume todos os parâmetros.

Tabela 2 – Parâmetros calibráveis do modelo SMAP-ONS.

Parâmetro	Significado
$Eb(0)$	escoamento básico inicial
$Tu(0)$	teor de umidade do solo inicial
$Sup(0)$	escoamento superficial inicial
$k1t$	constante de recessão do escoamento para planícies (dia)
$k2t$	constante de recessão do primeiro escoamento superficial (dia).
$k3t$	constante de recessão do escoamento da superfície/planícies (dia).
$k2t2$	constante de recessão do segundo escoamento superficial (dia).
kk	constante de recessão do escoamento básico (dia).
$Capc$	capacidade de campo (%)
Ai	abstração inicial (mm)
Str	capacidade de saturação do solo (mm)
$Crec$	parâmetro de recarga subterrânea (%)
$Ecof$	coeficiente de ajuste da evapotranspiração da bacia.
$Ecof2$	coeficiente de ajuste da evapotranspiração de planície de inundação.
$H1$	altura representativa para início do segundo escoamento superficial (mm)
H	altura representativa para transbordamento para planícies (mm)
$Pcof$	coeficiente de ajuste da precipitação

3.1.1 Processamento de Precipitação - ONS

Apesar desta etapa não fazer parte da tese, esta subseção apresenta a atual metodologia de otimização do modelo SMAP pelo ONS, importante para justificar mudanças propostas na tese. Em cada iteração, um valor unitário de precipitação alimenta o sistema. Assim, uma otimização busca, a partir de séries temporais passadas de chuvas observadas em estações pluviométricas, corrigir a alimentação do SMAP com uma chuva ótima.

Primeiramente é calculada uma precipitação média da bacia $Pb(t)$ por meio da Equação 3.27.

$$Pb(t) = P1(t)ke_1 + P2(t)ke_2 + \dots + Pn(t)ke_n \quad (3.27)$$

$$(3.28)$$

$$ke_1 + ke_2 + \dots + ke_n = 1$$

onde:

- $Pb(t)$: precipitação média observada na bacia, no instante de tempo t (mm/dia)
- $P1(t), P2(t), \dots, Pn(t)$ são as observações de cada estações consideradas na bacia.
- $ke_1; ke_2; \dots; ke_n$ são os coeficientes de representação espacial de cada estação.

Em seguida, o modelo calcula a precipitação média do dia t ($Pd(t)$), definida por uma média ponderada da precipitação observada dos dias próximos a t , conforme mostrado na seguinte equação:

$$Pd(t) = [Pb(t-n) \cdots Pb(t) Pb(t+1) Pb(t+2)] \cdot \begin{bmatrix} kt_{-n} \\ kt_{-n+1} \\ \vdots \\ kt_0 \\ kt_{t+1} \\ kt_{t+2} \end{bmatrix} \quad (3.29)$$

em que $kt_{-n}, kt_{-n+1}, kt_0, kt_{t+1}$ e kt_{t+2} são os coeficientes de representação do tempo e n é o número de dias no passado considerado nesta ponderação de tempo. Tais coeficientes são fixos, uma vez calibrados não fazem parte do processo iterativo de otimização e operação do SMAP pelo ONS.

Finalmente, a precipitação considerada $P(t)$ é obtida multiplicando $Pd(t)$ pelo fator $Pcof$, conforme mostrado em 3.30. $Pcof$ deve ser calibrado com os demais parâmetros, para garantir o balanço hídrico.

$$Precp(t) = Pd(t) \cdot Pcof \quad (3.30)$$

Após um processo de calibração, o ONS utiliza o SMAP para realizar uma previsão semanal de vazão. Neste processo, é alimentado o modelo recém-calibrado com as médias aritméticas das chuvas previstas nas quadrículas pertencentes à bacia. Portanto, a otimização apenas corrige o modelo para realizar a previsão partindo de um ponto mais

acurado. Neste caso, a Equação (3.27) é substituída da pela (3.31), onde contabiliza-se cada i -ésima quadrícula, em um total de n_q .

$$Pb(t) = \frac{1}{n_q} \sum_{i=1}^{n_q} Pi(t) \quad (3.31)$$

3.1.1.1 Calibração do SMAP-ONS

O atual modelo do ONS calibra o modelo SMAP semanalmente, ajustando a curva de vazão observada vs calculada para os últimos 30 dias. Para isso, utiliza o algoritmo de busca evolutivo bio-inspirado BAT, baseado no comportamento dos morcegos. As variáveis otimizadas são os coeficientes espaciais de cada estação (vide Equação 3.27) e a inicialização dos reservatórios. É importante mencionar que os parâmetros de bacia (vide Tabela 7) utilizados pelo ONS são fixos, e pré-estabelecidos para cada sub-bacia.

A Figura 9 apresenta o diagrama de otimização do SMAP-ONS. Através de diversas possíveis métricas de avaliação, o modelo SMAP fornece o valor da função objetivo, correlacionando as vazões geradas com as observadas. O algoritmo BAT, por sua vez, encontra os valores de inicialização dos reservatórios no trigésimo dia anterior e os coeficientes das chuvas. Desta forma, tem-se, no momento da previsão, um cenário da sub-bacia corrigido.

Para avaliação da função objetivo, existem diversas formas de mensurar o erro entre a vazão observada e a computada. O ONS faz o uso de uma combinação de duas métricas, visando uma melhor adequação operacional. O erro médio percentual absoluto (MAPE), Equação (3.32), tende a adequar a curva no período da seca. Já o coeficiente de eficiência de Nash-Sutcliffe (NSE), equação (3.33) é mais sensível a vazões altas. Desta maneira, faz-se o uso da distância multicritério (DM) combinando-se as duas métricas, através da (3.34) (ONS, 2012).

$$MAPE = \frac{100}{N} \times \sum_{i=1}^N \left| \frac{Q(i) - Qobs(i)}{Qobs(i)} \right| \quad (3.32)$$

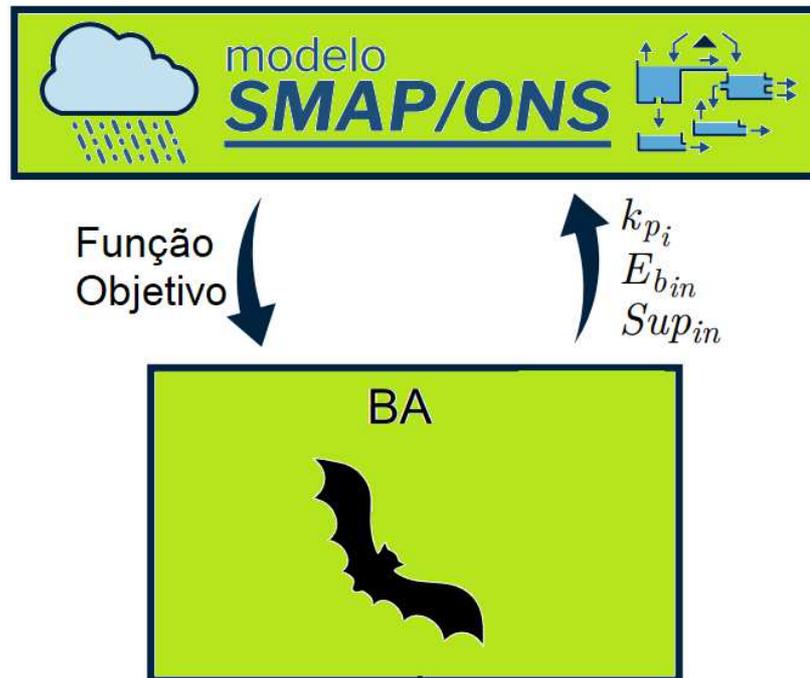
$$NSE = 1 - \frac{\sum_{i=1}^N [Q(i) - Qobs(i)]^2}{\sum_{i=1}^N [Qobs(i) - \overline{Qobs}]^2} \quad (3.33)$$

$$DM = \sqrt{MAPE^2 + (1 - NSE)^2} \quad (3.34)$$

3.2 AQUISIÇÃO E CONDICIONAMENTO DE DADOS

Assim como os modelos de previsão de vazão baseados em modelos hidrológicos, o sistema proposto necessita de dados meteorológicos e hidrológicos para calibração e

Figura 9 – Representação da calibração do SMAP-ONS



Fonte: adaptado de Leone(2021).

operação. Durante o desenvolvimento da tese, foram elaborados *softwares* para aquisição e processamento dos dados necessários. Alguns destes processos foram acrescidos de inovações em relação aos métodos utilizados comumente.

3.2.1 Precipitação

Atualmente a maioria dos modelos utiliza informações de precipitação observada provenientes de estações telemétricas. Como apresentado na seção anterior, o ONS alimenta a precipitação do SMAP a partir de um processo de otimização que pondera as estações da ANA (Agência Nacional de Águas) inseridas na bacia.

Utilizar dados observacionais de estações podem acarretar algumas desvantagens: (i) comumente há falhas nos dados, amostras faltantes; (ii) algumas bacias apresentam poucas estações, não cobrindo de forma satisfatória toda sua extensão; (iii) os produtos de previsão meteorológica são em formato *grid*, tornando a calibração e a operação dos modelos discordantes.

Com a finalidade de serem obtidos dados de precipitação melhor distribuídos ao longo da região de estudo e menos suscetíveis a erros, foi utilizado o produto de precipitação observada MERGE (ROZANTE *et al.*, 2010), fornecido pelo CPTEC/INPE.

O produto MERGE fornece precipitações diárias para quadrículas de dimensões 10 km x 10 km, resultado da combinação de observações de satélite com dados de estações em solo. Os dados são gratuitos, de fácil acesso e disponibilizados em tempo real no FTP do

INPE. Outra vantagem do MERGE em relação às estações da ANA é a compatibilidade com os produtos de previsão de tempo, que igualmente utilizam distribuição em *grid*. O acesso pode ser realizado através do seguinte modelo de endereço:

```
http://ftp.cptec.inpe.br/modelos/tempo/MERGE/GPM/DAILY/YYYY/MM
/MERGE_CPTEC_YYYYMMDD
```

Os dados fornecidos pelo produto MERGE são referentes apenas aos dados de precipitação observada. Ou seja, é utilizado para a calibração do modelo durante a etapa de treinamento. Para a realização de previsões pode ser utilizado, por exemplo, o produto BAM (*Brazilian Atmospheric Model*) (FIGUEROA *et al.*, 2016), que fornece previsões em um horizonte de 11 dias. O acesso aos dados do BAM pode ser realizado através do seguinte modelo de endereço:

```
http://ftp.cptec.inpe.br/modelos/tempo/BAM/TQ0666L064/recortes/regioes/
YYYY/MM/DD/00/BAM_CO_YYYYMMDD00
```

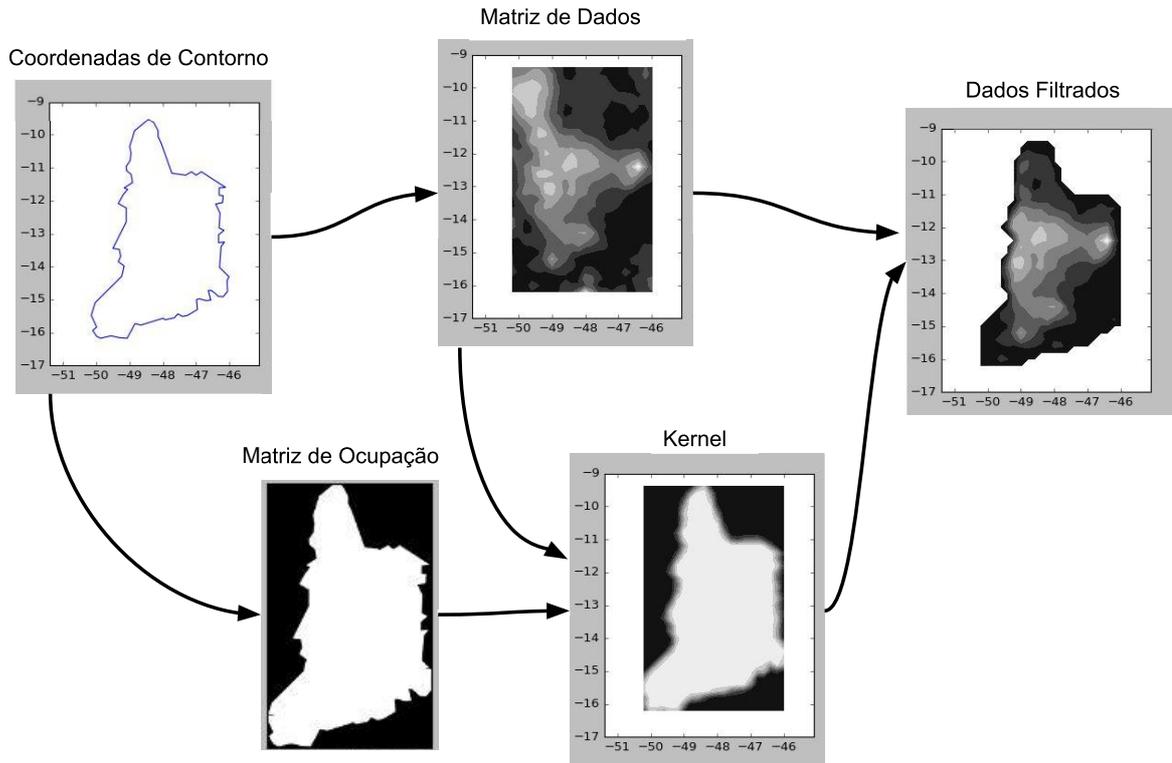
Para a utilização dos links acima, basta apenas substituir YYYY, MM e DD pelos números referentes ano, mês e dia respectivamente.

Tanto o MERGE quanto o BAM apresentam recortes com *grids* que englobam todo o território brasileiro. Para a utilização em modelos hidrológicos, é necessário um recorte sobre as áreas de escoamento dos reservatórios. Assim, foi desenvolvida uma metodologia de filtragem, aplicada com base nas coordenadas geográficas (latitude e longitude) que formam o contorno com os limites da região. A Figura 10 apresenta o processo de recorte dos dados.

- A partir dos valores máximo e mínimo de latitude e longitude das coordenadas de contorno, a **Matriz de Dados** é importada através da biblioteca *Py3grads*.
- A **Matriz de Ocupação** é formada através da clusterização da região interna das coordenadas. Assim, a região interna possui valor 1 e a externa 0.
- A Matriz de Ocupação é interpolada para ter as mesmas dimensões da Matriz de Dados, formando a **Matriz Kernel**.
- Por fim, a combinação das duas matrizes neutraliza as regiões fora do limite, resultando apenas em **Dados Filtrados**.

A partir da matriz de dados filtrados, é obtida a precipitação de entrada do modelo SMAP *Precp*. Com essa abordagem, não há a necessidade de ponderamento espacial, já que os modelos em *grid* englobam a região de forma homogênea. O valor médio é obtido

Figura 10 – Fluxograma do processamento dos dados de chuva



Fonte: elaborada pelo autor.

facilmente pela Equação 3.35, a matriz de dados filtrados também será utilizada por uma arquitetura de *Deep Learning*, que será descrita em tópicos seguintes.

$$Precp = \frac{\sum(DadosFiltrados)}{\sum(Kernel)} \quad (3.35)$$

3.2.2 Evapotranspiração

Como relatado na Seção 3.1, o SMAP precisa de informações de evapotranspiração potencial (Eto) diária para estimar as perdas d'água para a atmosfera. Esta informação não é fornecida por nenhuma agência governamental, sendo necessário um cálculo baseado em observações meteorológicas: temperatura, umidade, radiação, velocidade do vento, etc.

Por questão de simplicidade e facilidade no acesso aos dados, optou-se pela aplicação da equação Hargreaves e Samani (HARGREAVES; SAMANI, 1985), que depende apenas dos dados de temperaturas máxima e mínima diárias da região (T_{max}, T_{min}). Ainda, o estudo de Lima *et al.* (2016) analisou este método em diversas estações brasileiras. Hargreaves e Samani apresentou forte correlação positiva com a equação de Penman-Monteit, cuja precisão é recomendada pela *Food and Agriculture Organization* (FAO) e que utiliza 10 variáveis baseadas em observações meteorológicas.

Além disso, o modelo SMAP contém um coeficiente de calibração específico para

Evapotranspiração (Ecof), atenuando erros de super-estimação ou sub-estimação. Assim, não se justifica o uso de Penman-Monteit neste trabalho ao inserir grande complexidade para aquisição e manipulação dos dados. A equação de Hargreaves e Samani é descrita em 3.36. Além das temperaturas, o modelo também necessita do coeficiente de radiação, que varia conforme o mês do ano. Os valores de Q_o , para as latitudes que compõem as bacias do presente estudo são apresentados na Equação 3.37.

$$E_{to} = 0.0023 * Q_o[mes] * \sqrt{T_{max} - T_{min}} \left(17.8 + \frac{T_{max} + T_{min}}{2} \right) \quad (3.36)$$

$$Q_o = [15,9 \ 15,7 \ 15,0 \ 13,8 \ 12,4 \ 11,6 \ 11,9 \ 13,0 \ 14,4 \ 15,3 \ 15,7 \ 15,7]$$

Os dados de temperatura aplicados neste modelo são obtidos através de estações meteorológicas do Instituto Nacional de Meteorologia (INMET). Os dados de temperatura mínimas e máximas das estações podem ser obtidos em tempo real através do modelo de endereço a seguir (vide Figura 11), onde XXXX é código da respectiva estação:

<https://tempo.inmet.gov.br/TabelaEstacoes/XXXX>

Figura 11 – Serviço do INMET para fornecimento de observações meteorológicas

Estação: BRASILIA (A001)
Download CSV

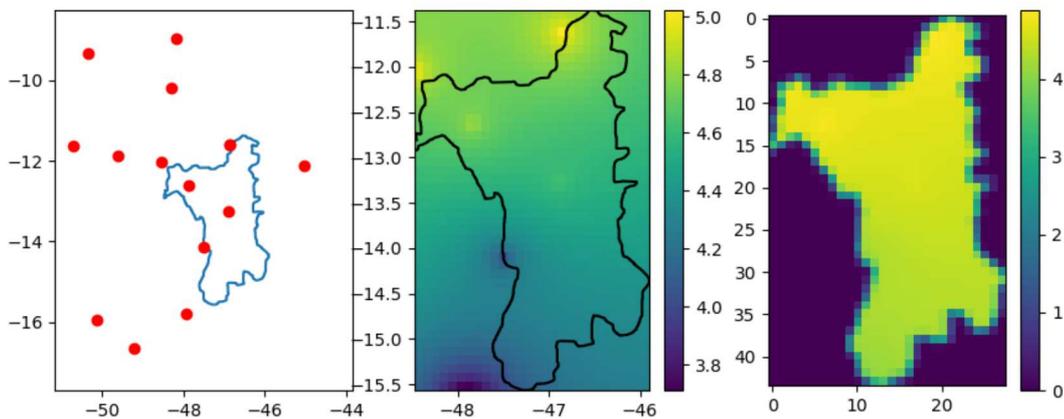
Data	Hora UTC	Temperatura (°C)			Umidade (%)			Pto. Orvalho (°C)			Pressão (hPa)			Vento (m/s)			Radiação KJ/m²	Chuva mm
		Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Inst.	Máx.	Mín.	Vel.	Dir.	Raj.		
01/12/2020	0000	21,0	23,3	21,0	62,0	63,0	53,0	13,5	13,8	13,2	886,8	886,8	886,4	1,0	26,0	2,9	0,0	
01/12/2020	0100	20,1	21,4	20,1	67,0	67,0	61,0	13,8	13,8	13,4	887,4	887,4	886,8	1,0	34,0	2,4	0,0	
01/12/2020	0200	19,6	20,1	18,7	69,0	73,0	67,0	13,8	14,3	13,5	887,4	887,6	887,4	0,7	57,0	1,9	0,0	
01/12/2020	0300	18,3	19,6	17,7	76,0	79,0	69,0	14,1	14,3	13,4	887,4	887,5	887,4	0,3	312,0	1,3	0,0	
01/12/2020	0400	17,1	18,4	17,1	82,0	83,0	75,0	14,1	14,4	13,6	886,9	887,4	886,9	0,9	287,0	1,7	0,0	
01/12/2020	0500	17,2	17,3	16,8	81,0	83,0	80,0	13,9	14,2	13,4	886,3	886,9	886,3	0,5	333,0	1,3	0,0	
01/12/2020	0600	17,3	17,4	16,7	78,0	82,0	78,0	13,5	13,9	13,4	885,9	886,3	885,9	0,9	25,0	1,6	0,0	
01/12/2020	0700	16,2	17,3	16,1	83,0	84,0	78,0	13,3	13,9	13,1	885,8	885,9	885,7	1,0	7,0	1,6	0,0	
01/12/2020	0800	16,4	16,5	15,9	86,0	86,0	81,0	14,0	14,0	13,2	886,3	886,3	885,8	0,7	155,0	2,4	0,0	
01/12/2020	0900	16,3	17,2	15,8	91,0	91,0	85,0	14,8	14,9	14,0	886,7	886,7	886,3	1,0	123,0	1,8	7,94	
01/12/2020	1000	22,0	22,0	16,2	64,0	91,0	63,0	14,9	16,6	14,5	887,5	887,5	886,7	2,1	88,0	3,4	471,48	

Fonte: <https://tempo.inmet.gov.br/TabelaEstacoes/A001>. Acessado em 02/01/2021;

Para facilitar o processo, foi desenvolvido um *software* em *Python* para aquisição automática dos dados de estações próximas às regiões de estudo. Para a aplicação do modelo SMAP, deve ser obtida a evapotranspiração unitária média, portanto, um método de ponderamento espacial das estações do INMET é proposto e aplicado.

