

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE PESQUISAS HIDRÁULICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS E SANEAMENTO
AMBIENTAL

JULIANO SANTOS FINCK

PREVISÃO EM TEMPO ATUAL DE NÍVEIS FLUVIAIS COM REDES NEURAIAS
ARTIFICIAIS:
Aplicação à bacia do rio Taquari-Antas/RS

Porto Alegre

2020

JULIANO SANTOS FINCK

PREVISÃO EM TEMPO ATUAL DE NÍVEIS FLUVIAIS COM REDES NEURAIIS
ARTIFICIAIS:
Aplicação à bacia do rio Taquari-Antas/RS

Dissertação apresentada ao programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de Mestre em Recursos Hídricos e Saneamento Ambiental.

Orientador: Prof. Dr. Olavo Correa Pedrollo

Porto Alegre

2020

CIP - Catalogação na Publicação

Santos Finck, Juliano
PREVISÃO EM TEMPO ATUAL DE NÍVEIS FLUVIAIS COM
REDES NEURAIS ARTIFICIAIS: Aplicação à bacia do rio
Taquari-Antas/RS / Juliano Santos Finck. -- 2020.
103 f.
Orientador: Olavo Correa Pedrollo.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Instituto de Pesquisas Hidráulicas,
Programa de Pós-Graduação em Recursos Hídricos e
Saneamento Ambiental, Porto Alegre, BR-RS, 2020.

1. Previsão hidrológica. 2. Redes Neurais
Artificiais. 3. Falta de sinal. 4. Registros de
treinamento. 5. Complexidade interna. I. Correa
Pedrollo, Olavo, orient. II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Rui Vicente Oppermann

Vice-Reitora: Prof.^a Dr.^a Jane Fraga Tutikian

INSTITUTO DE PESQUISAS HIDRÁULICAS

Diretor: Prof. Dr. Carlos André Bulhões Mendes

Vice Diretora: Prof.^a Dr.^a Edith Beatriz Camaño Schettini

PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS E SANEAMENTO
AMBIENTAL

Coordenador: Prof. Dr. Antônio Domingues Benetti

Vice Coordenador: Prof.^a Dr.^a Lúcia Helena Ribeiro Rodrigues

JULIANO SANTOS FINCK

PREVISÃO EM TEMPO ATUAL DE NÍVEIS FLUVIAIS COM REDES NEURAIAS

ARTIFICIAIS:

Aplicação à bacia do rio Taquari-Antas/RS

Dissertação apresentada ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul como requisito parcial à obtenção do título de mestre em Recursos Hídricos e Saneamento Ambiental.

Aprovado em: Porto Alegre, 16 de março de 2020.

Prof. Dr. Olavo Correa Pedrollo – PPGRHSA / UFRGS

Orientador

Prof. Dr. Guilherme Garcia de Oliveira – PPGSR/UFRGS

Examinador

Prof. Dr. Juan Martín Bravo – PPGRHSA/UFRGS

Examinador

Prof. Dr. ^a Nilza Maria dos Reis Castro – PPGRHSA/UFRGS

Examinadora

Prof. Dr. Antônio Domingues Benetti

Coordenador do PPGRHSA/IPH

Dedico esse trabalho à Ciência e à Universidade
Pública de qualidade. Que elas guiem nossa
Sociedade.

AGRADECIMENTOS

Ao Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental - PPGHRSA, pela oportunidade de realização de trabalhos em minha área de pesquisa.

Ao Instituto de Pesquisas Hidráulicas – IPH/UFRGS pela infraestrutura necessária para a realização dessa pesquisa e outras oportunidades, como o estágio de docência e o uso de laboratórios para aulas práticas.

À CPRM/SGB – Serviço Geológico Brasileiro – nas figuras de André Chagas e Márcia Pedrollo e Agência Nacional de Águas, por conceder os dados para realização dessa pesquisa e me receber em visitas a SUREG-PA.

Ao Prof. Dr. Olavo Correa Pedrollo, pelas tardes de orientação, conhecimento do tema do trabalho, e incentivar a carreira de professor-pesquisador, dosando importante humanidade nos seus conselhos em momentos complicados, culminando, por fim, em um profissional exemplar.

Aos colegas do grupo de pesquisa AHRNA, em especial a Juliana Andrade, Lais Teixeira e Luísa Lucchese por auxiliar em revisões e pela amizade no IPH.

Aos colegas do PPG, integrantes do WARP, Ana Paula Dalcin, Carlos Henrique, Gláucio Souza, Joseli Maia, Luana Pandolfo e Márcio Shigueaki pela troca mútua de saberes científicos e pessoais que contribuíram para agir com calma e cuidado.

Aos momentos tranquilos com The Zaguadados – Ayan Santos, Gean Michel, Rodrigo Paiva e eventuais participações de Vinícius Siqueira – que, na camaradagem, em tardes pós-IPH, trouxeram toques artísticos a uma rotina de estudos.

Aos amigos de longa data que me acompanham desde a graduação: Aline Meyer, Mel Guirro e Renata Rossoni. Em especial, àquele ser niilista-otimista que nunca se deve alimentar, que lembra que “ninguém existe de propósito, ninguém pertence a lugar algum, e todos morrerão. Assim, vamos assistir TV”: Iporã Brito Possanti.

À minha namorada, companheira e amiga que participou das noites desse processo que foi escrever essa dissertação, que transformou angústia em confiança, ansiedade em calma, estresse em gargalhada e desperta a Arte em mim: Sofia Victoria Urra Essedin. Que nossos sonhos se realizem e possamos desfrutar muitos outros dias ainda!

Ao CNPq pela provisão da bolsa de mestrado.

RESUMO

Modelos de previsão hidrológica baseados em Redes Neurais Artificiais (RNAs) constituem uma das melhores alternativas para sistemas de alerta de inundações, que fornecem tempo de reação à Proteção e Defesa Civil para reduzir danos. A seguinte pesquisa objetivou determinar a aplicabilidade de RNAs para prever em tempo atual níveis fluviais em estações do sistema de alerta da bacia do rio Taquari-Antas/RS (SACE-Taquari), usando exclusivamente dados da rede telemétrica existente. Oito séries de precipitação e cinco séries de níveis — ambas medindo com frequência horária durante o período de 26/11/2015 até 24/04/2019 — foram usadas para prever os níveis fluviais entre 4 e 24 horas à frente em Encantado, Estrela, Porto Mariante e Taquari. Para agregar dados de precipitação no tempo, foram comparados três filtros de média móvel: o de ponderação Uniforme (FMMU), Exponencial (FMME) e Gama (FMMG). Para lidar com a eventual indisponibilidade de dados, Árvores de RNAs foram utilizadas, contemplando os possíveis cenários de perda de sinal. O período de verificação foi o ano de 2017. Seis eventos foram avaliados visualmente; quatro deles pertencendo ao ano de 2017. Previsões com alcance de 8, 12 e 24 horas apresentaram, em média, coeficientes de Nash-Sutcliffe (NS) de 0,93, 0,89 e 0,71 e médias do erro absoluto (MEA) de 12, 16 e 20 cm, respectivamente. Os desempenhos obtidos confirmam a aplicabilidade das RNAs com a metodologia utilizada. FMME foi o pior dos filtros investigados para agregar dados de precipitação. A maior flexibilidade do novo FMMG possibilitou desempenho semelhante ao FMMU com menos registros, tendo sido selecionado para agregar os dados de precipitação. A abordagem das Árvores de RNAs possibilita emitir previsões a qualquer momento, dado que o cenário de indisponibilidade foi considerado durante o procedimento de treinamento. Porém, não substituem a urgência de manutenção das estações durante eventos extremos. De qualquer forma, entende-se que o investimento nas redes de monitoramento e nas técnicas de previsão utilizando RNAs é importante para aprimorar a previsão hidrológica em sistemas de alerta de evento críticos.

Palavras-chave: Falta de sinal, Complexidade Interna, Registros de Treinamento, Filtros de média móvel.