Conforme ilustra a Figura 12, são primeiramente selecionadas ne estações nos arredores da região de escoamento do reservatório. Para cada estação, são computadas as evapotranspirações $E[i]$ através da Equação 3.36. Com base nos dados de cada estação, é construída uma matriz contendo os valores médios de evapotranspiração no interior de quadrículas de dimensões 20 km x 20 km que compreende a região de estudo. Para cada ponto da matriz, a evapotranspiração em uma quadrícula $Eto_{n,m}$ é estimada através de uma ponderação inversamente proporcional às distâncias das estações selecionadas, conforme mostra a Equação 3.37. Onde lat_i e lon_i são as coordenadas da i -ésima estação, já $lat_{n,m}$ e $lon_{n,m}$ do ponto central de cada *grid*, correspondendo a n -ésima linha e m -ésima coluna.

Figura 12 – Obtenção da evapotranspiração sobre a área de escoamento.



Fonte: elaborada pelo autor.

$$Eto_{n,m} = \frac{\sum_{i=0}^{ne} \frac{E[i]}{\sqrt{(lat_i - lat_{n,m})^2 + (lon_i - lon_{n,m})^2}}}{\sum_{i=0}^{ne} \frac{1}{\sqrt{(lat_i - lat_{n,m})^2 + (lon_i - lon_{n,m})^2}}} \quad (3.37)$$

Para alimentar o modelo SMAP com um valor unitário, basta computar a média das quadrículas internas. Assim como conduzido para os dados de precipitação, anulam-se os valores da matriz referentes dos pontos externos ao contorno da região utilizando novamente a matriz *Kernel*, conforme (3.38).

$$\frac{\sum_{i=1}^m \sum_{i2=1}^n Eto_{i1,i2}}{\sum_{i=1}^m \sum_{i2=1}^n Kernel_{i1,i2}} \quad (3.38)$$

3.2.3 Vazões Afluentes

No treinamento dos modelos, utilizam-se vazões observadas para serem confrontadas com as calculadas. Em âmbito nacional, a Agência Nacional de Água (ANA) administra as informações hidráulicas de todo o sistema, fornecendo os dados gratuitamente.

A ferramenta desenvolvida em *Python* acessa diretamente as tabelas no sistema da ANA e importa para a ferramenta as vazões diárias de interesse, através da manipulação do endereço de acesso. A Figura 13 exemplifica o serviço utilizado.

Figura 13 – Sistema da Agência Nacional de Águas para o fornecimento de vazões no SIN.

The screenshot shows the SAR system interface. The main heading is 'Dados de operação dos reservatórios SIN'. Below it, there are filter options: 'Estado' (Todos Estados), 'Reservatório*' (LAJEADO), and 'Período*' (07/11/2020 to 01/03/21). A 'Buscar' button is present. Below the filters, there is a table of records with the following columns: Código do Reservatório, Reservatório, Cota (m), Afluência (m³/s), Defluência (m³/s), Vazão Vertida (m³/s), Vazão Turbinada (m³/s), Vazão Natural (m³/s), Volume Útil (%), Vazão Incremental (m³/s), and Data da Medição.

Código do Reservatório	Reservatório	Cota (m)	Afluência (m³/s)	Defluência (m³/s)	Vazão Vertida (m³/s)	Vazão Turbinada (m³/s)	Vazão Natural (m³/s)	Volume Útil (%)	Vazão Incremental (m³/s)	Data da Medição
19132	LAJEADO	211,78	1614,50	1063,00	0,00	1063,00	1127,30			07/11/2020
19132	LAJEADO	211,79	1330,21	1261,00	0,00	1261,00	1131,11			08/11/2020
19132	LAJEADO	211,77	1274,69	1413,00	0,00	1413,00	1109,97			09/11/2020
19132	LAJEADO	211,81	1294,21	1225,00	0,00	1225,00	1080,73			10/11/2020

Fonte: <<https://www.ana.gov.br/sar0/MedicaoSin>>. Acessado em 02/01/2021;

A ANA fornece diversos tipos de vazões nos principais reservatórios do SIN, conforme Tabela 3. Para operar modelos chuva-vazão, deve-se utilizar a vazão natural ou a vazão incremental natural, já que os mesmos não estão suscetíveis à operação das usinas. Como as séries de vazão incremental das UHE's em estudo não estão atualizadas no sistema na data da pesquisa, optou-se por utilizar a vazão natural, que considera toda a área de drenagem da usina que se encontra o reservatório, e o volume total de água que passaria em condições naturais pelo ponto do rio situado na barragem.

Tabela 3 – Vazões fornecidas pela ANA/ONS.

Tipo de Vazão Diária	Significado
Afluência	Média de vazão real que atingiu certo reservatório em determinado dia. É computada com base na defluência e na variação da altura de lâmina d'água (cota) da represa.
Defluência	Média de vazão real que atingiu certo reservatório em determinado dia. É estimada a partir da soma das vazões vertida e turbinada.
Vazão Vertida	Média do fluxo de água operada nos vertedouros, ou seja, água que a UHE não utilizou para geração.
Vazão Turbinada	Média da vazão utilizada para geração em determinado dia.
Vazão Natural	Cálculo da média de vazão que ingressaria no reservatório se não houvesse manipulação antrópica. Ou seja, compensa a operação em cidades e UHE's à montante.
Vazão Incremental	Cálculo da média de fluxo que ingressou ao reservatório pela área de escoamento incremental, referente a sub-bacia entre a UHE a montante e a UHE em estudo. Ou seja, despreza a parcela d'água que veio da UHE à montante, correlacionando a defluência da UHE a montante com a afluência da UHE em estudo.

3.3 PROPOSTA DE OTIMIZAÇÃO DO MODELO SMAP

Antes de operar o modelo SMAP, é necessário um processo de otimização de seus parâmetros de bacia e inicialização. O trabalho realiza dois tipos de calibração no modelo: global e local.

3.3.1 Otimização Global

A primeira etapa do processo de calibração e otimização dos modelos consiste na determinação de parâmetros de bacia globais do modelo SMAP. Esses parâmetros, vide Tabela 7, representam genericamente a região de estudo considerando um longo histórico de vazões observadas.

A calibração desses parâmetros, nesta etapa, consiste em um problema de otimização multimodal. Devido à característica multimodal do problema, foi aplicado um algoritmo populacional baseado em inteligência computacional.

Nesse tipo de técnica, cada indivíduo da população representa uma possível solução

para o problema. Dessa forma, cada indivíduo possui os valores possíveis para cada parâmetro do modelo e, então, são obtidas as vazões para aquele indivíduo. A vazão obtida é comparada com a que foi observada no período disponível no histórico através da métrica estatística MAE (Mean Absolute Error), descrita pela Equação 3.39. Onde Q_{calc} é a vazão calculada pelo modelo, e Q_{obs} observada no mesmo período.

$$MAE = \frac{\sum_{i=0}^n abs(Q_{calc,n} - Q_{obs,n})}{n} \quad (3.39)$$

Para o processo de otimização desta etapa, utilizou-se o algoritmo evolutivo Differential Evolution (DE) (STORN; PRICE, 1997), através da *toolbox* de otimização da biblioteca *SciPy* (SCIPY, 2020). Assim, como diversos outros algoritmos evolutivos, é baseado em uma população de soluções que, geração a geração, sofre mutação e *crossover*, vide algoritmo 1. Ao final do processo, têm-se os parâmetros globais do SMAP de uma região, que vão servir como base para a Otimização local, apresentada na próxima seção.

Algorithm 1: Pseudocódigo DE.

Dados Iniciais: NP: tamanho da população, F: fator de mutação,
 CR: probabilidade de cruzamento, NP = número de indivíduos,
 n= número de parâmetros; função objetivo f
 Gerar população inicial \mathbf{x} soluções aleatórias no espaço de busca;
while (*condição não satisfeita*) **do**
 | **for** ($j \leftarrow 1 : NP$) **do**
 | | gerar 3 inteiros randômicos, $r1, r2, r3$; $r1 \neq r2 \neq r3 \neq j$
 | | gerar 1 inteiro randômico i_{rand} entre 1 e n
 | | **for** ($i \leftarrow 1 : n$) **do**
 | | | **if** ($rand(0,1) < CR$ ou $i = i_{rand}$) **then**
 | | | | $u_{j,i} = x_{r1,i} + F * (x_{r2,i} - x_{r3,i})$
 | | | **else**
 | | | | $u_{j,i} = x_{j,i}$
 | | | **end**
 | | **end**
 | | **end**
 | | **if** ($f(\mathbf{u}_j) < f(\mathbf{x}_j)$) **then**
 | | | $\mathbf{x}_j = \mathbf{u}_j$
 | | **end**
end
return x_{best}

Fonte: elaborada pelo autor.

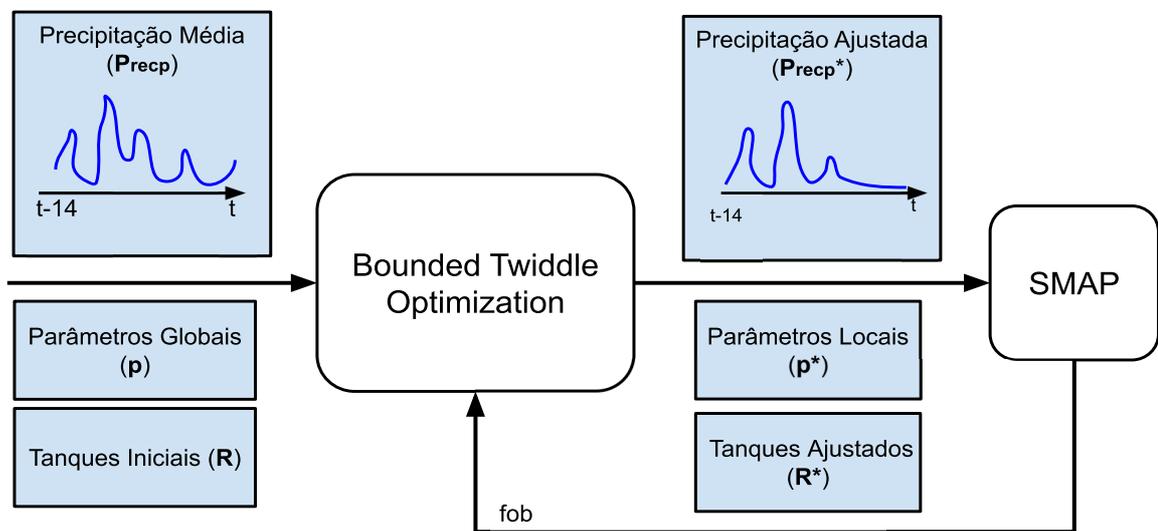
3.3.2 Otimização Local

Com a finalidade de aprimorar a assertividade do modelo, pequenas correções são realizadas através de uma otimização local do modelo SMAP, calibrando um histórico

dos 14 dias anteriores à previsão. Este processo possui dois objetivos: (i) corrigir os erros integrativos por pequenos ajustes nos valores iniciais dos reservatórios e chuva incidente no período; (ii) melhorar a caracterização da bacia no estado atual, permitindo pequenos ajustes dos parâmetros globais.

Como o SMAP gera entradas para o DL, todas as soluções partem dos parâmetros ótimos globais encontrados na etapa anterior, permitindo apenas pequenos ajustes finos. Assim, não haverá divergência abrupta na caracterização do SMAP, o que poderia confundir o reconhecimento de padrões da Rede Neural. A Figura 14 apresenta o diagrama da metodologia de ajuste local do Modelo SMAP.

Figura 14 – Ajuste local do Modelo SMAP



Fonte: elaborada pelo autor.

3.3.2.1 Bounded Twiddle Optimization

O algoritmo de busca local *Twiddle* é a técnica aplicada para a resolução da calibração nesta etapa. Proposto por Sebastian Thrun em um curso de inteligência artificial (THRUN, 2018), o Twiddle não é um algoritmo difundido na literatura. É um método de implementação simples, baseado em busca local e que não requer informações extras sobre a função que deve ser minimizada. Basicamente é um algoritmo que manipula parâmetro a parâmetro de uma solução de forma cíclica, aumentando ou diminuindo o passo conforme o impacto na função objetivo. Para uma melhor aplicação no modelo SMAP, foram propostas algumas mudanças no algoritmo original de modo a delimitar o processo. Este processo visa ajustes suaves, sem divergências significativas entre diferentes rodadas.

1. Uma normalização foi realizada com base nos valores dos limites inferior e superior

(**lb**, **ub**) de cada variável. Assim, as variáveis serão limitadas entre 0 e 1, havendo uniformidade entre todas.

2. A taxa de atualização para cada variável é restrita entre $dxmin$ e $dxmax$. Assim, não há dissipação ou explosão de nenhuma taxa, havendo uma convergência suavizada. Problemas como o SMAP, altamente multimodal e com variáveis altamente vinculadas, podem convergir para situações absurdas em teoria, nas quais apesar de uma boa métrica de resultado, perderia a caracterização real da bacia.

Algorithm 2: Bounded Twiddle Optimization.

```

x = scaling(x0; (lb, ub); (0, 1));
best_err = fob(inv_scaling(x));
while sum(dx) > threshold do
  for i in range(len(x)) do
    xbest = x[i];
    x[i] = x[i] + dx[i];
    x[i] = constrain(x[i]; 0; 1);
    err = fob(inv_scaling(x));
    if err < best_err then
      best_err = err;
      dx[i] = dx[i]*(1+rate);
      dx[i] =constrain(dx[i]; dxmin; dxmax);
    else
      x[i] = xbest - dx[i];
      x[i] =constrain(x[i]; 0; 1);
      err = fob(inv_scaling(x));
      if err < best_err then
        best_err = err;
        dx[i] = dx[i]*(1+rate);
        dx[i] =constrain(dx[i]; dxmin; dxmax);
      else
        x[i] = xbest;
        dx[i] = dx[i]*(1-rate);
        dx[i] = constrain(dx[i]; dxmin; dxmax);
      end
    end
  end
end
return inv_scaling(x)

```

Fonte: elaborada pelo autor.

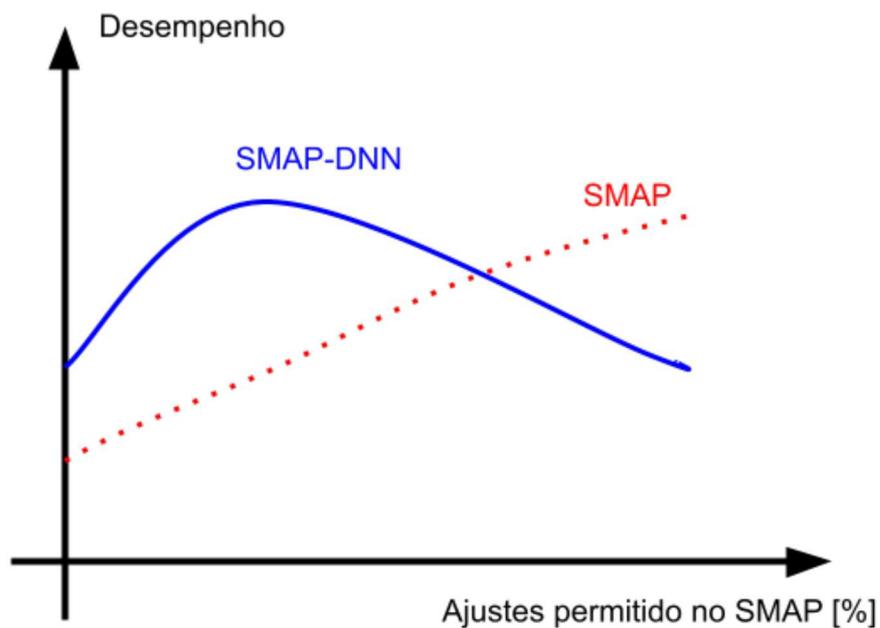
O algoritmo BTO (Bounded Twiddle Optimization) está descrito em detalhes no pseudo-código 2. x_0 é a solução inicial, x é a solução em escala atual, dx é a matriz de variações de x , $dxmin$ e $dxmax$ são definidos como 10^{-4} e 10^{-2} , respectivamente, e $fob()$

é a função que calcula o MAE entre a vazão observada e calculada através dos parâmetros em x .

3.3.2.2 BTO aplicado ao SMAP

Na aplicação do BTO no ajuste local do SMAP, visam-se pequenos ajustes nos parâmetros de bacia, condições iniciais dos reservatórios e ajuste na magnitude da série de precipitação de entrada. Assim, os erros integrativos inerentes aos modelos geológicos são amortizados, fora que os parâmetros tornam-se mais compenetrados ao atual cenário. O relaxamento das restrições foi baseado em um estudo empírico, através da observação do comportamento descrito qualitativamente na Figura 15.

Figura 15 – Comportamento das restrições da otimização local no SMAP e no modelo Híbrido



Fonte: Elaborada pelo autor.

Como característico de problemas de otimização, quanto menos restritivo a canalização do ajuste do SMAP, melhor será o seu resultado. No caso do modelo híbrido, ao melhorar a precisão do SMAP, obtém-se um melhor resultado, já que o mesmo fornecerá entradas mais acuradas para a RNA. Mas, quando são permitidas variações elevadas, estas soluções tendem a perder o padrão e dificultar o treinamento, como se diferentes modelos de bacias hidrográficas gerassem entradas de treinamento para a mesma rede neural. Assim, foram escolhidas as variações descritas a seguir:

- Entrada de precipitação **Precp**: $\pm 20\%$ dos valores observados. Essa modificação visa corrigir possíveis erros de medição e garantir o balanço hídrico.

- Parâmetros da bacia hidrográfica \mathbf{p} : $\pm 5\%$ dos valores obtidos na calibração global (primeira etapa).
- Estados iniciais dos reservatórios \mathbf{R} : $\pm 10\%$ dos valores obtidos para este dia na última iteração do algoritmo. Este ajuste, nas variáveis de estado inicial, visa compensar os erros integrativos que acompanharam o modelo até o momento da previsão.

Cada uma das 37 variáveis envolvidas no Ajuste Local poderiam possuir limites particulares. Mas, devido à explosão combinatorial, os limites foram escolhidos empiricamente e de forma generalista. Vale a pena salientar que outros algoritmos de busca local poderiam ser utilizados nessa etapa, desde que não necessitem de uma função objetivo analítica e derivável, possibilite o uso de uma solução inicial específica e respeite restrições de canalização. A escolha do BTO, além de satisfazer as premissas acima, é possível um controle de sensibilidade do passo para uma convergência suavizada e equilibrada entre as variáveis. Uma convergência equilibrada tende a manter o padrão das soluções de forma satisfatória.

3.4 MODELO HÍBRIDO SMAP-DNN

Esta Seção apresenta o produto final do trabalho, baseado na fusão do modelo analítico com uma arquitetura *deep learning*. Para um melhor entendimento do processamento dos dados realizados pela rede neural, a Seção 3.4.1 apresenta, em suma, o funcionamento das principais camadas utilizadas.

3.4.1 Deep Learning

Deep learning é um tipo específico de aprendizado de máquina, que consiste no conjunto de algoritmos capazes de aprender a partir de conjuntos de dados (BENGIO *et al.*, 2017). Com o avanço no poder de processamento dos computadores, o uso das redes neurais profundas vem crescendo exponencialmente, aplicadas em diversas áreas. As redes neurais profundas buscam modelar abstrações em alto nível a partir de várias camadas de processamento. Normalmente são definidas como redes neurais vastas, que possuem mais de uma camada oculta.

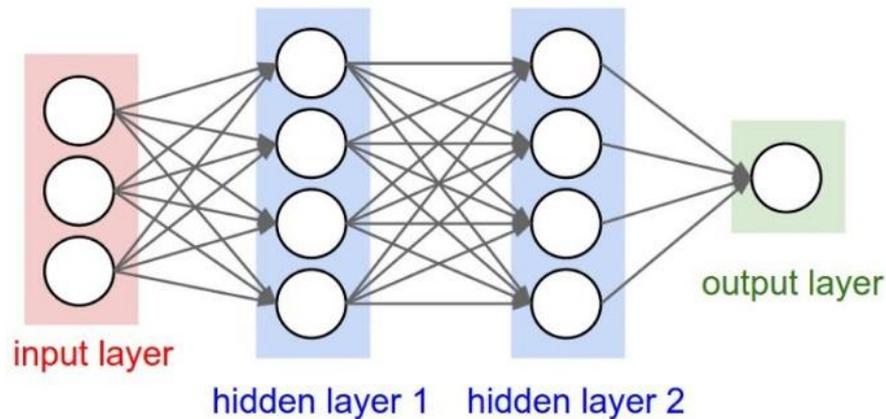
Nesta seção serão expostas algumas definições importantes, que fazem parte da compreensão da arquitetura desenvolvida neste trabalho.