ABSTRACT

Hydrological forecasting models based on Artificial Neural Networks (ANN) constitute one of the best alternatives for early warning systems — an approach to provide more reaction time to the Civil Defence for damage reduction. This study aims to evaluate the applicability of ANNs to forecast in real-time water levels in stations of the Taquari-Antas early warning system, using solely data of its hourly telemetric network. Eight rainfall and five water level series — whose data period spans from 26/11/2016 until 24/04/2019— were used to forecast water levels up to 24 hours ahead at Encantado, Estrela, Porto Mariante and Taquari. To aggregate rainfall data in time, three moving average filters were compared: Uniformly Weighted (UWMA), Exponentially Weighted (EWMA), and the novel Gamma-Weighted (GWMA). ANN Trees were used to cope with eventual scenarios of signal loss. The verification period was the year of 2017. Six events were visually evaluated; four of them pertaining to the year of 2017. Forecasts at 8, 12 and 24 hours ahead presented, on average, Nash-Sutcliffe coefficients (NS) of 0.93, 0.89 and 0.71 and EMA of 12, 16 and 20 cm, respectively. The obtained performances confirm the applicability of ANNs with the applied methodology. EWMA appeared to be the worst filter at aggregating rainfall data. The greater flexibility of the novel GWMA allowed a similar performance to the UWMA with fewer entries, thus, being selected to aggregate rainfall data. The approach of the Tree of ANNs makes it possible to issue forecasts at any time, given that the unavailability scenario was considered during the training procedure. However, they do not substitute the urgency in maintenance of the station during extreme events. In any case, investing in monitoring networks and in forecasting techniques using ANNs is important to enhance hydrological forecasting in early warning systems.

Keywords: Signal loss, Internal complexity, Training registers, Moving-average filters.

LISTA DE ILUSTRAÇÕES

Figura 1. Componente de um sistema de alerta de eventos críticos.	7
Figura 2. SACE CPRM/SGB.	8
Figura 3. Rede neural biológica e alguns componentes.	14
Figura 4. Neurônios e proposições lógicas: conjunção, disjunção e negação.	15
Figura 5. O perceptron óptico.	16
Figura 6. A máquina ADALINE.	16
Figura 7. Sobreajustamento de uma rede.	20
Figura 8. Representação da topografia da bacia do rio Taquari-Antas. A linha vermelha apresenta o trecho mais crítico a inundações.	39
Figura 9. Precipitação mensal média, temperatura mensal máxima, temperatura mensal média e temperatura mensal mínima da série da estação Bento Gonçalves do INMET entre 01/01/1961 e 31/12/2013.	40
Figura 10. Precipitação mensal média na bacia do rio Taquari-Antas/RS no período de 1977-2006, calculadas com 15 estações pluviométricas.	41
Figura 11. Cotas máximas de inundação em Lajeado/RS, período de 1873-2018.	42
Figura 12. Número de ocorrências de inundação por mês em Lajeado/RS, período de 1940-2018.	43
Figura 13. Representações gráficas do (A.) Mapa geológico regional das unidades geológico-ambientais obtido no GeoSGB da CPRM/SGB (2020) e do (B.) Mapa pedológico do Brasil seguindo o Sistema Brasileiro de Classificação de Solos (SiBCS, 2006) obtido no GeoINFO da EMBRAPA (2020).	45
Figura 14. Evolução no uso do solo na bacia do rio Taquari-Antas de 1985 para 2018.	46
Figura 15. Distribuição das estações utilizadas.	46
Figura 16. Comparação entre dados de Muçum e Linha José Júlio.	48
Figura 17. Visualização gráfica da repartição dos períodos de treinamento e validação do período de verificação (2017) para a estação de Estrela.	50
Figura 18. Diagrama unifilar das estações utilizadas, sintetizando informações sobre a área controlada e com as redes neurais pretendidas.	51
Figura 19. Fluxograma da Metodologia. <i>NaN – Not a number</i> ; CM – Modelo Completo; PM – Modelo Parcial.	53
Figura 20. Visualização dos seis eventos selecionados para demonstrar o desempenho dos modelos. E1 e E2 fazem parte dos conjuntos de treinamento e validação. E3, E4, E5 e E6 fazem parte do conjunto de verificação.	54
Figura 21. Precipitação com FMMU, FMME e FMMG - Passo Tainhas e Localidade (ENC, EST, PM, TA).	56
Figura 22. Pesos do FMMG para Passo Tainhas nos alcances de 4, 14 e 24 horas.	57
Figura 23. Análise dos filtros nos níveis de Muçum e diferença, t e $t+alc$, entre níveis das localidades de previsão (ENC, EST, PM, TA) - Pesquisa dos filtros FMMU, FMME e FDM.	57
Figura 24. Nível – Análise de filtros nos níveis da localidade e diferença (t e $t+alc$) entre níveis das localidades de previsão (ENC, EST, PM, TA) - Pesquisa dos filtros FMMU, FMME e FDM.	58
Figura 25. Pesquisa do número de registros de treinamento para a localidade de Encantado para os alcances de 4, 14 e 24 horas.	59
Figura 26. Pesquisa do número de registros de treinamento para localidade de Taquari para os alcances de 4, 14 e 24 horas.	60
Figura 27. Pesquisa da complexidade interna para a RNA da localidade de Encantado para os alcances de 4, 6, 12, 16 e 24 horas.	61
Figura 28. Pesquisa da complexidade interna para a RNA da localidade de Taquari para os alcances de 4, 6, 12, 16 e 24 horas.	62

Figura 29. Ilustração da rna_ENC, referente à Tabela 8. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).....	63
Figura 30. Ilustração da rna_EST, referente à Tabela 10. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).....	64
Figura 31. Ilustração da rna_PM, referente à Tabela 12. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).....	65
Figura 32. Ilustração da rna_TA, referente à Tabela 14. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).....	66
Figura 33. Efeito da retirada dos dados de precipitação do posto Vacaria (pPV), representado pelo PM2 (modelo parcial 2) em relação ao modelo completo (CM) para Encantado nas métricas NS, MEA, E10 e E90.	68
Figura 34. Efeito da retirada dos dados de precipitação, representado pelo PM11 (modelo parcial 11) em relação ao modelo completo (CM) desenvolvido para Encantado nas métricas NS, MEA, E10 e E90.	68
Figura 35. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Estrela nas métricas NS, MEA, E10 e E90.....	69
Figura 36. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Porto Mariante nas métricas NS, MEA, E10 e E90.	69
Figura 37. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Taquari nas métricas NS, MEA, E10 e E90.....	70
Figura 38. Previsão em tempo atual feita em ENC para o Evento 1 (participou do treinamento).....	71
Figura 39. Previsão em tempo atual feita em EST para o Evento 1 (participou do treinamento).....	72
Figura 40. Previsão em tempo atual feita em PM para o Evento 1 (participou do treinamento).....	72
Figura 41. Previsão em tempo atual feita em TA para o Evento 1 (participou do treinamento).....	72
Figura 42. Previsão em tempo atual feita em ENC para o Evento 2 (participou do treinamento).....	73
Figura 43. Previsão em tempo atual feita em EST para o Evento 2 (participou do treinamento).....	73
Figura 44. Previsão em tempo atual feita em PM para o Evento 2 (participou do treinamento).....	74
Figura 45. Previsão em tempo atual feita em TA para o Evento 2 (participou do treinamento).....	74
Figura 46. Previsão em tempo atual feita em ENC para o Evento 3 (não presente do treinamento). 75	
Figura 47. Previsão em tempo atual feita em EST para o Evento 3 (não presente do treinamento).. 75	
Figura 48. Previsão em tempo atual feita em PM para o Evento 3 (não presente do treinamento).. 75	
Figura 49. Previsão em tempo atual feita em TA para o Evento 3 (não presente do treinamento). ... 76	
Figura 50. Previsão em tempo atual feita em ENC para o Evento 4 (não presente do treinamento). 76	
Figura 51. Previsão em tempo atual feita em EST para o Evento 4 (não presente do treinamento).. 76	
Figura 52. Previsão em tempo atual feita em PM para o Evento 4 (não presente do treinamento). ... 77	
Figura 53. Previsão em tempo atual feita em TA para o Evento 4 (não presente do treinamento). ... 77	
Figura 54. Previsão em tempo atual feita em ENC para o Evento 5 (não presente do treinamento). 78	
Figura 55. Previsão em tempo atual feita em EST para o Evento 5 (não presente do treinamento).. 78	
Figura 56. Previsão em tempo atual feita em PM para o Evento 5 (não presente do treinamento)... 78	
Figura 57. Previsão em tempo atual feita em TA para o Evento 5 (não presente do treinamento). ... 79	
Figura 58. Previsão em tempo atual feita em ENC para o Evento 6 (não presente do treinamento). 80	
Figura 59. Previsão em tempo atual feita em EST para o Evento 6 (não presente do treinamento).. 80	
Figura 60. Previsão em tempo atual feita em PM para o Evento 6 (não presente do treinamento). ... 80	
Figura 61. Previsão em tempo atual feita em TA para o Evento 6 (não presente do treinamento). ... 81	