3.4.1.1 Camada Densa

Camadas Densas são camadas totalmente conectadas, todos os neurônios recebem informações de todos os neurônios das camadas anteriores e não há conexões entre unidades

da mesma camada. São as camadas mais fundamentais da área de estudo de redes neurais artificiais. A Figura 16 apresenta uma ilustração de um RNA com duas camadas ocultas da categoria Densa.

Figura 16 – Exemplo de RNA com camadas densas.



Fonte:(ALLIBHAI, 2018).

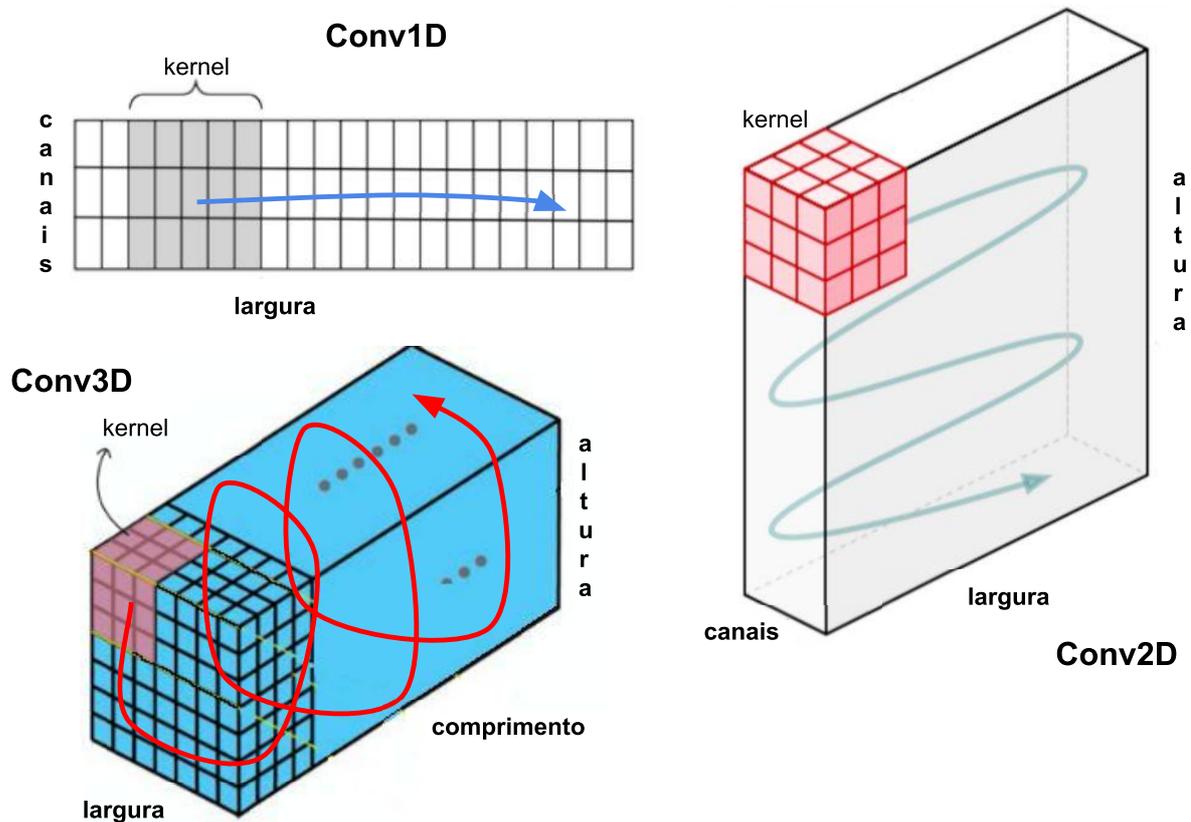
3.4.1.2 Camada Convolucionais

As camadas convolucionais, como o próprio nome diz, conectam neurônios através de convoluções. A ligação de cada neurônio da camada posterior é resultado de uma segmentação de conjuntos através do deslocamento do kernel, que varre a informação de entrada agrupando neurônios espacialmente próximos. O kernel é uma matriz treinável n-dimensional que multiplica um determinado subconjunto de entrada, os valores são somados, passam por uma função de ativação e se conectam em um neurônio na camada seguinte. Uma mesma camada convolucional pode ter vários kernels, cada um é considerado um filtro e produz um “feature map” de saída independente. Existem três tipos de camadas convolucionais (LITTLE, 2020). A Figura 17 exemplifica as 3 operações convolucionais, sendo importante mencionar que a Conv3D processa 4 dimensões, a dimensão “canais” não foi ilustrada:

- Conv1D: O kernel se move em apenas uma direção, sendo que a entrada possui 2 dimensões (largura, canais). O uso mais comum é na extração de *features* em séries temporais.
- Conv2D: O kernel se move em duas direções, sendo que a entrada possui 3 dimensões (largura, altura, canais). A maior aplicação é no reconhecimento de padrões em imagens.

- Conv3D: O kernel se move em três direções, sendo que a entrada possui 4 dimensões (largura, altura, comprimento, canais). É utilizado principalmente em extração de características de vídeos e nuvens de pontos.

Figura 17 – Ilustração dos processos convolucionais.



Fonte: Adaptado de (LITTLE, 2020).

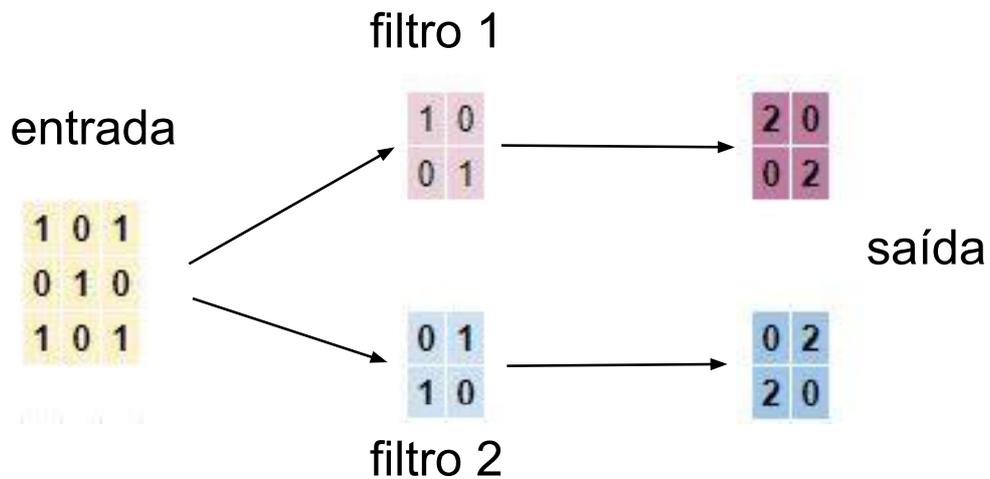
3.4.1.2.1 Parâmetro Filtro

Os Filtros são as matrizes que compõem o *kernel*. Cada filtro retorna uma saída independente, em uma dimensão paralela para a próxima camada. Estes filtros são calibradas pelo processo de treinamento, compondo os hiper-parâmetros da camada convolucional. A Figura 18 apresenta um exemplo simples com dois filtros 2×2 em uma entrada 3×3 .

3.4.1.2.2 Parâmetro Stride

Stride refere-se às unidades que o kernel salta durante a varredura do processo de convolução, que por padrão é de apenas uma unidade em cada dimensão da entrada. A escolha dos valores do *stride* influencia diretamente na dimensão da camada de saída, realizando uma operação de *downsample*. Cada iteração da convolução é conectada a um neurônio de saída, portanto, quanto maior a dimensão do stride menos convoluções serão necessárias para varrer a entrada de dados. A Figura 19 exemplifica o processo.

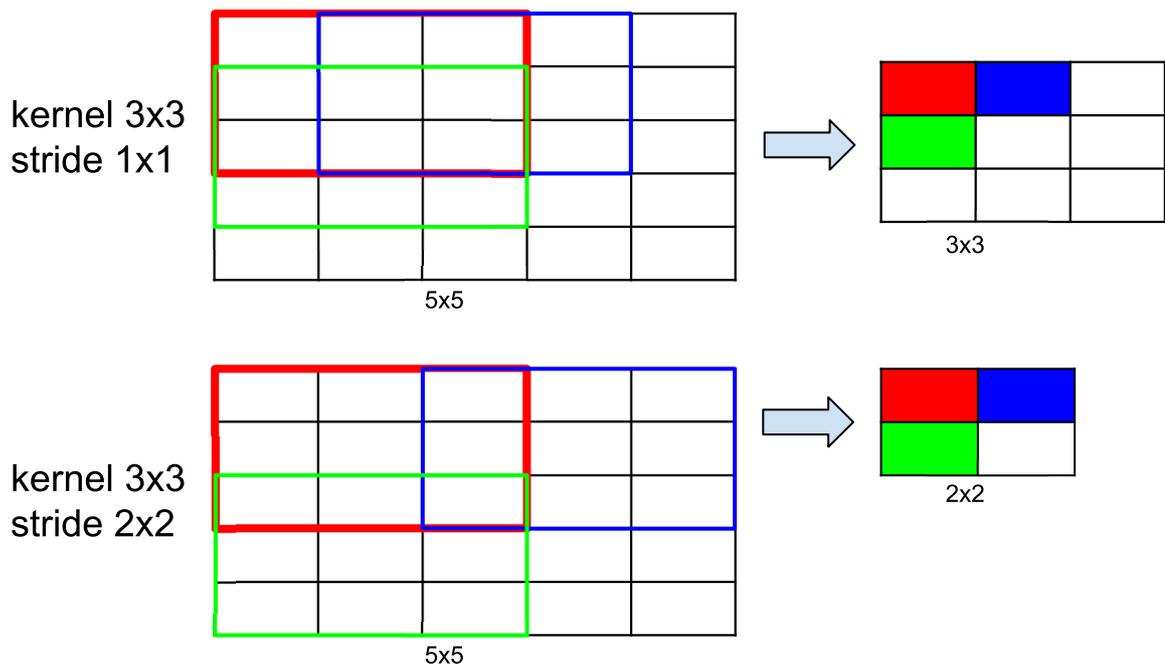
Figura 18 – Exemplo da aplicação dos filtros na camada convolucional 2D



Fonte:

Elaborada pelo autor.

Figura 19 – Parâmetros Stride da Convolução



Fonte: Elaborada pelo autor.

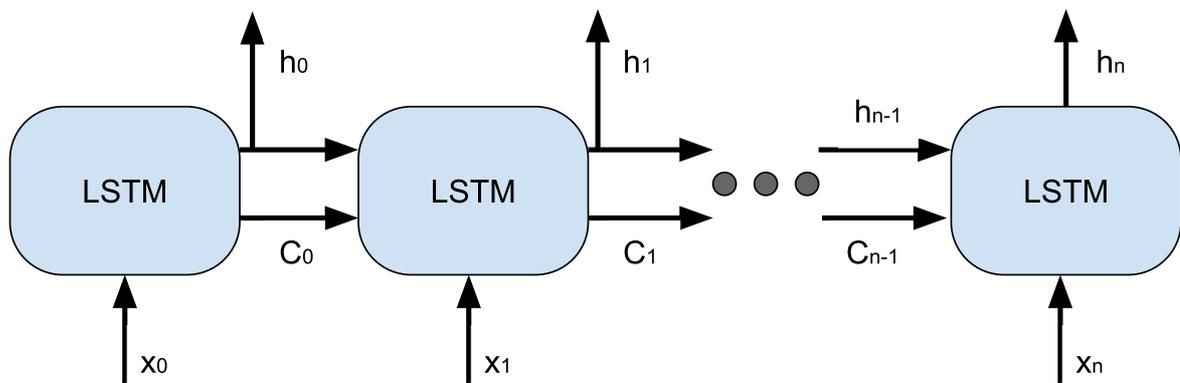
3.4.1.3 Camada LSTM

A camada LSTM (*Long short-term memory*) é uma categoria de camada recorrente. As redes recorrentes foram projetadas para reconhecer sequências de dados, ou seja, há uma relação ordenada entre as entradas. Este tipo de arquitetura permite que a informação persista, já que uma célula nesta camada recebe informações da antecessora e transmite à sucessora.

A LSTM foi criada como uma evolução das camadas recursivas, evitando o problema das dependências de longo prazo. Quando maior o número de entradas, mais difícil o treinamento. O fato dos neurônios da mesma camada estarem conectados, há uma degeneração do gradiente durante o treinamento (GERS *et al.*, 1999).

A LSTM inseriu uma célula de memória em cada unidade. Essas células formam uma cadeia que define quando a informação deve fluir, ser alterada ou ser removida. Como se em cada unidade LSTM houvesse um portão que bloqueia ou libera a passagem da informação.

Figura 20 – Ilustração de uma camada LSTM



Fonte: Elaborada pelo autor.

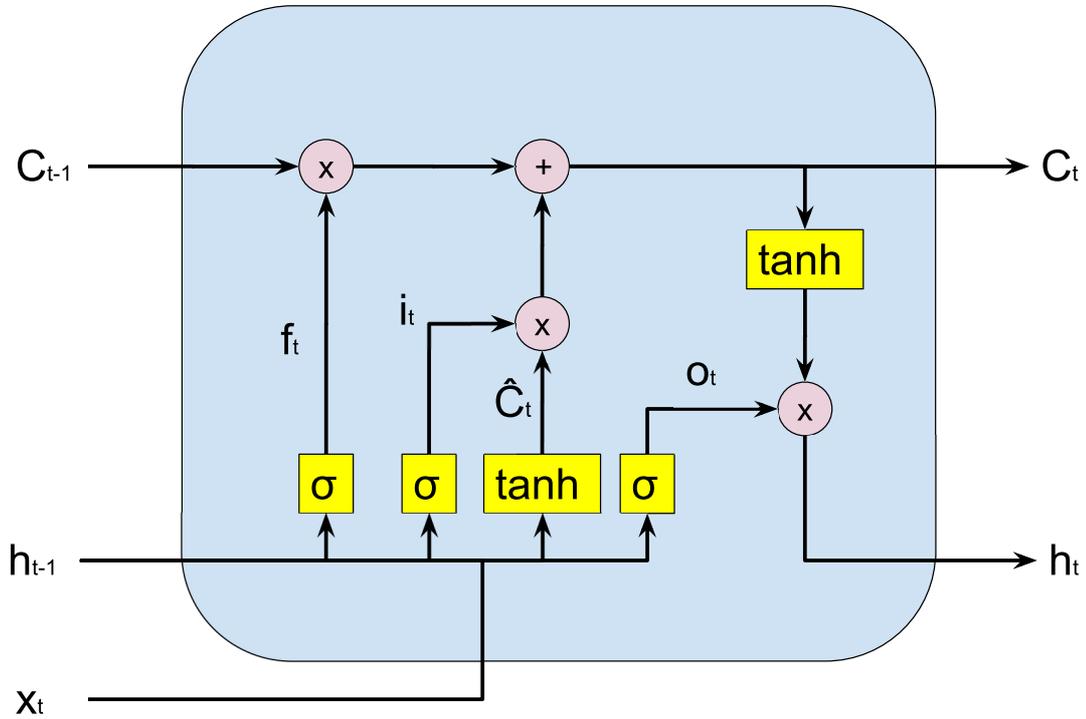
A Figura 20 apresenta uma camada LSTM que ilustra o processo descrito. Cada i -ésima unidade LSTM recebe uma entrada da série x_i e fornece uma saída h_i para a próxima camada da RNA. Cada saída h_i também é fornecida como entrada à unidade LSTM paralela posterior. Por fim, uma cadeia de portões através de saídas C_i controla o fluxo de informações através de estados de memória de cada unidade (JUNIOR, 2019).

O fluxograma matemático da unidade LSTM está exposto na Figura 21. Primeiramente calcula-se f_t , através da equação (3.40), que confronta a entrada da célula x_t com a saída da célula anterior h_{t-1} por uma função de ativação sigmoideal. Assim, um valor entre 0 a 1 decidirá o quanto o estado de memória propagado da célula anterior C_t deve ser esquecido(atenuado). Uma analogia ao problema do trabalho pode ser realizada, ponderando o quanto da informação de precipitação anterior impactará na previsão de vazão de determinado dia.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3.40)$$

O próximo passo é atualizar o estado da célula C_t , para isso utilizam-se as equações (3.41) e (3.42) para calcular i_t e \hat{C} . O produto dessas duas variáveis é somado ao produto de C_{t-1} e f_t através da equação (3.43).

Figura 21 – Parâmetros da unidade LSTM



Fonte: Elaborada pelo autor.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3.41)$$

$$\hat{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (3.42)$$

$$C_t = f_t C_{t-1} + i_t \hat{C}_t \quad (3.43)$$

Por fim, é computada a saída da célula h_t , pelas equações 3.44 e 3.45. Os vetores W e b de todas as equações da camada, referem-se aos hiper-parâmetros calibráveis, pesos e bias.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3.44)$$

$$h_t = o_t * \tanh(C_t) \quad (3.45)$$

O uso da camada LSTM se justifica pelo fato do modelo híbrido representar um sistema integrativo, que processa séries temporais cujos valores passados impactam fortemente a resposta presente. Uma camada *feedforward* comum teria menor competência em correlacionar temporalmente as entradas.

3.4.2 Arquitetura Proposta

Diversos estudos, envolvendo arquiteturas *Deep Learning* para previsão de vazões, foram conduzidos ao longo da pesquisa. Ao utilizar apenas redes neurais, o treinamento apresentou ótima acurácia, embora tenha apresentado menor robustez em operação devido à ocorrência de *overfitting*. Neste caso, somente a precipitação e as vazões históricas não foram suficientes para caracterizar de forma satisfatória a região. A proposta da pesquisa é a exploração do acoplamento de redes neurais para melhorar os cenários obtidos pelo modelo SMAP. Este acoplamento fornece uma estimativa das condições atuais da bacia, diminui o espaço de busca e discrimina melhor as entradas, tornando o processo de treinamento mais robusto.

Após o ajuste local descrito anteriormente, tem-se uma boa estimativa do estado atual dos aspectos hidrológicos da região. Através das variáveis de estado intermediária do SMAP (Reservatórios: $[R_{sub}(t), R_{solo}(t), R_{sup}(t), R_{sup2}(t)]$), é modelada a distribuição da água nas camadas do solo. Essas informações complementam a vazão atual da região, como, por exemplo: (i) se a descarga atual é predominantemente uma vazão superficial ou dos lençóis freáticos, (ii) se o solo está saturado e as próximas chuvas tendem a escorrer na superfície abruptamente, etc.

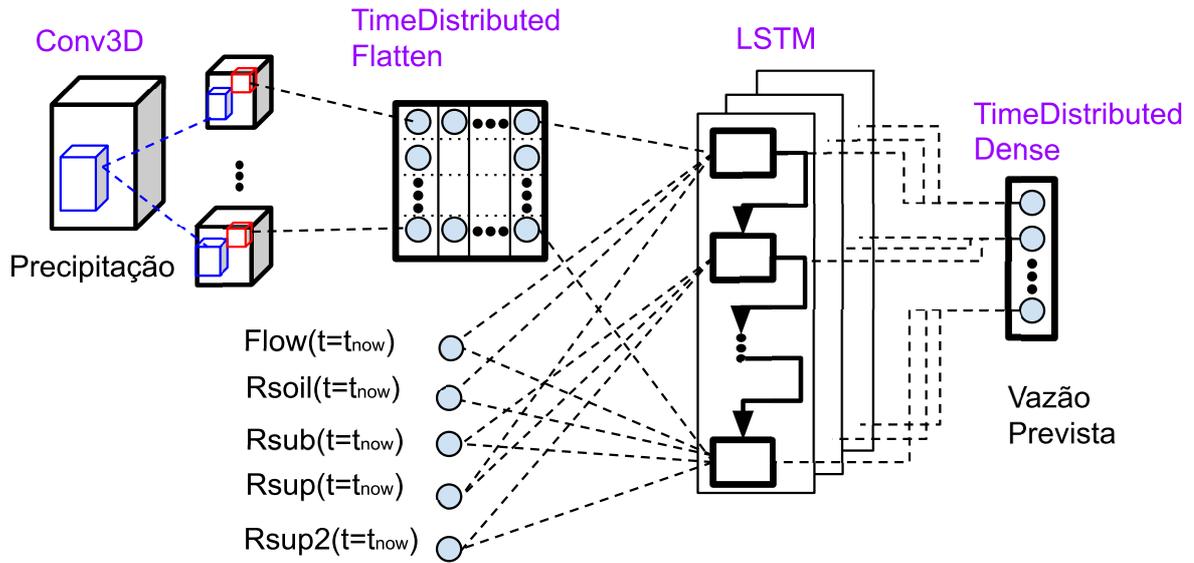
Além das informações do estado atual do SMAP no momento da previsão, a arquitetura desenvolvida possui como parâmetros de entrada: (i) A última vazão observada para ser utilizada como ponto de partida para os dados de saída da rede; (ii) previsões de precipitação para o número desejado de dias (vide Figura 7). Como a ONS trabalha com o SMAP em sistema de previsão de 7 dias, o mesmo horizonte foi utilizado.