LISTA DE TABELAS

Tabela 1. Tabela de referências em aplicação de RNAs, e trabalhos de previsão na área de estudo. EMQ, Erro Médio Quadrado; MER, Média do Erro Relativo; REMQ: Raiz do Erro Médio Quadrado; r: coeficiente de correlação de Pearson; E90: Percentil 90% do erro. MEA, Erro Absoluto Médio.....	25
Tabela 2. Uso do solo por categorias na bacia do rio Taquari-Antas nos anos de 1985, 2015 e 2018. ID é o número de identificação da classe utilizado no MapBiomas v4.0.	45
Tabela 3. Informações das estações do SACE-Taquari. Fluviométrica – “f”. Pluviométrica – “p”. .	47
Tabela 4. Valores de mínimo e máximo nas séries de cota para ENC, EST, PM e TA.	51
Tabela 5. Defasagem horária referente à maior correlação entre os dados de nível: estação (defasagem em horas, coeficiente de Pearson associado, distância em km até estação jusante).	55
Tabela 6. Número de registros de treinamento adotado para cada localidade.	61
Tabela 7. Número de neurônios internos adotado para cada localidade.	62
Tabela 8. Tabela de apresentação, variáveis de entrada e saída para rna_ENC (Encantado). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, EN – Encantado, f – Cota, p – Precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.	63
Tabela 9. Métricas de avaliação da verificação para a rna_ENC. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.	63
Tabela 10. Tabela de apresentação, variáveis de entrada e saída para rna_EST (Estrela). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.	64
Tabela 11. Métricas de avaliação da verificação para a rna_EST. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.	64
Tabela 12. Tabela de apresentação, variáveis de entrada e saída para rna_PM (Porto Mariante). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, PM – Porto Mariante, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.	65
Tabela 13. Métricas de avaliação da verificação para a rna_PM. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.	65
Tabela 14. Tabela de apresentação, variáveis de entrada e saída para rna_TA (Taquari). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, PM – Porto Mariante, TA – Taquari, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.	66
Tabela 15. Métricas de avaliação da verificação para a rna_TA. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. –	

coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.....	66
Tabela 16. Tabela-resumo quanto aos modelos desenvolvidos nessa pesquisa. CM é modelo completo. PM é modelo parcial. PM sem 1 variável é quando se retira uma variável, tal que PM5 é aquele modelo em que a 5ª variável foi retirada.....	67
Tabela 17. Modelos que compõe a Árvore de RNAs “rna_ENC”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.	97
Tabela 18. Modelos que compõe a Árvore de RNAs “rna_EST”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.	98
Tabela 19. Modelos que compõe a Árvore de RNAs “rna_PM”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.	100
Tabela 20. Modelos que compõe a Árvore de RNAs “rna_TA”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.	102