Foi estudada uma arquitetura baseada em camadas convolucionais 3D e camada LSTM, conforme teorias descritas anteriormente. Esta RNA realiza previsões de vazão afluente, e seu design está exibido na Figura 22. Como pode ser visualizado, a presença da sequência convolucional e camada da LSTM proporcionam inúmeras camadas de processamento à rede, sendo enquadrada em uma arquitetura *Deep Learning*. A seguir será apresentada a justificativa de cada camada.

As entradas da arquitetura são:

- Quatro entradas referentes à umidade armazenada nos reservatórios no dia atual que será realizada a previsão: $[R_{sub}(t), R_{solo}(t), R_{sup}(t), R_{sup2}(t)]$.

Figura 22 – Esquema da arquitetura desenvolvida



Fonte: Elaborada pelo autor.

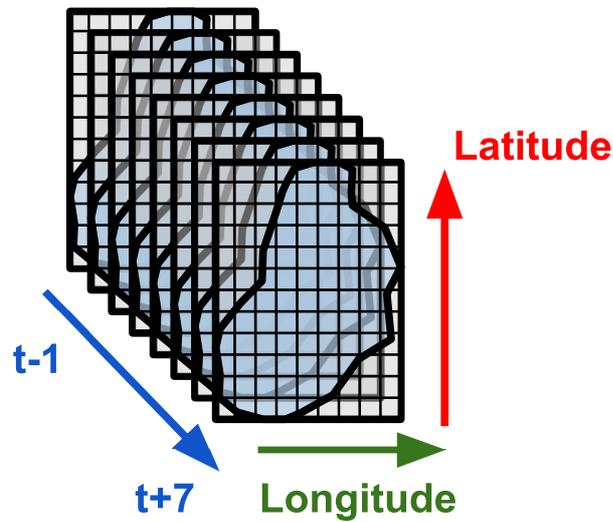
- Uma entrada referente à última medição de vazão natural $[Q_{obs}(t)]$, para a RNA levantar a curva de saída a partir deste ponto.
- Precipitação em formato grid na bacia, sendo b dias anteriores e a dias de previsão de chuva, formando um paralelepípedo de dados de entrada.

Fora a contribuição referente ao acoplamento do modelo hidrológico clássico SMAP à uma arquitetura DL, este trabalho trouxe uma inovação no contexto de processamento de dados de chuva em *grid*. Diferente do SMAP que recebe um valor médio da chuva na região, o DL processa as matrizes de dados de chuva filtrados (vide Figura 10). O objetivo é extrair os padrões espaço-temporais de cada *grid* de precipitação, para compreender de maneira mais efetiva os efeitos dessa chuva sobre a vazão.

Os dados de precipitação diária filtrados para as coordenadas de contorno, em formato de matriz, são concatenados do dia anterior ($t-1$) até sete dias a frente ($t+7$). São considerados dados de precipitação do dia anterior, visto que a dinâmica da bacia em um determinado momento pode estar associada com a precipitação observada em tempos anteriores. A Figura 23 mostra a entrada de dados de precipitação do modelo, formando um arranjo retangular em três dimensões.

Como cada quadrícula tem 10 km, tem-se um grande volume de dados de entrada. Uma sub-bacia de grande porte ultrapassa facilmente 1000 dados de precipitação para apenas um dia. Esta enorme quantidade de dados de entrada inviabiliza o treinamento de uma rede normal *feedforward* densa. Assim, para contornar o problema, tem-se o uso da camada Conv3D.

Figura 23 – Estrutura da entrada de precipitação.



Fonte: Elaborada pelo autor.

A camada Conv3D, filtra as informações mais relevantes. As dimensões do *kernel*, juntamente com as do parâmetro *stride*, reduzem drasticamente a dimensão do problema, separando e processando blocos de dados. Ao chegar na camada Time Distributed Flatten apenas as principais características dessa chuva estarão disponíveis, sendo que as informações espaciais serão misturadas e vetorizadas em uma dimensão, mantendo apenas a separação temporal das *features* de chuva. A quantidade de neurônios, dimensões da convolução e do stride é definida conforme a dimensão da bacia.

Neste momento, estas principais características espaço-temporais da chuva serão processadas juntamente com o estado atual da bacia, fornecido pela última vazão observada e volume armazenado nos reservatórios do SMAP. O procedimento é realizado pela camada LSTM, atual estado da arte em processamento de séries temporais em redes neurais. As *features* da chuva, separadas temporalmente, alimentam as unidades LSTM. Cada instância LSTM recebe os valores do estado da bacia. A saída desta camada fornece a vazão prevista.

3.4.3 Treinamento

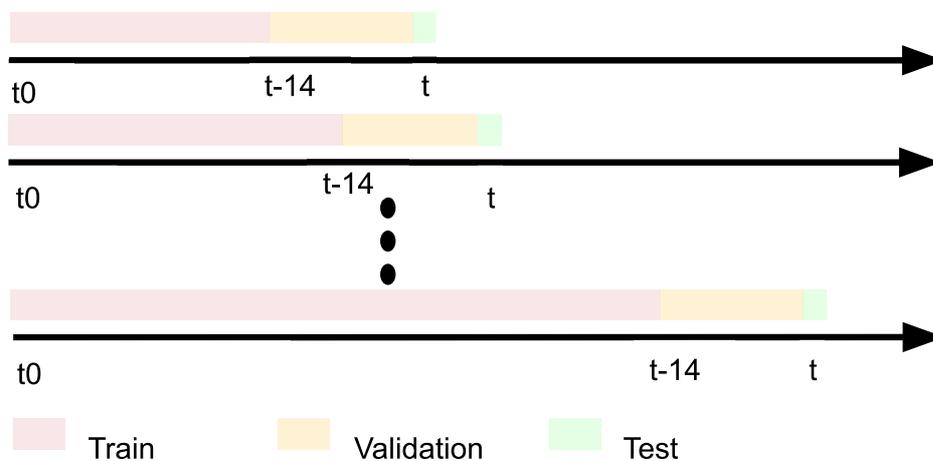
Visando uma possível fase operacional, o processo de treinamento da RNA adotou uma estratégia de *reinforcement learning* com atualização semanal. A cada sete dias, o banco de dados é atualizado e ocorre um retreinamento. Para validar a metodologia e simular uma operação a longo prazo, utilizou-se *cross-validation* (SCIKIT-LEARNING, 2019). Os conjuntos de treinamento, validação e teste são divididos em ordem cronológica. Se os conjuntos fossem divididos aleatoriamente, as características integrativas e a dependência do sistema com o tempo poderiam reduzir o potencial de previsão da ferramenta.

Foi utilizado o *cross-validation* com *Time Series Split*, onde a etapa de treinamento começa com uma parte dos dados. Os dados são adicionados com amostras mais recentes em cada execução e a rede neural é retreinada. Ao final do processo, é obtido o desempenho médio dos conjuntos de teste para todas as iterações.

A partir do histórico de dados, a rede é sempre validada com as últimas 14 amostras, enquanto o restante do histórico é usado para treinamento. O número de amostras na validação foi definido empiricamente e visando reduzir a distância entre o momento da previsão e o conjunto de treinamento. Assim que o desempenho da validação estagnar, o treinamento é interrompido. A Figura 24 apresenta o processo.

Ao final de cada treinamento, o teste é aplicado no instante t para prever vazões com 7 dias de antecedência. Neste estudo, as simulações apresentadas visam a uma comparação entre a vazão prevista pela arquitetura desenvolvida e a observada para o mesmo período. Assim, a execução da ferramenta será realizada para prever intervalos de 1 semana, até que sejam obtidos dados para comparação ao longo de todo o período especificado.

Figura 24 – Cross-Validation



Fonte: Elaborada pelo autor.

O método *Adaptive Moment Estimation* (ADAM), (KINGMA; BA, 2014), foi o otimizador usado para treinar a arquitetura. A métrica utilizada para validação também foi o Erro Médio Absoluto (MAE), como na calibração do SMAP.

A cada iteração, a RNA obtida na iteração anterior é retreinada com paciência de 10 épocas, ou seja, o processo é interrompido após 10 épocas sem melhorias no conjunto de validação.

Visando um treinamento mais robusto, para inicialização dos pesos na primeira iteração do treinamento, utilizou-se a seguinte estratégia:

- A Camada convolucional é inicializada com todos os pesos iguais a uma constante. O valor dessa constante é o valor inverso do número de neurônios do *kernel*. Assim, ela evita que os pesos dos *grids* de entrada de chuva sejam inicializados de forma caótica, saturando ou desaparecendo com os respectivos gradientes. A busca de filtros convolucionais mais suaves, em teoria, incrementa a robustez e o reconhecimento de padrão da chuva.
- Para as demais camadas, os pesos foram inicializados no padrão normal do *keras*, por uma distribuição aleatória normalizada.

3.4.4 Métricas de Validação

A comparação entre os resultados obtidos pela arquitetura proposta e as vazões observadas na região será feita por métricas estatísticas. As métricas consideradas são as seguintes: *Mean Absolute Percentage Error* (MAPE), *Mean Absolute Error* (MAE), *Nash-Sutcliffe coefficient of efficiency* (NSE), *Root Mean Square Error* (RMSE) e *Pearson's correlation coefficient* (r).

O coeficiente NSE é o mais utilizado para avaliar o desempenho de modelos hidrológicos. O coeficiente r indica a correlação entre os dados de vazão observada e prevista, o MAE representa a média do erro absoluto e o MAPE representa a média do erro absoluto em relação ao observado, em porcentagem. Finalmente, o RMSE representa um erro quadrático médio, que pondera os erros para intensificar os maiores. As métricas estatísticas descritas são calculadas conforme as seguintes expressões:

$$MAPE = \frac{100}{N} \times \sum_{i=1}^N \left| \frac{Q(i) - Q_{obs}(i)}{Q_{obs}(i)} \right| \quad (3.46)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Q(i) - Q_{obs}(i)| \quad (3.47)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [Q(i) - Q_{obs}(i)]^2} \quad (3.48)$$

$$r = \frac{\sum_{i=1}^N (Q(i) - \bar{Q})(Q_{obs}(i) - \bar{Q}_{obs})}{\sqrt{\sum_{i=1}^N [Q(i) - \bar{Q}]^2} \sqrt{\sum_{i=1}^N [Q_{obs}(i) - \bar{Q}_{obs}]^2}} \quad (3.49)$$

$$NSE = 1 - \frac{\sum_{i=1}^N [Q(i) - Q_{obs}(i)]^2}{\sum_{i=1}^N [Q_{obs}(i) - \bar{Q}_{obs}]^2} \quad (3.50)$$

onde Q representa a previsão de fluxo, Q_{obs} o fluxo observado, \bar{Q} e \bar{Q}_{obs} as respectivas médias e N a quantidade de dados. As métricas consideradas têm limites e valores ótimos diferentes, conforme mostrado na Tabela 4.

Tabela 4 – Limites das Métricas

Métrica	Mínimo	Máximo	Valor Ótimo
MAPE	0	$+\infty$	0
MAE	0	$+\infty$	0
RMSE	0	$+\infty$	0
r	0	1	1
NSE	$-\infty$	1	1

Para a calibração do modelo SMAP e treinamento da arquitetura, foi aplicado o MAE, dado que esta métrica não é influenciada por períodos de seca como o MAPE nem por períodos de cheia como RMSE. No entanto, todas as métricas são usadas para comparar o desempenho da previsão e da calibração.

3.5 ARQUITETURA DE SOFTWARE

Esta seção descreve a arquitetura de software utilizada para implementação das metodologias da tese. Apesar da abordagem poder ser programada utilizando diversas outras plataformas, o detalhamento a seguir pode facilitar a replicação do trabalho.

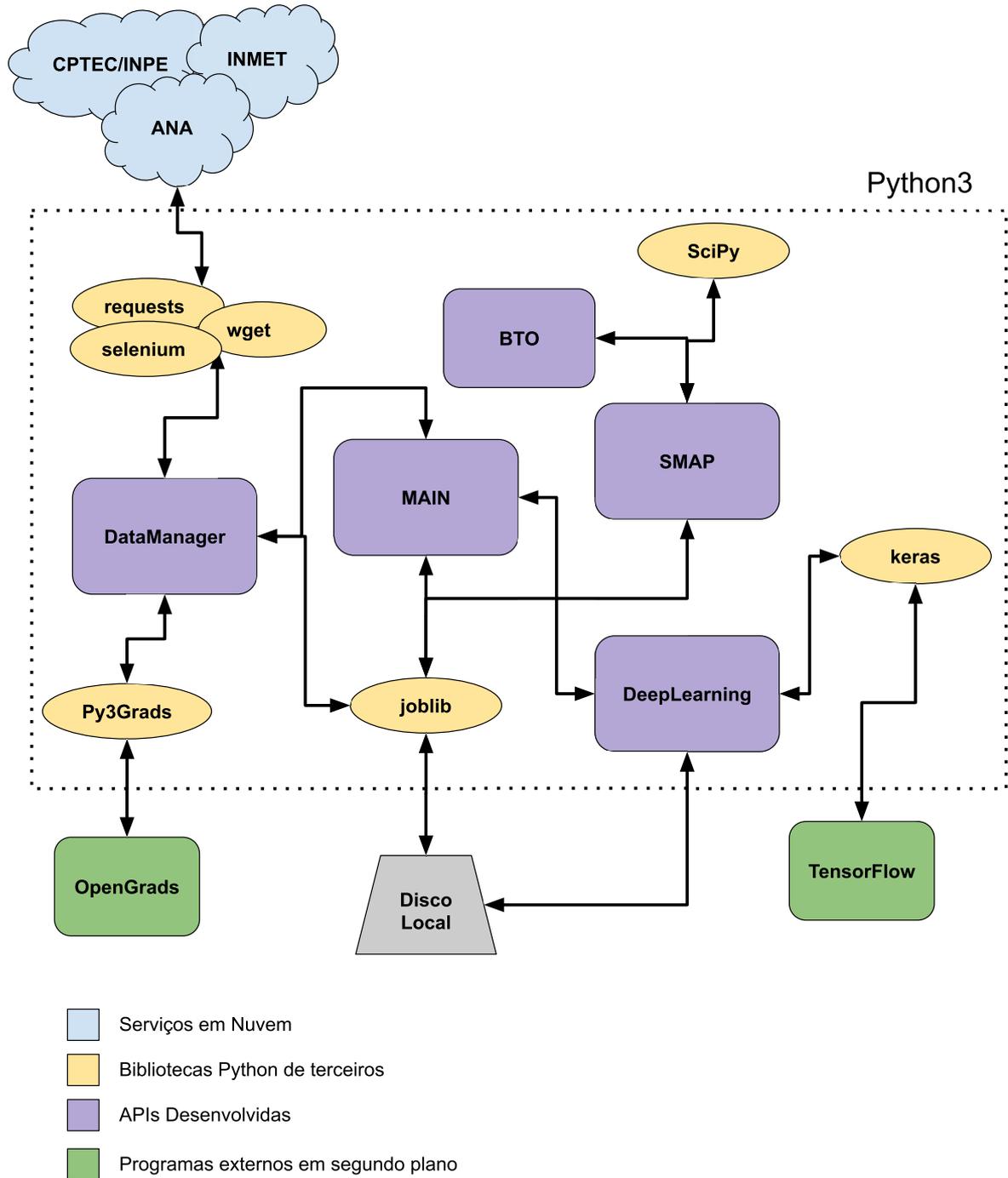
Toda a implementação foi realizada em Python 3.6, através da distribuição Anaconda, criando API's e utilizando bibliotecas existentes para a plataforma. Fora isso, alguns programas são executados em paralelo, acionados dentro das API's em Python, para condicionar os dados necessários e realizar alguns processamentos. Em Resumo, a arquitetura de *software* desenvolvida está representada na Figura 25, representando as principais API's, programas e bibliotecas utilizadas, sub-divididas em: (i) Serviços utilizados em servidores externos, (ii) Bibliotecas em Python já existentes, (iii) API's desenvolvidas em Python e (iv) programas externos executados em paralelo na máquina. Cada tópico será apresentado nas próximas sub-seções.

A Tabela 5 apresenta os principais pacotes utilizados e suas versões, facilitando a replicação do ambiente *Anaconda* no qual se realizou esta pesquisa. Contudo, não é necessário utilizar as versões apresentadas para aplicação da metodologia, mas em um mesmo ambiente *Python* pode haver conflitos entre versões de pacotes.

Tabela 5 – Principais Pacotes utilizados

Pacote	Versão
conda	4.10.3
joblib	0.14.0
keras-gpu	2.2.4
matplotlib	2.2.2
numpy	1.13.3
py3grads	1.0
python	3.6.5
python-wget	3.2
requests	2.26.0
scipy	1.2.0
selenium	3.141.0
tensorflow	1.12.0

Figura 25 – Arquitetura de Software



Fonte: Elaborada pelo autor.

3.5.1 Acesso aos servidores na Nuvem

Este tópico foi detalhado na Seção 3.2. Basicamente consiste em servidores de órgãos governamentais que fornecem em tempo real os dados necessários para operação do projeto. Algumas bibliotecas em *Python* foram necessárias para este acesso aos dados,

sendo:

- **wget:** Biblioteca utilizada para o *download* dos arquivos dos servidores do CP-TEC/INPE. Basta informar o URL com o arquivo e o endereço para armazenamento no disco local.
- **requests:** Biblioteca necessária para a leitura dos arquivos dos servidores da ANA. A partir da manipulação do url com os dados de interesse, toda a informação da página é transferida para variáveis de memória do ambiente python.
- **selenium:** Consiste em uma biblioteca *webdriver* que permite automatizar a navegação em páginas da web, simulando a operação humana no navegador Google-chrome. Este robô foi necessário para aquisição dos dados do INPE. Para o acesso às tabelas com as informações de interesse, uma sequência de cliques em botões e validações na página são requisitadas. Este processo é realizado sem a necessidade de uma interface gráfica de navegação.

3.5.2 Operação de Dados de Ciência Geológica

O trabalho utiliza dados de carácter geológico tanto na fase de calibração quanto operação. Neste contexto, o CPTEC/INPE fornece em seu servidor dados de precipitação, podendo ser no formato binário, GRIB, NetCDF ou HDF-SDS. Há uma necessidade de *softwares* específicos para manipular tais dados, os quais consistem em mapas e variáveis climáticas espacialmente distribuídas. A pesquisa utiliza a ferramenta **OpenGrads**, um programa gratuito formado por uma interface e um *firmware* de manipulação e visualização de dados de ciência geológica. A documentação e o *download* do *software* estão no site:

<http://opengrads.org/>

Para permitir a inserção destes dados no ambiente *Python*, foi utilizado a biblioteca **Py3Grads**, que estabelece uma conexão entre o *kernel* do Python e o *firmware* do *Grads*. Assim, dentro de um programa *Python* é possível acessar o *prompt* de comando do *Grads*, enviando comandos para manipular os dados necessários e importá-los para dentro do ambiente. Não há necessidade da interface gráfica do *Grads*, apenas o firmware é acessado. A biblioteca *Py3Grads* não é fornecida nos servidores do *Anaconda* e do *PIP3*, portanto para utilização é necessário acessar o link abaixo e instalá-lo manualmente:

<https://github.com/meridionaljet/py3grads>

3.5.3 Treinamento e Operação de Redes Neurais

O uso de *Deep Learning* é uma das principais abordagens do trabalho. Existem diversas bibliotecas para treinamento e manuseio de redes neurais para o ambiente *Python*. A ferramenta **Keras** foi escolhida por ser de fácil operação e altamente difundida e documentada. Devido à complexidade da arquitetura DL e a vastidão dos dados de treinamento, faz-se necessário o uso da versão **Keras-GPU**, que força o processamento ser realizado na placa de vídeo do computador. Se o treinamento fosse realizado na CPU da máquina, demandaria dezenas de vezes mais tempo, impossibilitando a pesquisa.

O processamento das redes neurais não é realizado no ambiente *Python*, o *Keras* age como operador do programa **TensorFlow**, uma plataforma completa de código aberto para Redes Neurais, executada paralela e externamente na máquina.

3.5.4 Dados Locais

Devido à grande extensão dos dados utilizados na pesquisa, torna-se inviável o acesso aos mesmos nos servidores em toda execução. Assim, os dados são baixados para o disco local da máquina. Para uma maior facilidade, após os dados dos modelos estarem condicionados para um UHE, eles são armazenados em disco local isoladamente, uma vez que os dados já estão recortados e preparados para utilização como entrada.

Para manipulação de dados locais, utilizou-se a biblioteca **joblib**, que salva, em formato “.pkl” no disco local, um conjunto de variáveis no ambiente *Python*. Após importar o arquivo de volta para o *Python*, as variáveis aparecem exatamente como foram salvas: com nome, formato e informações iguais, facilitando a manipulação das mesmas.

Assim, os dados de precipitação, evapotranspiração e vazão de cada UHE tornam-se acessíveis de forma instantânea. Quando há uma necessidade de atualização com dados recentes, basta abrir o arquivo “.pkl” com as variáveis, concatenar os dados novos, e salvar novamente. A biblioteca **joblib** foi amplamente utilizada na pesquisa, salvando resultados, dados de entrada, soluções dos processos de otimização, etc.

3.5.5 Aplicações Desenvolvidas

Esta sub-seção apresenta como foi implementada a metodologia proposta na plataforma *Python*, apresentando os principais programas desenvolvidos. O programa **MAIN** contém o controle centralizado de todos os módulos, chamando a execução dos demais e controlando a visualização dos resultados.

3.5.5.1 DataManager

O módulo **DataManager** realiza todo o condicionamento dos dados de entrada dos modelos, contendo:

- A verificação do banco de dados local, se há necessidade de novos downloads e preparação para a execução dos modelos no período desejado.
- A automação da obtenção dos dados dos servidores da ANA, CPTEC/INPE e INMET, através das bibliotecas *requests*, *selenium* e *wget*.
- O recorte e processamento dos dados geológicos através do *py3grads*.
- Os cálculos da Evapotranspiração e precipitação média.

3.5.5.2 SMAP

Neste Módulo estão implementados todos os métodos para calibração e operação do modelo SMAP, contendo:

- Método com o equacionamento do SMAP-ONS.
- Programa de otimização global do SMAP, que utiliza a biblioteca **Scipy** para execução do *Differential Evolution* sobre o modelo.
- Rotina para preparação e chamada da otimização local do SMAP.

3.5.5.3 BTO

Classe com a implementação genérica do *Bounded Twiddle Optimization*, contendo a normalização e a implementação do algoritmo.

3.5.5.4 Deep Learning

Programa com implementação das rotinas referentes às redes neurais, incluindo:

- Declaração da arquitetura de camadas
- Normalização de Dados
- Desnormalização de Dados
- Estratégia de treinamento, *cross-validation* e *reinforcement*.
- Abrir e salvar os modelos já com os hiper-parâmetros treinados em formato “.hdf5”

3.6 Resumo

Este capítulo apresentou toda o embasamento metodológico da Tese. (i) Os fundamentos teóricos sobre o SMAP e *Deep Learning*; (ii) As contribuições secundárias relacionadas às técnicas de aquisição e filtragem de dados meteorológicos; (iii) propostas de melhoria no modelo SMAP; (iv) a arquitetura do modelo híbrido final; (v) a arquitetura de *software* utilizada no trabalho.

4 RESULTADOS

Este Capítulo apresenta os resultados, análises e discussões sobre a metodologia proposta. Foram elaborados em uma máquina com processador Intel core i7-9750H, com 16 GB de memória RAM e placa de vídeo RTX 2060 com 6 GB. Vale a pena salientar que devido à complexidade computacional, tanto dos tópicos de otimização quanto do treinamento da rede neural, há grande demanda de memória e esforço computacionais. Sem uma placa de vídeo, fica inviável o treinamento dos modelos híbrido, apenas possível executá-lo operacionalmente.

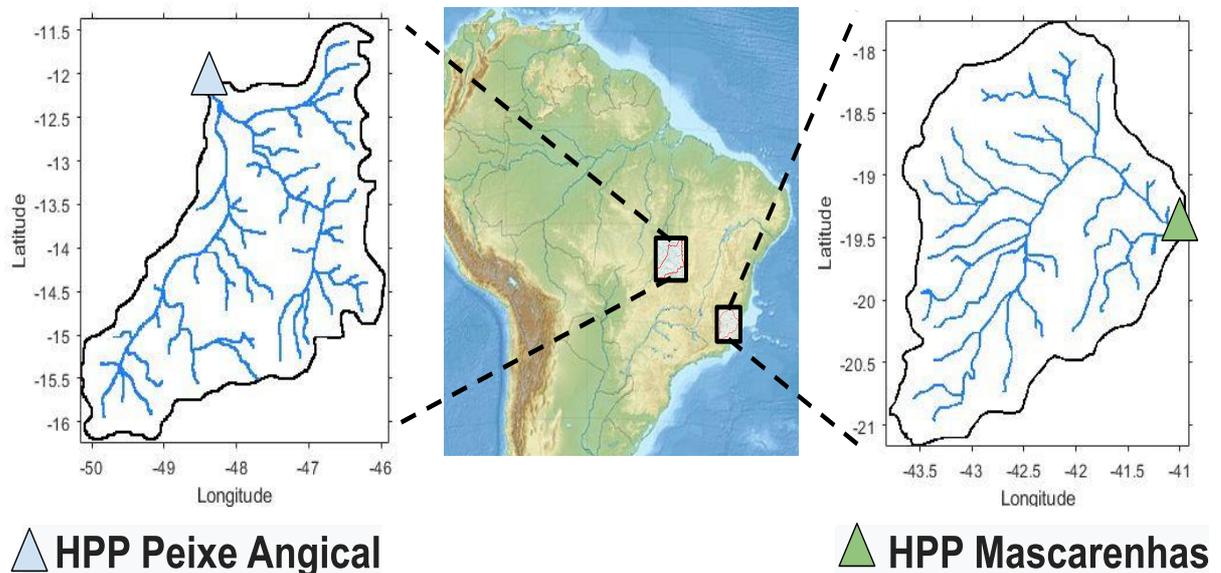
4.1 DESCRIÇÃO DOS ESTUDOS DE CASO

Para validar a arquitetura desenvolvida, foram consideradas duas bacias hidrográficas brasileiras nas quais estão instaladas as UHE Peixe Angical e UHE Mascarenhas. Estas usinas fazem parte do Grupo EDP Brasil, contratante do projeto P&D que originou a pesquisa. O objetivo é aplicar a metodologia de previsão de vazões diárias naturais sobre seus reservatórios. A UHE Mascarenhas está localizada no estado do Espírito Santo em uma região de clima tropical costeiro. Faz parte da bacia do Rio Doce, e sua área de drenagem possui aproximadamente 75.500 km^2 . Sua operação foi iniciada em 1974 e tem capacidade instalada de 198 MW. A UHE Peixe Angical está localizada no estado do Tocantins, em uma região predominantemente semi-árida. Faz parte da bacia do Tocantins - Araguaia, com área de drenagem de aproximadamente 121.500 km^2 . Sua operação foi iniciada em 2002 e possui capacidade instalada de 498,8 MW. A figura 26 mostra as áreas de drenagem de cada bacia e suas respectivas localizações. As vazões relacionadas à cada UHE são fornecidas pela Agência Nacional de Águas (ANA - *Agência Nacional de Águas*).

Cada *grid* de dados de precipitação fornecida pelo produto MERGE possui cerca de 10 km de lado. Assim, uma abundância de dados pluviométricos é identificada dentro de cada bacia, devido à extensão territorial da área de drenagem. São identificados 1.116 dados diários de precipitação para a região da UHE Mascarenhas e 2.420 para a UHE Peixe Angical.

O excesso de dados diários para cada região justifica a aplicação da camada Conv3D para tratamento e processamento, já que a mesma irá realizar um processamento de forma clusterizada em regiões da bacia. A definição das características de cada camada depende dos dados de cada bacia. Dessa forma, os parâmetros arquitetônicos para cada bacia foram obtidos empiricamente, apresentados na Tabela 6. Os valores abaixo podem ser um bom ponto de partida para implementação em novas usinas, apenas adaptando à área da bacia: as dimensões da entrada, do *kernel* e do *stride* devem ser associadas de modo que, no processo de convolução, nenhuma linha ou coluna de dados seja ignorada.

Figura 26 – Localização das regiões de estudo e áreas de drenagem das bacias hidrográficas: UHE Peixe Angical e UHE Mascarenhas.



Fonte: Elaborada pelo autor.

Tabela 6 – Arquitetura de cada bacia

	Conv3D					LSTM		Saída	
	Entrada	Kernel	Filters	Strides	Ativação	Units	Ativação	Neurons	Ativação
Mascarenhas	(8, 36, 31)	(2, 12, 11)	1	(1, 12, 10)	Relu	12	Tanh	7	Sigmoid
Peixe	(8, 50, 44)	(2, 18, 16)	1	(1, 16, 14)	Relu	12	Tanh	7	Sigmoid

4.2 CALIBRAÇÃO GLOBAL DO MODELO SMAP

O primeiro passo na aplicação da arquitetura desenvolvida é a calibração do modelo SMAP de forma Global, considerando um longo histórico de vazões, precipitações e evapotranspirações. O objetivo desta etapa é a definição dos parâmetros que modelam a bacia hidrográfica.

Foi considerado para esta etapa o período de 09/2012 a 11/2016. A tabela 7 mostra os limites mínimo e máximo de cada parâmetro considerado na calibração e os valores obtidos nesta etapa para cada bacia hidrográfica.

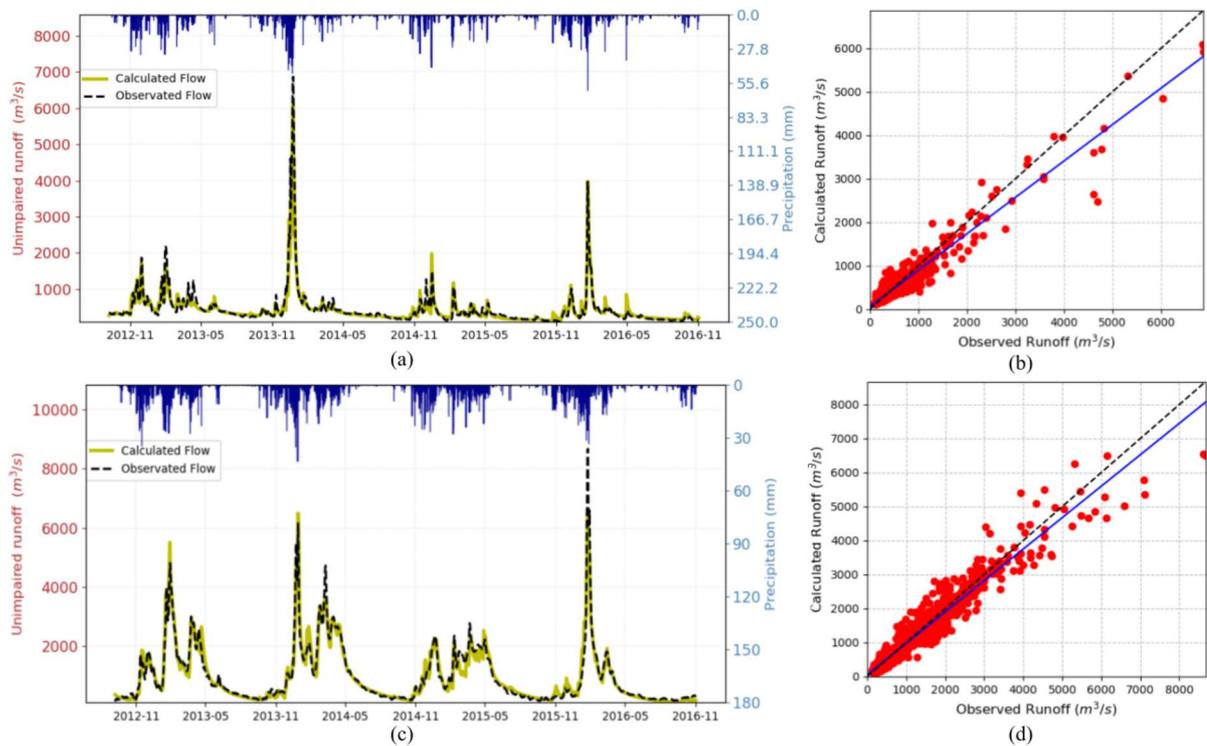
A Figura 27 mostra o comportamento da vazão calibrada e observada temporalmente, além dos gráficos de dispersão dos dados para cada bacia hidrográfica. As métricas estatísticas dos dados para cada região são apresentadas na Tabela 8.

Os resultados demonstram que o modelo SMAP é uma ferramenta com excelente capacidade de modelar o processo chuva-vazão. Com os parâmetros globais encontrados, entra-se em uma fase *online* da metodologia.

Tabela 7 – Resultados dos Parâmetros Globais do SMAP.

Parâmetro	UHE Mascarenhas	UHE Peixe Angical	Limite Inferior	Limite Superior
$Eb(0)$	290.1	361.1	0	1000
$Tu(0)$	50.7	11.31	0	100
$Sup(0)$	236.1	61.3	0	1000
$k1t$	5.32	2.72	0.1	10
$k2t$	5.01	9.37	0.1	10
$k3t$	29.3	10.6	0.1	60
$k2t2$	27.7	15.1	0.1	30
kkt	181.5	88.7	0.1	250
$Capc$	46	21.1	20	60
Ai	0.01	0.01	0.01	10
Str	157.7	234.5	0	300
$Crec$	8.56	0.29	0.1	100
$Ecof$	0.74	0.701	0.7	1.3
$Ecof2$	1.02	1.23	0.7	1.3
$H1$	88.2	53.61	0	200
H	3.4	13.9	0	200
$Pcof$	0.816	0.991	0.7	1.3

Figura 27 – Comparação entre o escoamento observado e calculado pelo modelo SMAP na etapa de calibração. (a) and (b) UHE Mascarenhas; (c) and (d) UHE Peixe Angical.



Fonte: Elaborada pelo autor.

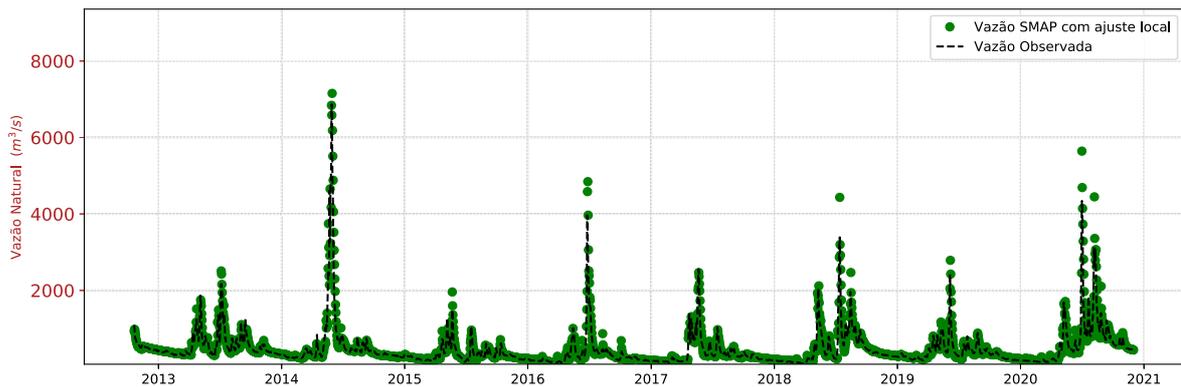
Tabela 8 – Desempenho calibração global do SMAP.

UHE	MAPE	MAE	RMSE	NSE	r
Mascarenhas	15.8	69.7	150.6	0.898	0.965
Peixe Angical	17.3	167.1	247.4	0.935	0.966

4.3 OTIMIZAÇÃO LOCAL DO SMAP

Nesta seção são apresentados os resultados do ajuste local do SMAP, sendo que para cada dia do banco de dados de treinamento da RNA, houve uma calibração local pelo *Twiddle* envolvendo os 14 dias precedentes. As Figuras 28 e 29 apresentam as vazões encontradas pelo SMAP comparadas com as vazões observadas, cada ponto verde é o dia final do processo de otimização, lembrando que as variáveis de estado que forneceram tais vazões (valores dos tanques) serão utilizadas como entrada do modelo híbrido.

Figura 28 – Ajuste Local - UHE Mascarenhas



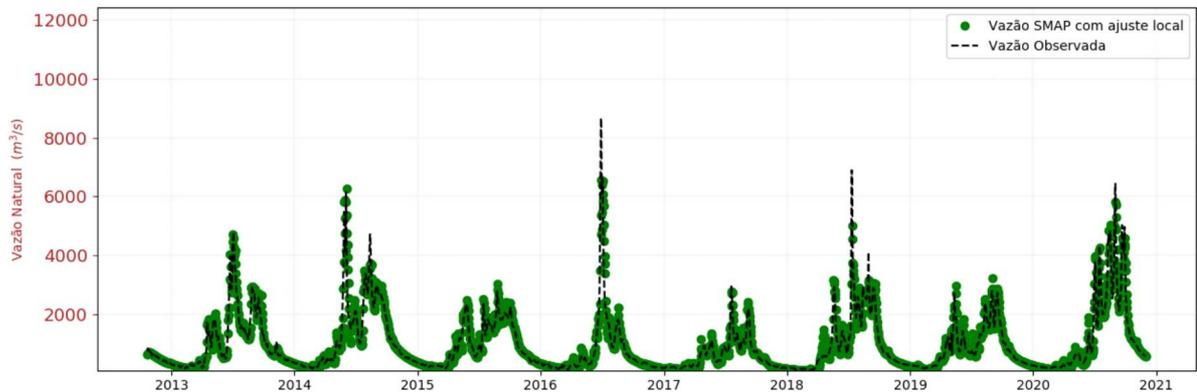
Fonte: Elaborada pelo autor.

Tabela 9 – Desempenho da calibração local do SMAP.

UHE	MAPE	MAE
Mascarenhas	12.1	59.1
Peixe Angical	13.9	127.1

Como pode ser observado, na Tabela 9 ao se permitirem pequenas correções no modelo SMAP de forma *online*, há um aumento de desempenho em relação ao modelo com parâmetros globais. Tais resultados poderiam ser mais acurados com um maior relaxamento das restrições, mas a perda de padrão poderia comprometer a RNA posteriormente. Os dados que alimentam o sistema *Deep Learning*, que geraram as vazões da Figura 28 e 29, estão expostos nas Figuras 30 e 31 respectivamente.

Figura 29 – Ajuste Local - UHE Peixe Angical.



Fonte: Elaborada pelo autor.

4.4 SMAP-DNN

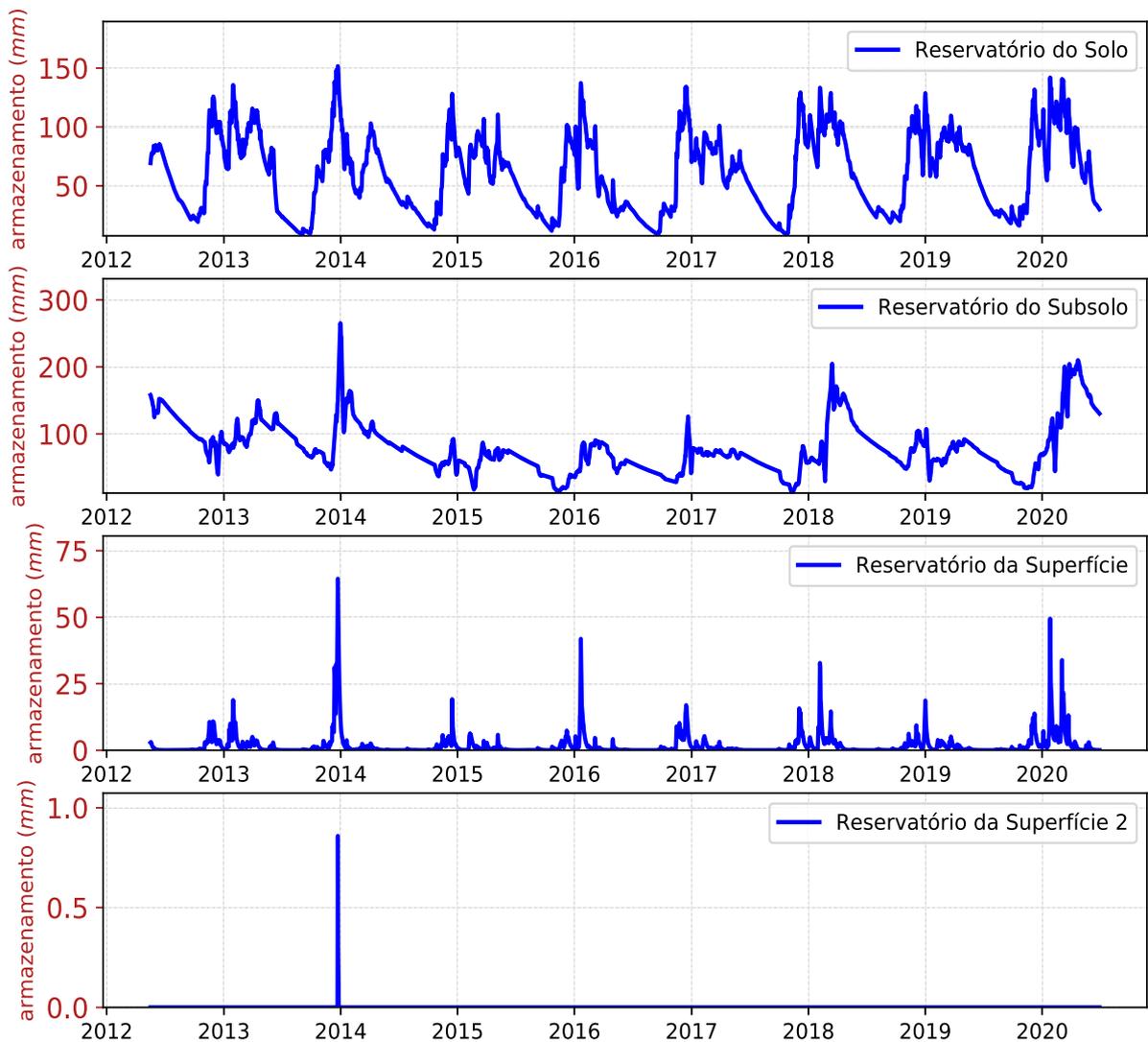
Inicialmente, para gerar o banco de dados de treinamento da DNN, todo o histórico foi calibrado pelo algoritmo Twiddle para obter os estados sub-ótimos de reservatório em cada momento t . Esses dados são referentes ao estado dos reservatórios no período de operação do sistema.

Assim, para validar e simular uma operação do modelo, utilizou-se o *cross-validation*. A previsão foi realizada em intervalos de 7 dias, começando em 11/2016. Assim, na primeira execução, a previsão é feita do tempo t ao tempo $t + 7$. Na próxima rodada, a previsão é efetuada para o tempo $t + 8$ a $t + 15$ e, assim, sucessivamente até 05/2020. Portanto, a cada nova execução, o conjunto de treinamento aumenta em 7 amostras, e os conjuntos de validação e teste são deslocados. Neste teste, utilizou-se a chuva observada do MERGE, sendo uma aproximação do que aconteceu. Em operação real, é necessário o uso de dados de previsão, que trazem uma maior estocasticidade. Assim, os dados MERGE são utilizados nesta etapa de validação e cumprem o papel de precipitação prevista neste período, visto que a validação deve ser feita em um período em que as vazões observadas estejam disponíveis para comparação com as saídas da DNN.

Para validar a metodologia, as vazões calculadas serão comparadas com a vazão observada para o mesmo período. Além disso, a arquitetura proposta (SMAP-DNN) será comparada com duas outras estratégias de previsão:

- O modelo SMAP, após o processo de calibração local, efetuará uma previsão baseada nas precipitações e evapotranspirações futuras, com os parâmetros encontrados neste processo.
- Outro modelo DNN, com a mesma estratégia CONV3D-LSTM. Mas, não receberá as

Figura 30 – Tanques Ajustados - UHE Mascarenhas.



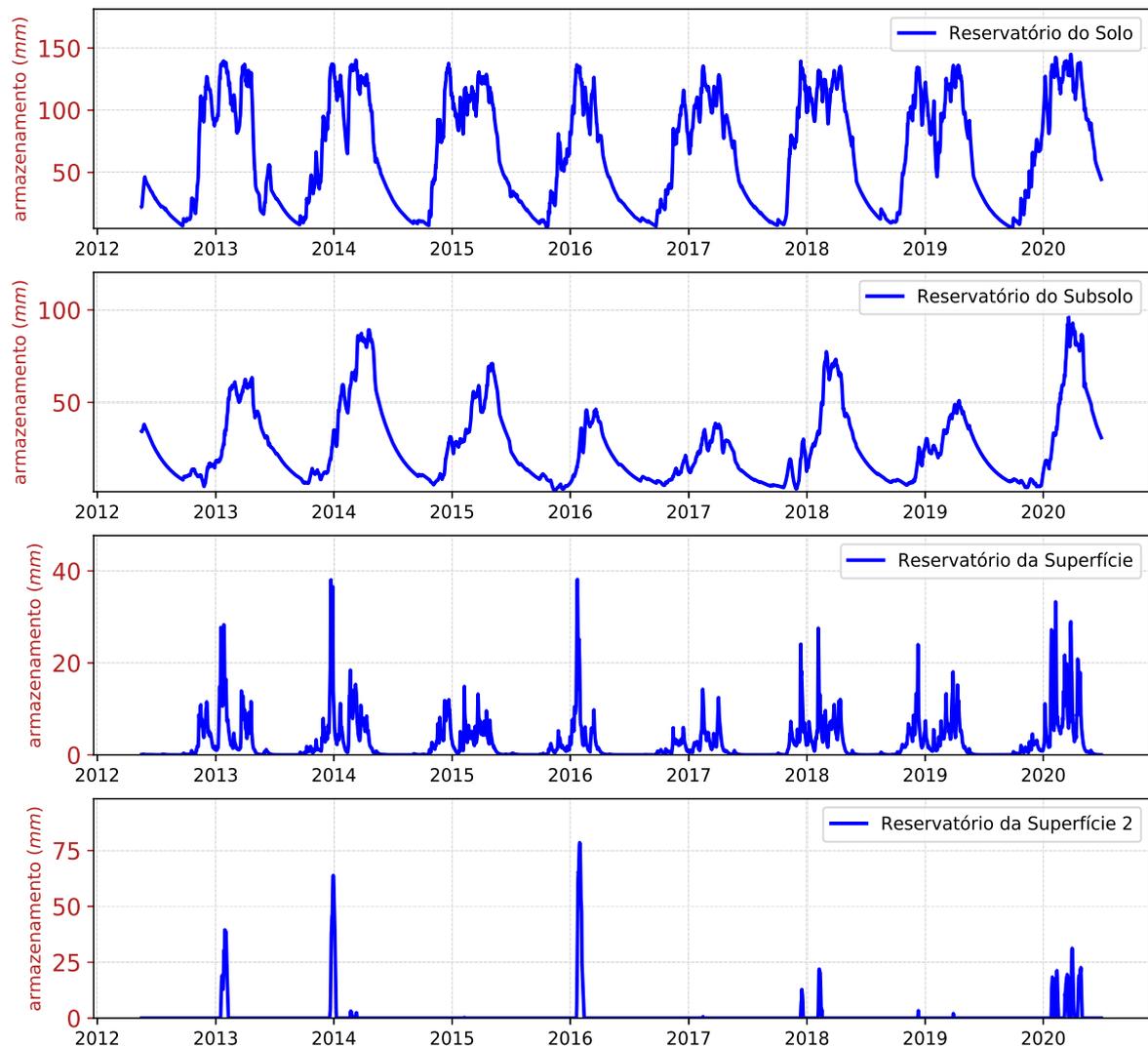
Fonte: Elaborada pelo autor.

entradas do modelo SMAP. Portanto, as entradas são apenas os *grids* de precipitação e última vazão observada. Esta versão terá mesmo otimizador, camadas e métrica que a versão híbrida.

Devido à natureza estocástica do otimizador ADAM e do processo de inicialização de pesos, estratégias envolvendo RNAs foram realizadas 10 vezes cada. Os algoritmos *Twiddle* e SMAP possuem uma característica determinística, portanto, o modelo SMAP foi operado apenas uma vez a cada intervalo de previsão. A distribuição dos resultados obtidos por cada técnica é mostrada na Figura 32.

Os resultados obtidos demonstram que os modelos DNN possuem melhor capacidade de previsão do que o modelo SMAP. Além disso, os gráficos mostram que usar as variáveis

Figura 31 – Tanques Ajustados - UHE Peixe Angical



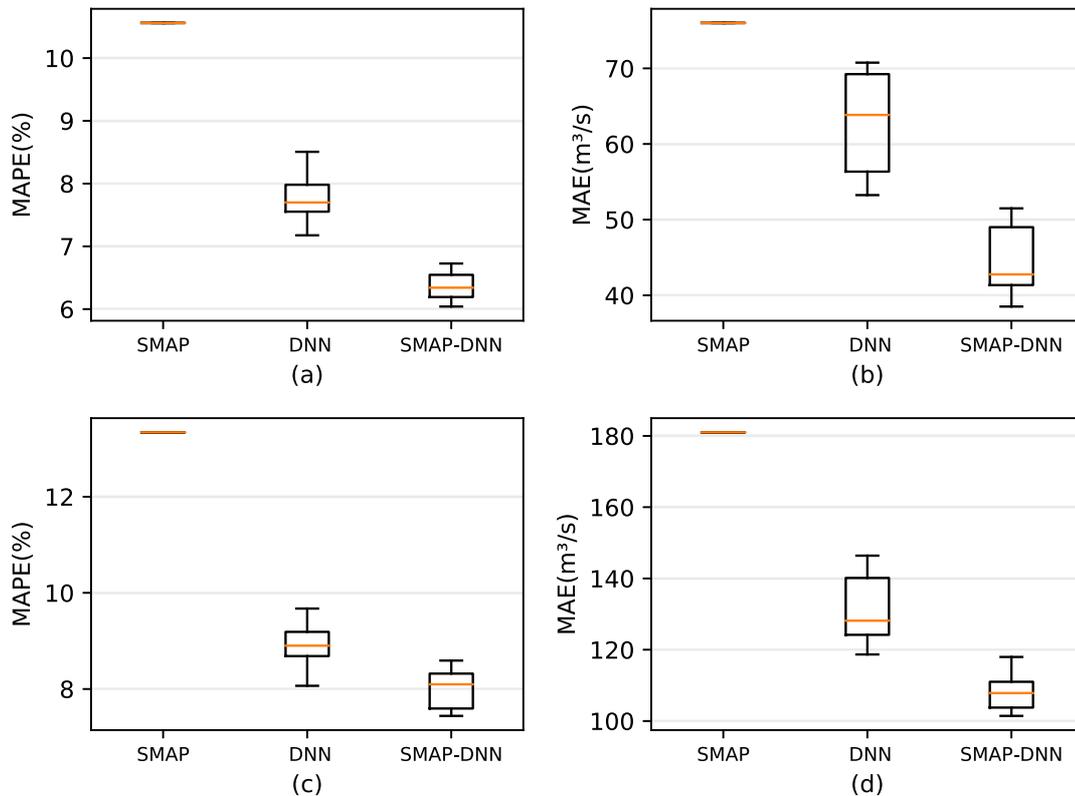
Fonte: Elaborada pelo autor.

de estado do SMAP como dados de entrada para o DNN, melhora a capacidade de previsão. As variáveis de estado intermediárias do SMAP complementam a vazão observada com as características dinâmicas da bacia no momento da previsão.

Os resultados da UHE Peixe Angical mostram que a aplicação da estratégia SMAP-DNN reduz os valores MAPE e MAE. Porém, os resultados do UHE Mascarenhas apresentaram reduções mais expressivas. Essa diferença entre os resultados das duas regiões pode ser explicada pelos seguintes aspectos:

- O modelo SMAP pode representar melhor a região da UHE Mascarenhas, devido às características intrínsecas da bacia hidrográfica;
- Dados climáticos para uma região podem ser mais precisos; principalmente por

Figura 32 – Comparação entre SMAP, DNN and DNN-SMAP. (a) e (b) UHE Mascarenhas; (c) e (d) UHE Peixe Angical.



Fonte: Elaborada pelo autor.

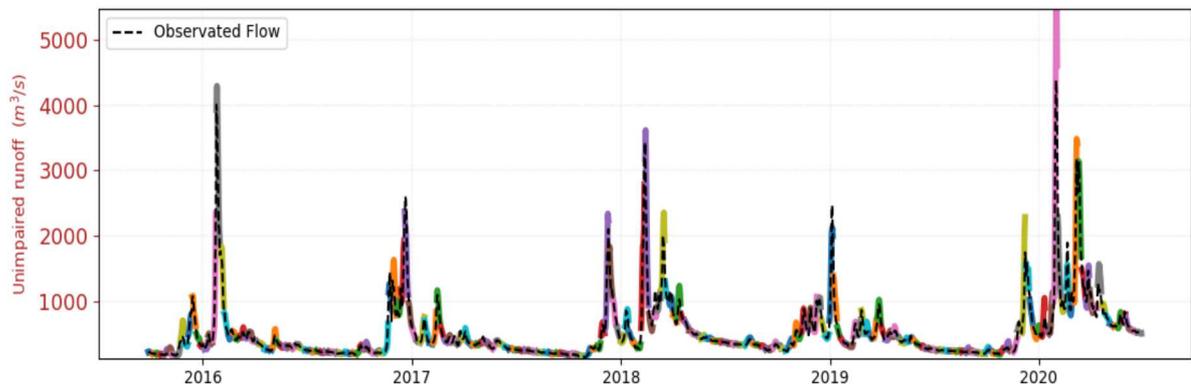
Mascarenhas estar no Sudeste, em uma região mais desenvolvida e povoada, coberta por mais estações meteorológicas.

- Características de uma bacia hidrográfica podem dificultar a obtenção de pontos ideais por meio do processo de calibração;

Apesar da diferença entre os resultados das duas bacias, ambos demonstram que a arquitetura proposta é eficaz para o processo de previsão de vazão no curto prazo. Os dados de vazão obtidos pela arquitetura proposta são apresentados na Figura 33 e 34 para a UHE Mascarenhas e UHE Peixe Angical, respectivamente. A linha pontilhada preta representa a vazão observada e as linhas coloridas representam os intervalos de 7 dias gerados em sequência nesta etapa de previsão. Os resultados mostram haver grande assertividade nos períodos de estiagem e boa previsão da tendência nos períodos de cheias. Este comportamento também pode ser observado nos gráficos de dispersão da Figura 35, e em poucos momentos houve erros de tendência.

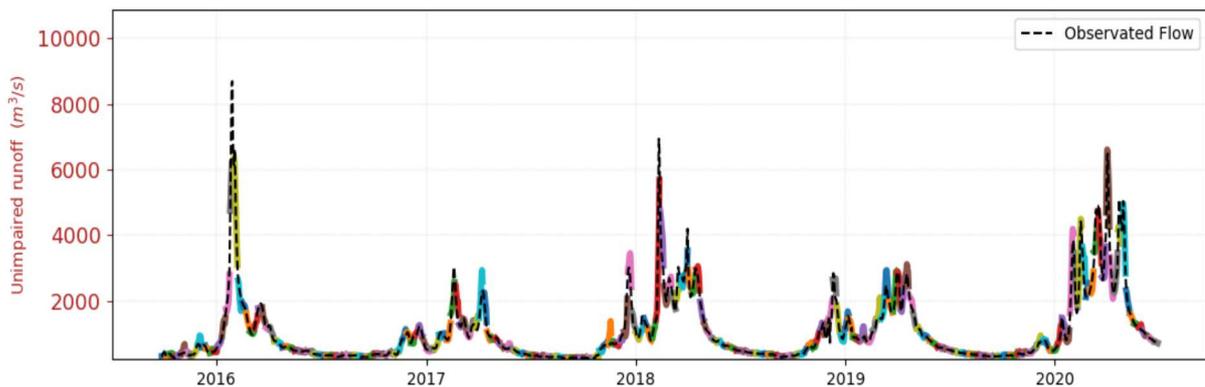
Análises estatísticas também foram aplicadas para avaliar a melhor previsão obtida para cada região. O comportamento desta previsão foi analisado segundo o número de dias

Figura 33 – Simulação de Previsão de Escoamento para UHE Mascarenhas.



Fonte: Elaborada pelo autor.

Figura 34 – Simulação de Previsão de Escoamento para UHE Peixe Angical.

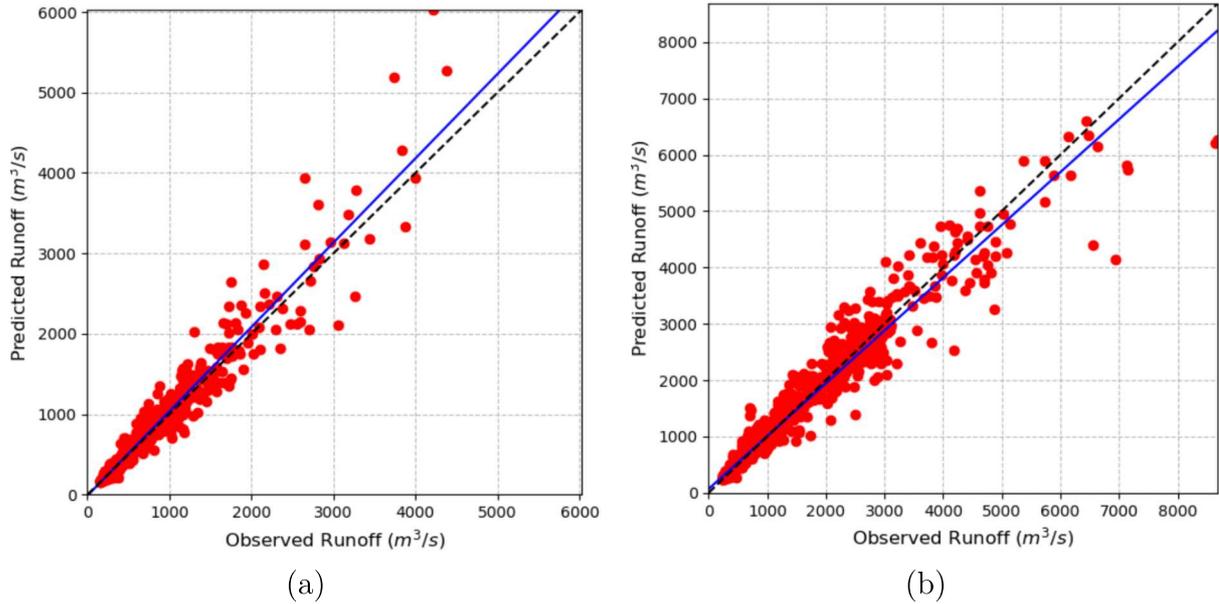


Fonte: Elaborada pelo autor.

à frente, conforme mostrado na Tabela 10. Os resultados mostram que a correlação dos dados da UHE Mascarenhas é mais forte do que a dos dados da UHE Peixe Angical. Além disso, todas as métricas mostram que a previsão de Mascarenhas teve melhor desempenho. No entanto, os resultados obtidos para ambas as bacias demonstram que a arquitetura desenvolvida é uma excelente ferramenta para a previsão de vazões afluentes. A tabela 10 também mostra que as previsões para o primeiro dia têm um erro muito baixo e alta correlação entre os dados de fluxo observados e calculados. Como esperado, o desempenho tende a diminuir à medida que se distancia do momento da execução. Mesmo assim, estes últimos também apresentam forte correlação entre os dados.

A partir das análises realizadas, pode-se observar que as previsões para os primeiros dias são mais acuradas, portanto, podem ser úteis em programas de otimização de despacho e alocação de geradores.

Figura 35 – Gráfico de dispersão de dados de escoamento observados e previstos. (a) UHE Mascarenhas; (b) UHE Peixe Angical.



Fonte: Elaborada pelo autor.

Tabela 10 – Resultados Estatísticos das Previsões.

UHE Mascarenhas								
Métrica	Total	Dia1	Dia2	Dia3	Dia4	Dia5	Dia6	Dia7
MAPE	6.047	1.94	4.193	5.527	6.345	7.321	8.228	8.778
MAE	38.51	14.06	32.25	39.63	39.78	41.85	51.90	50.12
RMSE	85.24	32.41	90.03	94.65	85.71	75.1	116.61	102.19
r	0.978	0.997	0.986	0.979	0.977	0.982	0.957	0.969
NSE	0.952	0.995	0.958	0.945	0.951	0.96	0.916	0.939

UHE Peixe Angical								
Métrica	Total	Dia1	Dia2	Dia3	Dia4	Dia5	Dia6	Dia7
MAPE	7.408	1.594	3.708	6.191	7.948	9.596	10.991	11.829
MAE	92.02	23.12	41.11	73.32	101.02	122.55	141.06	141.82
RMSE	210.27	65.45	86.44	168.43	247.62	285.12	319.24	299.59
r	0.97	0.997	0.996	0.983	0.967	0.957	0.945	0.946
NSE	0.94	0.994	0.991	0.966	0.932	0.912	0.889	0.893

4.4.1 Análise de tendência

O método foi elaborado para fornecer 7 dias de previsões. Neste horizonte, é possível identificar que as tendências de variações futuras foram bem acompanhadas. Um reconhecimento prévio das tendências das variações semanais da vazão contém informações importantes para a planta. O processo de planejar a manutenção e a operação de turbinas

e vertedouros com antecedência, otimiza a dinâmica de toda a usina. Um estudo de reconhecimento de tendências foi realizado a partir dos resultados anteriores, vide Figuras 33 e 34, estando representado nas matrizes confusão nas Tabelas 11 e 12. Nesta análise, avaliando a assertividade do modelo, foram classificados os eventos de cada semana de previsão em: (i) grande variação positiva; (ii) grande variação negativa; e (iii) pequena variação.

Para classificar os eventos, realizaram-se regressões lineares tanto da série observada, quanto dos resultados da RNA, a cada sete amostras. As retas partem da última vazão observada (dia realizada a previsão), vide Figura 36. Analisando e comparando os coeficientes angulares dessas retas, elaboraram-se as matrizes confusão. O critério de classificação foi:

- Variação negativa: quando o coeficiente da reta é menor que o negativo da metade do desvio padrão de todas as séries semanais de vazão observada.
- Pequena variação: quando o módulo o coeficiente da reta é menor que a metade do desvio padrão de todas as séries semanais de vazão observada.
- Variação positiva: quando o coeficiente da reta é maior que a metade do desvio padrão de todas as séries semanais de vazão observada.

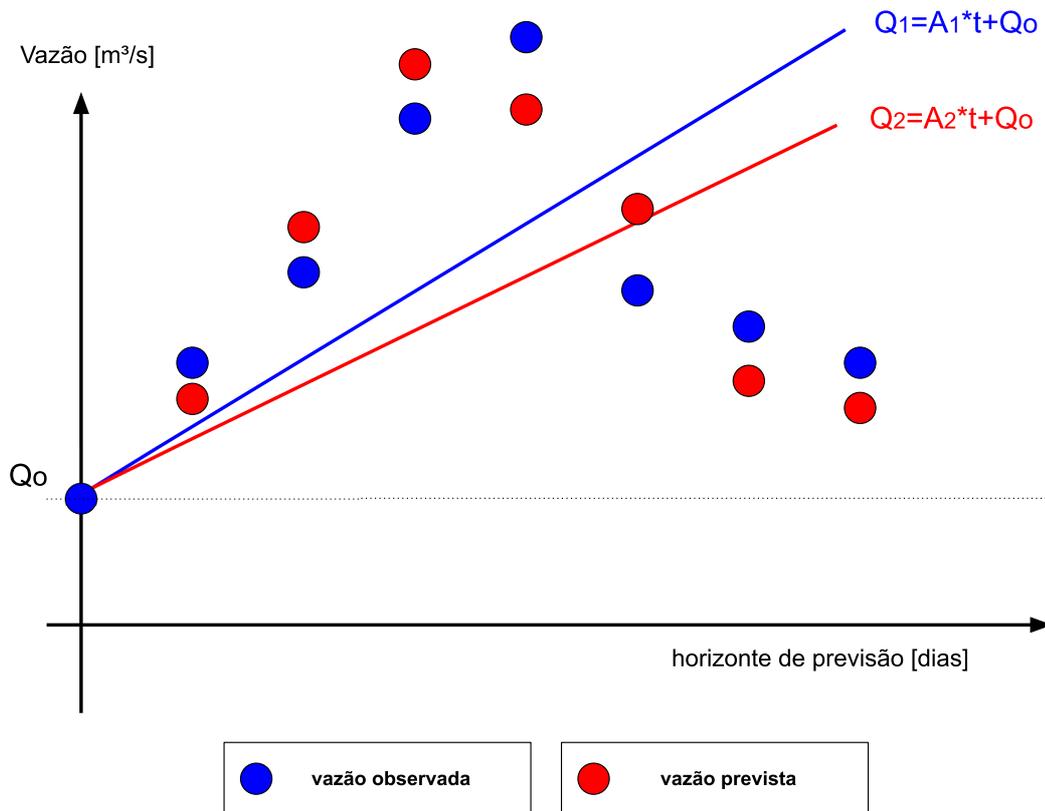
Quando as linearizações da vazão observada e da prevista possuem coeficientes A_1 e A_2 próximos, significa que os volumes totais de água foram semelhantes, e que houve uma boa assimilação da tendência da vazão. As Figuras 37 e 38 exibem o comportamento destas tendências de variação da vazão observada para as UHEs Mascarenhas e Peixe respectivamente.

Desta maneira, as Tabelas 11 e 12 foram elaboradas comparando as tendências observadas e previstas pelo modelo híbrido, por matrizes confusão. Através do sistema de classificação proposto, os resultados apresentaram uma ótima acurácia, com poucos falsos positivos. Houve 97% de assertividade na UHE Mascarenhas e 92% em Peixe Angical, sendo que não houve erros de classificação de “variação positiva” como “variação negativa” ou vice-versa.

Tabela 11 – Matriz Confusão UHE Mascarenhas

$Q_{real} \setminus Q_{calc}$	Variação Negativa	Variação Pequena	Variação Positiva
Variação Negativa	20	2	0
Variação Pequena	2	204	0
Variação Positiva	0	3	17

Figura 36 – Linearização das curvas de vazão



Fonte: Elaborada pelo autor.

Tabela 12 – Matriz Confusão UHE Peixe

$Q_{real} \setminus Q_{calc}$	Variação Negativa	Variação Pequena	Variação Positiva
Variação Negativa	25	3	0
Variação Pequena	4	183	8
Variação Positiva	0	4	21

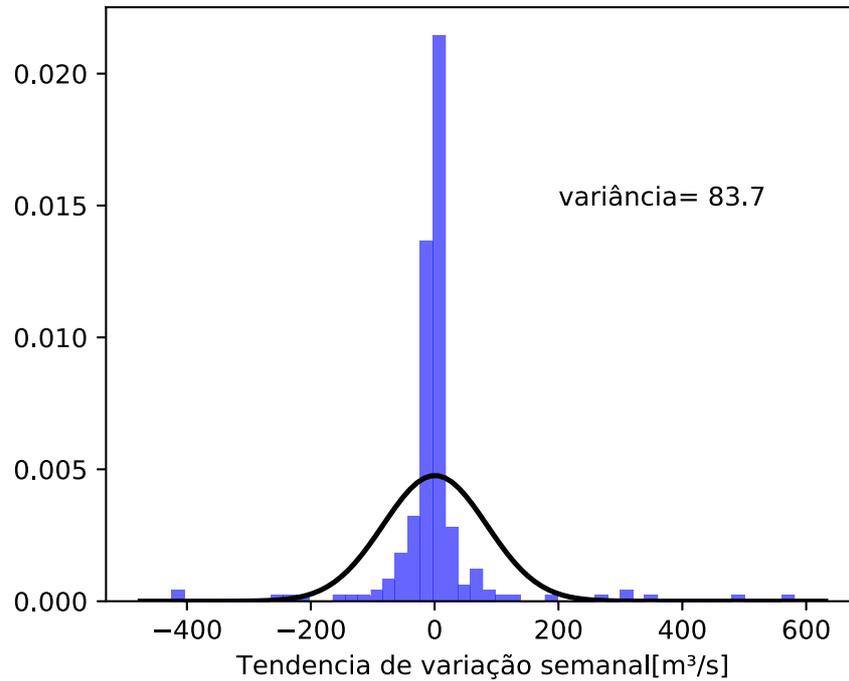
4.4.2 Análise estocástica da Precipitação

As simulações de operação do modelo apresentadas anteriormente, foram realizadas com o produto MERGE, baseado em chuva observada pré-processada. Em uma operação real, os erros inerentes às medições pluviométricas são superiores, podendo reduzir o desempenho dos resultados. Para avaliar a robustez do modelo perante a estocasticidade das previsões de chuva, os resultados do *cross-validation* foram realizados novamente com as seguintes mudanças no conjunto de teste:

Serão realizados testes em 3 amostras em cada rodada do *cross-validation*:

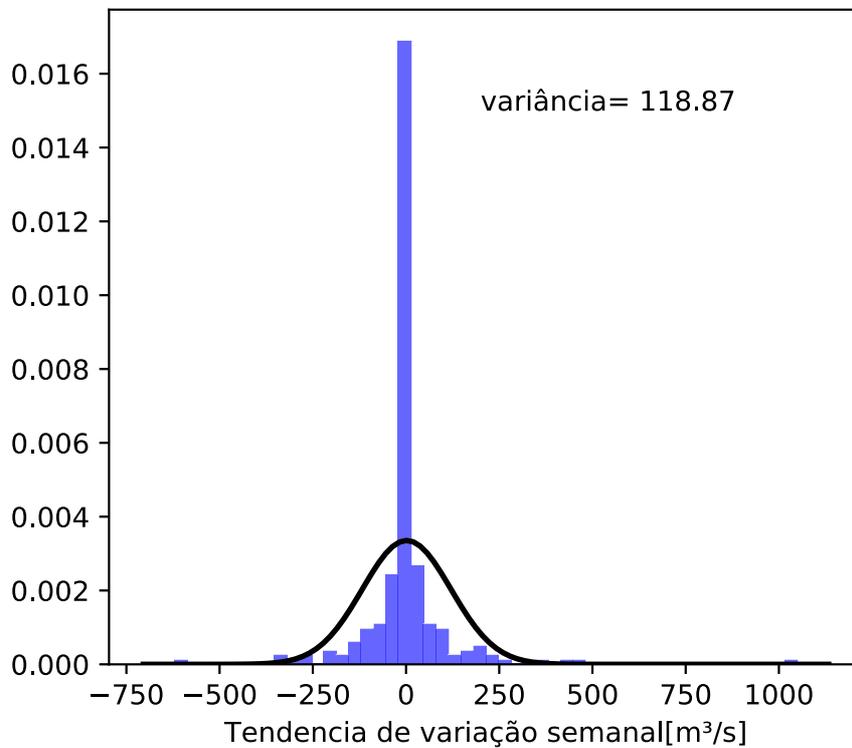
1. Uma amostra referente a previsão com a chuva futura do produto MERGE, assim como realizado na seção anterior.

Figura 37 – Histograma normalizado das tendências de variação da vazão observada na UHE Mascarenhas



Fonte: Elaborada pelo autor.

Figura 38 – Histograma normalizado das tendências de variação da vazão observada na UHE Peixe Angical.



Fonte: Elaborada pelo autor.

2. Uma amostra adicionando um erro escalar de 20% sobre todas as quadrículas do MERGE para todo o horizonte de previsão, hiper-estimando as precipitações.
3. Uma amostra subtraindo um erro escalar de 20% sobre todas as quadrículas do MERGE para todo o horizonte de previsão, sub-estimando as precipitações.

O objetivo desta análise é avaliar a robustez da rede neural treinada, se variações da chuva poderiam causar “*outliers*” ou anomalias no resultado. As Figuras 39 e 40 apresentam os resultados gráficos para UHE Mascarenhas e UHE Peixe Angical respectivamente.

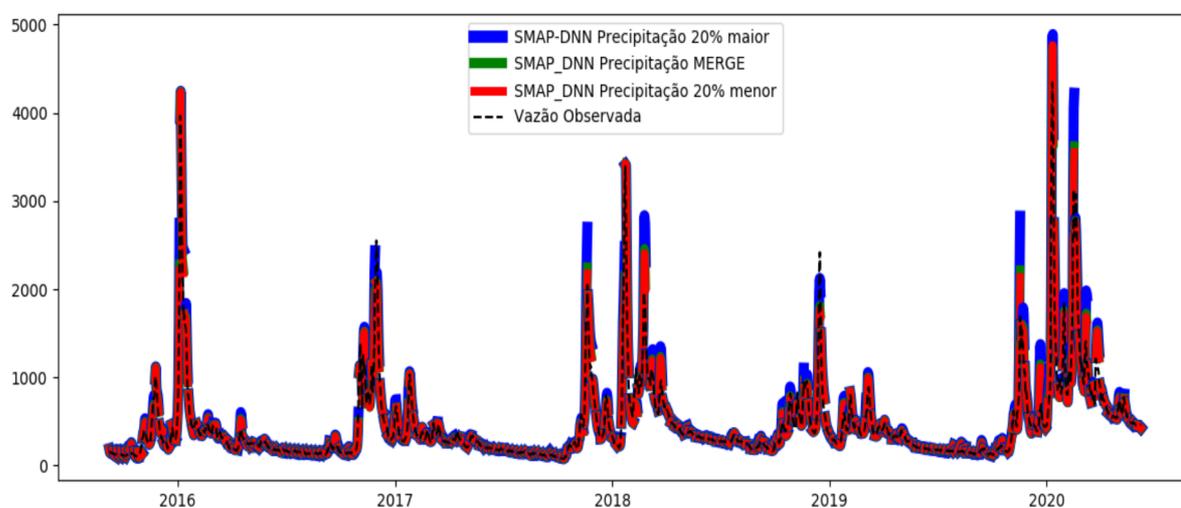


Figura 39 – Simulação de Previsão Estocástica para UHE Mascarenhas.

Fonte: Elaborada pelo autor.

Conforme as curvas apresentadas nas Figuras 39 e 40, não houve anomalias perceptíveis ao variar o volume total de chuva em mais ou menos 20%. Nos períodos de seca, como era de se esperar, as curvas ficam sobrepostas. E no período de grandes tempestades, as curvas com chuva super-estimada (em azul) ficam ligeiramente superiores. Estes resultados mostram que o treinamento é robusto, com reconhecimento de padrão generalista e com “*overfitting*” reduzido. O modelo híbrido conseguiu correlacionar de forma satisfatória a distribuição espaço-temporal da precipitação com o atual estado da bacia, estimando a vazão futura.

Para avaliar quantitativamente tais resultados, foi elaborada novamente a matriz confusão da tendência dos resultados, assim como efetuado na seção anterior. Assim, pode-se visualizar o quanto os erros sistemáticos simulados alteraram a detecção de tendência do modelo. Os resultados estão expostos nas Tabelas 13 a 18 .

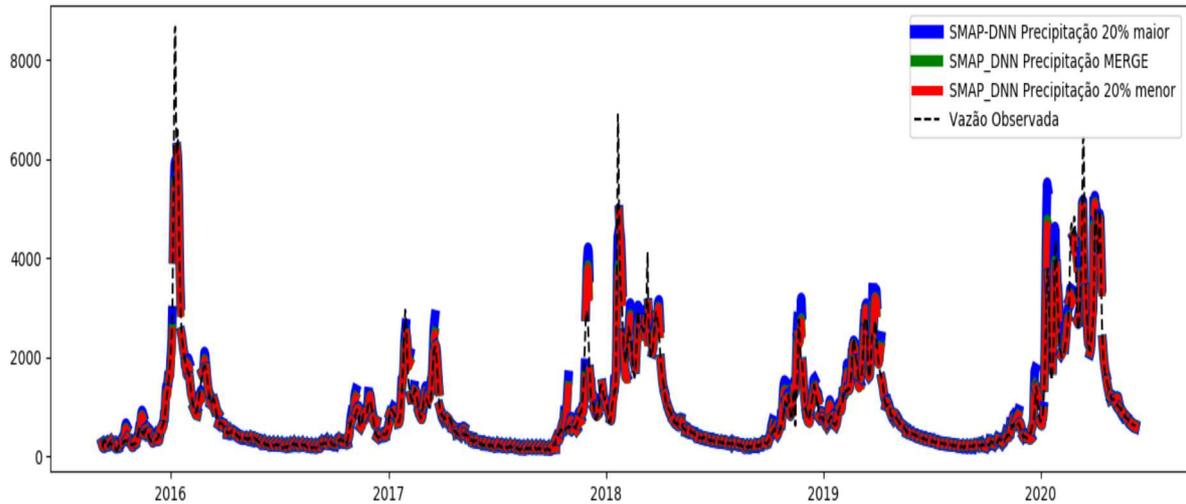


Figura 40 – Simulação de Previsão Estocástica para UHE Peixe Angical.
Fonte: Elaborada pelo autor.

Tabela 13 – Matriz Confusão UHE Mascarenhas - Chuva sem adição de erros

Q _{real} \ Q _{calc}	Variação Negativa	Variação Pequena	Variação Positiva
Variação Negativa	18	5	0
Variação Pequena	3	202	3
Variação Positiva	0	2	15

Tabela 14 – Matriz Confusão UHE Mascarenhas - Chuva subestimada.

Q _{real} \ Q _{calc}	Variação Negativa	Variação Pequena	Variação Positiva
Variação Negativa	19	4	0
Variação Pequena	4	201	3
Variação Positiva	0	3	14

Tabela 15 – Matriz Confusão UHE Mascarenhas - Chuva superestimada

Q _{real} \ Q _{calc}	Variação Negativa	Variação Pequena	Variação Positiva
Variação Negativa	17	6	0
Variação Pequena	0	200	8
Variação Positiva	0	0	17

Tabela 16 – Matriz Confusão UHE Peixe - Chuva sem adição de erros

Q _{real} \ Q _{calc}	Varição Negativa	Varição Pequena	Varição Positiva
Varição Negativa	24	6	0
Varição Pequena	3	181	6
Varição Positiva	0	4	24

Tabela 17 – Matriz Confusão UHE Peixe - Chuva subestimada

Q _{real} \ Q _{calc}	Varição Negativa	Varição Pequena	Varição Positiva
Varição Negativa	24	6	0
Varição Pequena	3	181	6
Varição Positiva	0	5	23

Tabela 18 – Matriz Confusão UHE Peixe - Chuva superestimada

Q _{real} \ Q _{calc}	Varição Negativa	Varição Pequena	Varição Positiva
Varição Negativa	21	9	0
Varição Pequena	2	176	12
Varição Positiva	0	4	24

Os erros sistemáticos de previsão simulados, pouco alteraram a distribuição das matrizes confusão, demonstrando que o modelo apresentou robustez na assertividade de tendências. Vale a pena salientar que ao simular erros aleatórios nas quadrículas de chuva, a saída do método praticamente não se altera. Com uma média do erro igual a zero, e devido ao grande número de quadrículas, os erros positivos e negativos se compensam tornando a volume médio próximo do constante. Assim, apenas erros de escala foram simulados.

4.5 Resumo

Este capítulo apresentou os resultados que validam as propostas da Tese: (i) O processo de obtenção dos parâmetros globais de cada bacia estudada; (ii) o resultado referente ao ajuste global, demonstrando o potencial de ajuste das vazões de saída para um cenário local mais fidedigno; (iii) resultados de validação cruzada, simulando o modelo operado durante dois anos semanalmente; (iv) análise de assertividade na tendência das variações de vazão; (v) análise estocástica simulando erros de precipitação.

5 CONCLUSÕES

A previsão da disponibilidade de recursos hídricos é um estudo importante para diversos setores, incluindo o elétrico. Uma boa estimativa de vazões permite planejar a operação das UHEs do sistema elétrico interligado. Esse planejamento inclui a previsão de geração e vertimento a curto e longo prazo. Este trabalho propôs uma ferramenta inovadora para melhorar a assertividade das previsões de vazão de água com 7 dias de antecedência em uma bacia hidrográfica.

A metodologia acoplou o modelo matemático de chuva-vazão SMAP e uma arquitetura de Deep Learning Conv3D - LSTM. O algoritmo Twiddle foi aplicado para um ajuste local suave do modelo SMAP. Foram propostas alterações no algoritmo para garantir um processo gradual e respeitando as restrições. Assim, os dados do reservatório foram disponibilizados para treinamento de arquitetura de forma mais precisa.

A camada convolucional 3D (Conv3D) mostrou grande potencial para a obtenção de características espaço-temporais da chuva. A ferramenta desenvolvida foi testada para reservatórios de duas grandes usinas em bacias brasileiras com características divergentes. Os resultados mostraram que a união dos dois modelos apresentou melhor assertividade em relação à aplicação dos métodos desacoplados. A arquitetura proposta foi construída visando definir apenas um cenário para cada execução e este é o foco deste estudo. Devido à natureza aleatória do processo de treinamento do DNN, análises estocásticas podem ser desenvolvidas, como o uso de Regressão de Quantis no estágio de treinamento.

A comparação entre as vazões observadas e previstas pelo modelo demonstrou a existência de forte correlação, validando o estágio de previsão. Além disso, análises estatísticas foram aplicadas às previsões para cada dia da semana. Os resultados mostraram que a assertividade da previsão é maior no início da previsão. Porém, os últimos dias têm apresentado resultados aceitáveis para o planejamento e conforme os limites ideais das métricas estatísticas consideradas.

Visando uma informação qualitativa, foi avaliada a precisão da previsão de tendência da vazão. Os resultados mostraram que os modelos possuem grande assertividade ao classificar a tendência de variação das descargas nos reservatórios das UHEs, o que pode ser útil para planejamento da operação de turbinas, vertedouros e manutenções. Para avaliar a robustez do modelo perante os possíveis erros de previsão de chuva, foram simulados cenários de subestimação e superestimação pluviométrica. Concluiu-se que tais erros não geraram erros anômalos, e impactaram de forma desprezível as tendências de saída do modelo.

Para as simulações apresentadas, foram utilizados dados de precipitação observados em formato de grade fornecidos pelo produto MERGE / CPTEC. Para a aplicação da rede como ferramenta de previsão, são necessários dados de previsão meteorológica. A aplicação

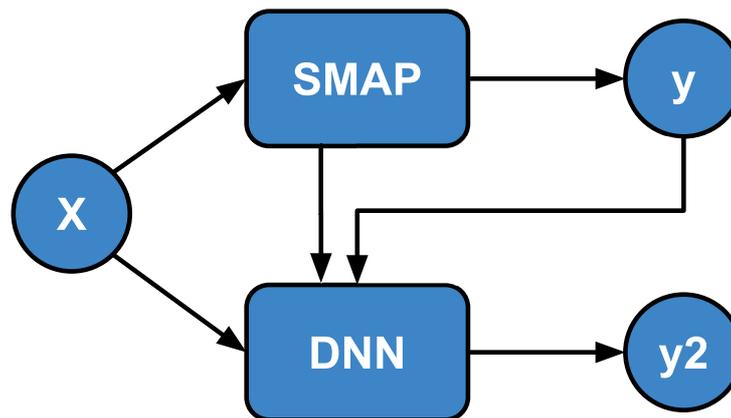
dos dados MERGE representa uma vantagem desse ponto de vista, visto que os dados da previsão possuem este mesmo formato. No Brasil existem modelos gratuitos de previsão de chuva como o BAM e o ETA, mas devido à grande complexidade e dimensão destes dados, o trabalho não explorou resultados com os mesmos, já que para uma estatística aceitável seriam necessárias centenas de cenários.

5.1 TRABALHOS FUTUROS

Como apresentado no documento, a arquitetura SMAP-DNN apresenta diversas possibilidades, desde a forma do acoplamento, camadas *Deep Learning*, número de neurônios, funções de ativação, etc. A solução final apresentou excelentes resultados, mas que podem ser melhorados. Algumas possibilidades foram levantadas, mas por questões de tempo e complexidade computacional não foram implementadas.

A primeira sugestão seria utilizar tanto o acoplamento SMAP-DNN de pós-processamento série quanto o processamento paralelo utilizado na Tese (vide Figura 41). Em vez do SMAP fornecer apenas suas variáveis intermediárias de estado, forneceria em conjunto a saída de vazão prevista. Como mostrado nos Resultados, a acurácia do SMAP na previsão de afluência é bem inferior ao *Deep Learning*. A pergunta é: a estimativa do SMAP em adicional ajudaria de alguma forma, seria irrelevante ou confundiria o modelo híbrido?

Figura 41 – Possibilidade de acoplamento SMAP-DNN Série e Paralelo.

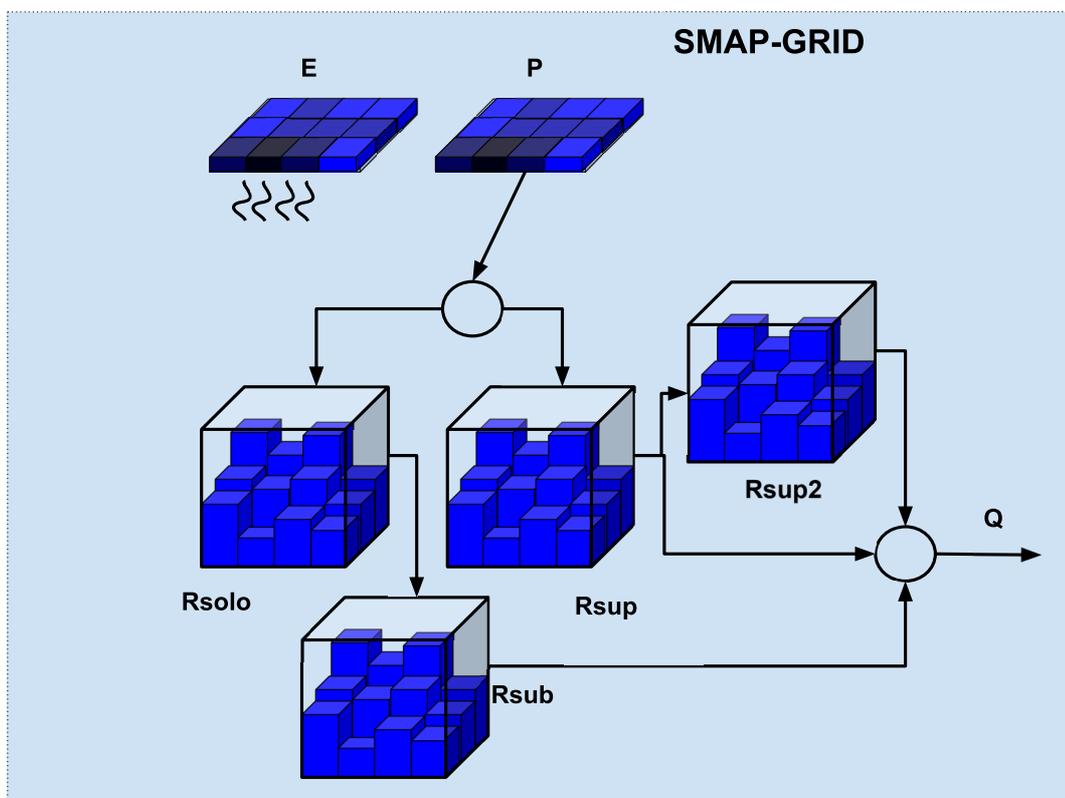


Fonte: Elaborada pelo autor.

Um outra possibilidade consiste na modificação do modelo SMAP para padrão distribuído. Como as entradas de chuva e evapotranspiração são representações espaciais em formato matricial, os reservatórios do modelo SMAP podem ser fracionados (vide Figura 42). Cada quadrícula representa um modelo SMAP independente, cujas saídas

individuais são ponderadas para produzir uma vazão única no ponto de interesse da sub-bacia. Portanto, como resultado, o modelo Híbrido receberia imagens representando a distribuição da água nas camadas do solo sobre o território. Esta possibilidade chegou a ser estudada, aparentando ser promissora, mas devido ao enorme esforço computacional não houve continuidade. Seria necessária uma computação paralela em nuvem para lidar com os milhares de parâmetros do problema de otimização resultante, ou alguma estratégia de aprendizado não supervisionado para clusterizar os *grids* diminuindo a dimensão do problema.

Figura 42 – SMAP-GRID



Fonte: Elaborada pelo autor.

REFERÊNCIAS

- ALLIBHAI, E. *Building A Deep Learning Model using Keras*. 2018. Disponível em: <<https://towardsdatascience.com/building-a-deep-learning-model-using-keras-1548ca149d37>>. Acesso em: 10/08/2020.
- ALMEIDA, L.; SERRA, J. C. V. Modelos hidrológicos, tipos e aplicações mais utilizadas. *Revista da FAE*, v. 20, n. 1, p. 129–137, 2017.
- BARCELLOS, P. d. C. L.; CATALDI, M. Flash flood and extreme rainfall forecast through one-way coupling of wrf-smap models: Natural hazards in rio de janeiro state. *Atmosphere*, Multidisciplinary Digital Publishing Institute, v. 11, n. 8, p. 834, 2020.
- BELAYNEH, A.; ADAMOWSKI, J.; KHALIL, B.; OZGA-ZIELINSKI, B. Long-term spi drought forecasting in the awash river basin in ethiopia using wavelet neural network and wavelet support vector regression models. *Journal of Hydrology*, Elsevier, v. 508, p. 418–429, 2014.
- BELVEDERESI, C.; DOMINIC, J. A.; HASSAN, Q. K.; GUPTA, A.; ACHARI, G. Predicting river flow using an ai-based sequential adaptive neuro-fuzzy inference system. *Water*, Multidisciplinary Digital Publishing Institute, v. 12, n. 6, p. 1622, 2020.
- BENGIO, Y.; GOODFELLOW, I.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press Massachusetts, USA:, 2017. v. 1.
- BHUIYAN, H. A.; MCNAIRN, H.; POWERS, J.; MERZOUKI, A. Application of hec-hms in a cold region watershed and use of radarsat-2 soil moisture in initializing the model. *Hydrology*, Multidisciplinary Digital Publishing Institute, v. 4, n. 1, p. 9, 2017.
- BOU, A. F.; SÁ, R. V. D.; CATALDI, M. Flood forecasting in the upper uruguay river basin. *Natural Hazards*, Springer, v. 79, n. 2, p. 1239–1256, 2015.
- CAMPOS, L. A. Comparação de meta-heurísticas na calibração do modelo smap/ons. *Trabalho de Conclusão - Engenharia Elétrica - UFJF*, 2021.
- CAPONI, G. R.; SILVA, B.; SANTOS, A. H. M. Calibração automática de um modelo chuva-vazão mensal por algoritmos genéticos. In: *21st Simpósio Brasileiro de Recursos Hídricos*. [S.l.: s.n.], 2011.
- CHANG, C.-H.; LEE, H.; HOSSAIN, F.; BASNAYAKE, S.; JAYASINGHE, S.; CHISHTIE, F.; SAAH, D.; YU, H.; SOTHEA, K.; BUI, D. D. A model-aided satellite-altimetry-based flood forecasting system for the mekong river. *Environmental modelling & software*, Elsevier, v. 112, p. 112–127, 2019.
- CHEN, C.; HUI, Q.; PEI, Q.; ZHOU, Y.; WANG, B.; LV, N.; LI, J. Crml: A convolution regression model with machine learning for hydrology forecasting. *IEEE Access*, IEEE, v. 7, p. 133839–133849, 2019.
- FIGUEROA, S. N.; BONATTI, J. P.; KUBOTA, P. Y.; GRELL, G. A.; MORRISON, H.; BARROS, S. R.; FERNANDEZ, J. P.; RAMIREZ, E.; SIQUEIRA, L.; LUZIA, G. *et al.* The brazilian global atmospheric model (bam): performance for tropical rainfall forecasting and sensitivity to convective scheme and horizontal resolution. *Weather and Forecasting*, v. 31, n. 5, p. 1547–1572, 2016.

FLEMING, S. W.; GOODBODY, A. G. A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the us west. *IEEE Access*, IEEE, v. 7, p. 119943–119964, 2019.

GERS, F. A.; SCHMIDHUBER, J.; CUMMINS, F. Learning to forget: Continual prediction with lstm. IET, 1999.

GHOORBANI, M. A.; KHATIBI, R.; MEHR, A. D.; ASADI, H. Chaos-based multigene genetic programming: A new hybrid strategy for river flow forecasting. *Journal of hydrology*, Elsevier, v. 562, p. 455–467, 2018.

HARGREAVES, G. H.; SAMANI, Z. A. Reference crop evapotranspiration from temperature. *Applied engineering in agriculture*, American Society of Agricultural and Biological Engineers, v. 1, n. 2, p. 96–99, 1985.

JUNIOR, J. R. F. *Redes Neurais Recorrentes — LSTM*. 2019. Disponível em: <<https://medium.com/@web2ajax/redes-neurais-recorrentes-lstm-b90b720dc3f6>>. Acesso em: 24/06/2021.

KARPLUS, W. J. The spectrum of mathematical modeling and systems simulation. *ACM SIGSIM Simulation Digest*, ACM New York, NY, USA, v. 9, n. 1, p. 32–38, 1977.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

KWON, M.; KWON, H.-H.; HAN, D. A hybrid approach combining conceptual hydrological models, support vector machines and remote sensing data for rainfall-runoff modeling. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 11, p. 1801, 2020.

LIMA, J. C. d.; ARRAES, F. D. D.; OLIVEIRA, J. B. d.; NASCIMENTO, F. A. L. d.; MACÊDO, K. G. d. Parametrização da equação de hargreaves e samani para estimativa da evapotranspiração de referência no estado do ceará, brasil. *Revista Ciência Agronômica*, SciELO Brasil, v. 47, p. 447–454, 2016.

LITTLE, Z. *Conv1D, Conv2D and Conv3D*. 2020. Disponível em: <<https://xzz201920.medium.com/conv1d-conv2d-and-conv3d-8a59182c4d6>>. Acesso em: 10/10/2020.

LOOS, S.; SHIN, C. M.; SUMIHAR, J.; KIM, K.; CHO, J.; WEERTS, A. H. Ensemble data assimilation methods for improving river water quality forecasting accuracy. *Water research*, Elsevier, v. 171, p. 115343, 2020.

LOPES, J. E. G.; JR, B. B.; CONEJO, J. Smap—a simplified hydrologic model. *Applied modeling in catchment hydrology/ed. by VP Singh*, Littleton, Colo.: Water Resources Publications, 1982., 1982.

MARCON, G. F. Modelagem chuva-vazão utilizando o modelo matemático smap para um trecho da sub-bacia rio carreiro-rs. 2019.

MENG, S.; XIE, X.; LIANG, S. Assimilation of soil moisture and streamflow observations to improve flood forecasting with considering runoff routing lags. *Journal of Hydrology*, Elsevier, v. 550, p. 568–579, 2017.

MOROMISATO, G. D. Y. *Programação Dinâmica Aplicada ao Cálculo da Energia Firme de Usinas Hidrelétricas*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Brasil, 2012.

NASCIMENTO, L. S. V.; JR, D. S. R.; MARTINS, E. S. P. Avaliação do algoritmo evolutivo mopso na calibração multiobjetivo do modelo smap no estado do ceará. *Diretoria da ABRH*, p. 85, 2009.

NELLES, O. *Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*. [S.l.]: Springer Nature, 2020.

ONS. Aplicação de Modelos de Previsão de Vazões a Curto Prazo na Sub-bacia do Alto Rio Paranaíba: Relatório final. ONS 153/12. [S. l.: s. n.], 20 ago. 2012. 161 p. Revisão: 1. Brasil, 2012.

ONS. Manual de metodologia do Aplicativo SMAP. Brasil, 2017.

PAIVA, L. F. G. D.; MONTENEGRO, S. M.; CATALDI, M. Prediction of monthly flows for três marias reservoir (são francisco river basin) using the cfs climate forecast model. *RBRH*, SciELO Brasil, v. 25, 2020.

PAPACHARALAMPOUS, G.; TYRALIS, H.; LANGOUSIS, A.; JAYAWARDENA, A. W.; SIVAKUMAR, B.; MAMASSIS, N.; MONTANARI, A.; KOUTSOYIANNIS, D. Probabilistic hydrological post-processing at scale: Why and how to apply machine-learning quantile regression algorithms. *Water*, Multidisciplinary Digital Publishing Institute, v. 11, n. 10, p. 2126, 2019.

PATIL, A.; RAMSANKARAN, R. Improving streamflow simulations and forecasting performance of swat model by assimilating remotely sensed soil moisture observations. *Journal of Hydrology*, Elsevier, v. 555, p. 683–696, 2017.

PEREIRA, W. J. X. Calibração de um modelo hidrológico de escala diária para bacias hidrográficas do semiárido cearense. 2017.

RAMPINELLI, C.; COSTA, M. Calibração do modelo chuva-vazão smap em uma estação do semiárido nordestino com os algoritmos pso ag e nelder mead e seus efeitos na definição das séries de vazões médias mensais. In: *21st Simpósio Brasileiro de Recursos Hídricos*. [S.l.: s.n.], 2015.

RASOULI, K.; HSIEH, W. W.; CANNON, A. J. Daily streamflow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, Elsevier, v. 414, p. 284–293, 2012.

REZAIE-BALF, M.; KIM, S.; FALLAH, H.; ALAGHMAND, S. Daily river flow forecasting using ensemble empirical mode decomposition based heuristic regression models: Application on the perennial rivers in iran and south korea. *Journal of Hydrology*, Elsevier, v. 572, p. 470–485, 2019.

ROKAYA, P.; MORALES-MARIN, L.; LINDENSCHMIDT, K.-E. A physically-based modelling framework for operational forecasting of river ice breakup. *Advances in Water Resources*, Elsevier, v. 139, p. 103554, 2020.

ROZANTE, J. R.; MOREIRA, D. S.; GONCALVES, L. G. G. de; VILA, D. A. Combining trmm and surface observations of precipitation: technique and validation over south america. *Weather and forecasting*, v. 25, n. 3, p. 885–894, 2010.

SAMUEL, J.; ROUSSEAU, A. N.; ABBASNEZHADI, K.; SAVARY, S. Development and evaluation of a hydrologic data-assimilation scheme for short-range flow and inflow forecasts in a data-sparse high-latitude region using a distributed model and ensemble kalman filtering. *Advances in Water Resources*, Elsevier, v. 130, p. 198–220, 2019.

SARAIVA, I. S. Regionalização de parâmetros de modelos de simulação hidrológica para emprego em bacias hidrográficas sem monitoramento fluviométrico. *Mestrado em meio ambiente e recursos hídricos, Universidade Federal de Minas Gerais, Belo Horizonte*, 2010.

SCIKIT-LEARNING. *Cross-validation: evaluating estimator performance*. 2019. Disponível em: <https://scikit-learn.org/stable/modules/cross_validation.html>. Acesso em: 24/08/2020.

SCIPY. *SciPy Reference Guide, Differential Evolution*. 2020. Disponível em: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.differential_evolution.html>. Acesso em: 24/08/2020.

SHANG, Y.; XU, Y.; SHANG, L.; FAN, Q.; WANG, Y.; LIU, Z. A method of direct, real-time forecasting of downstream water levels via hydropower station reregulation: A case study from gezhouba hydropower plant, china. *Journal of Hydrology*, Elsevier, v. 573, p. 895–907, 2019.

SILVA, F. d. N. R. da; ALVES, J. L. D.; CATALDI, M. Climate downscaling over south america for 1971–2000: application in smap rainfall-runoff model for grande river basin. *Climate Dynamics*, Springer, v. 52, n. 1, p. 681–696, 2019.

SILVA, R. B. *Ajuste do modelo SMAP via técnicas de inteligência computacional*. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, Brasil, 2019.

SILVA, S.; JÚNIOR, H. C.; ROCHA, V.; TORRACA, H. Avaliação do uso de dados de satélite para previsão de vazões utilizando o modelo hidrológico smap/ons. *Proceedings of the XXIII SBRH 2019*, ABRH Porto Alegre, 2019.

SONG, C. M. Hydrological image building using curve number and prediction and evaluation of runoff through convolution neural network. *Water*, Multidisciplinary Digital Publishing Institute, v. 12, n. 8, p. 2292, 2020.

STORN, R.; PRICE, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, Springer, v. 11, n. 4, p. 341–359, 1997.

THRUN, S. *Twiddle - Artificial Intelligence for Robotics [Streaming video]*. 2018. Disponível em: <<https://www.youtube.com/watch?v=2uQ2BSzDvXs>>. Acesso em: 24/08/2020.

WIJAYARATHNE, D. B.; COULIBALY, P. Identification of hydrological models for operational flood forecasting in st. john's, newfoundland, canada. *Journal of Hydrology: Regional Studies*, Elsevier, v. 27, p. 100646, 2020.

- WÖHLING, T.; BURBERY, L. Eigenmodels to forecast groundwater levels in unconfined river-fed aquifers during flow recession. *Science of The Total Environment*, Elsevier, v. 747, p. 141220, 2020.
- XIE, T.; ZHANG, G.; HOU, J.; XIE, J.; LV, M.; LIU, F. Hybrid forecasting model for non-stationary daily runoff series: A case study in the han river basin, china. *Journal of Hydrology*, Elsevier, v. 577, p. 123915, 2019.
- XU, L.; CHEN, N.; ZHANG, X.; CHEN, Z. An evaluation of statistical, nmme and hybrid models for drought prediction in china. *Journal of hydrology*, Elsevier, v. 566, p. 235–249, 2018.
- YASEEN, Z. M.; SULAIMAN, S. O.; DEO, R. C.; CHAU, K.-W. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, Elsevier, v. 569, p. 387–408, 2019.
- YIN, S.; TANG, D.; JIN, X.; CHEN, W.; PU, N. A combined rotated general regression neural network method for river flow forecasting. *Hydrological Sciences Journal*, Taylor & Francis, v. 61, n. 4, p. 669–682, 2016.
- ZHOU, T.; JIANG, Z.; LIU, X.; TAN, K. Research on the long-term and short-term forecasts of navigable river's water-level fluctuation based on the adaptive multilayer perceptron. *Journal of Hydrology*, Elsevier, p. 125285, 2020.
- ZHOU, Y.; GUO, S.; XU, C.-Y.; CHANG, F.-J.; YIN, J. Improving the reliability of probabilistic multi-step-ahead flood forecasting by fusing unscented kalman filter with recurrent neural network. *Water*, Multidisciplinary Digital Publishing Institute, v. 12, n. 2, p. 578, 2020.
- ZHU, S.; HRNJICA, B.; PTAK, M.; CHOIŃSKI, A.; SIVAKUMAR, B. Forecasting of water level in multiple temperate lakes using machine learning models. *Journal of Hydrology*, Elsevier, p. 124819, 2020.
- ZUO, G.; LUO, J.; WANG, N.; LIAN, Y.; HE, X. Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting. *Journal of Hydrology*, Elsevier, v. 585, p. 124776, 2020.