LISTA DE ABREVIATURAS E SIGLAS

- ADALINE - Máquina de Adaptação Linear
- AHSUL - Administração Hidroviária do Sul
- ANEEL - Agência Nacional de Energia Elétrica
- AR - Modelo autorregressivo
- ARIMA - Modelo autorregressivo integrado com média móvel
- ARIMAX - Modelo autorregressivo integrado com variáveis exógenas
- CM - Modelo completo
- CPRM/SGB - Companhia de Pesquisas de Recursos Minerais/Serviço Geológico Brasileiro, é uma empresa pública vinculada ao Ministério de Minas e Energia
- CPTEC - Centro de Previsão de Tempo e Estudos Climáticos
- d - Diferença de níveis
- DRH/SEMA - Departamento de Recursos Hídricos/Secretaria Estadual do Meio Ambiente e Infraestrutura
- MEA - Média do Erro Absoluto
- E10 - Percentil 10% da distribuição dos erros
- E25 - Percentil 25% da distribuição dos erros
- E50 - Percentil 50% da distribuição dos erros
- E75 - Percentil 75% da distribuição dos erros
- E90 - Percentil 90% da distribuição dos erros
- ENC - Encantado
- EST - Estrela
- ETA - Modelo regional de previsão meteorológica ETA, utilizado pelo CPTEC
- f - Série fluviométrica
- FDM - Filtro de Diferenças Móveis
- FMME - Filtro de Média Móvel Exponencial
- FMMG - Filtro de Média Móvel Gama
- FMMU - Filtro de Média Móvel Uniforme
- FFNN - *Feed Forward Neural Network*, uma rede neural que funciona por propagação unidirecional, diferenciando-se de outras do tipo recorrentes
- HadCM3 - Hadley Centre Coupled Model, version 3 (Modelo de Circulação Geral da atmosfera desenvolvido no Hadley Centre no Reino Unido)

HEC-HMS - Hydrologic Modelling System, modelo hidrológico desenvolvido no Centro de Engenharia Hidrológica (Hydrologic Engineering Center) dos Estados Unidos

IPH - Instituto de Pesquisas Hidráulicas

INPE - Instituto Nacional de Pesquisas Espaciais

MGB-IPH - Modelo de Grandes Bacias (conceitual chuva-vazão) desenvolvido no IPH

MLP - Perceptron de Múltiplas Camadas

NS - Coeficiente de eficiência de Nash Sutcliffe

p - Série pluviométrica

PM - Porto Mariante

PM# - Modelo parcial número #

PPGRHSA - Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental

PPGSR - Programa de Pós-Graduação em Sensoriamento Remoto

PR - Estado do Paraná, Brasil

RBF - Funções de Base Radiais

RLM - Regressão Linear Múltipla

RLS - Mínimos quadrados recursivos

RNA - Rede Neural Artificial

RS - Estado do Rio Grande do Sul, Brasil

SACE - Sistema de Alerta de Eventos Críticos

SC - Estado de Santa Catarina, Brasil

SED - Sistema Especialista Difuso

SISMADEN - Sistema de Monitoramento de Desastres

TA - Taquari

UFRGS - Universidade Federal do Rio Grande do Sul

UHE - Usina Hidroelétrica

UNIVATES- Universidade do Vale do Taquari

SUMÁRIO

RESUMO	v
ABSTRACT	vi
LISTA DE ILUSTRAÇÕES	vii
LISTA DE TABELAS	ix
LISTA DE ABREVIATURAS E SIGLAS	xi
SUMÁRIO	xiii
1. INTRODUÇÃO	1
1.1. Contextualização	1
1.2. Motivação	2
1.3. Objetivo	3
2. REVISÃO BIBLIOGRÁFICA	4
2.1. Ocupação de áreas suscetíveis no Brasil	4
2.2. Sistemas de alerta	6
2.3. Previsão	9
2.4. Modelos hidrológicos	11
2.5. Redes neurais artificiais	14
2.5.1. Histórico	14
2.5.2. Aproximador universal	19
2.5.3. Treinamento, validação e verificação	20
2.5.4. Filtros	21
2.6. Métricas para avaliação	22
2.7. Exemplos de Aplicações das RNAs	24
2.7.1. Internacional – revisões	29
2.7.2. Trabalhos no Brasil	32
2.7.3. Alguns trabalhos envolvendo previsão no Taquari	34
3. MATERIAIS E MÉTODOS	39
3.1. Área de estudo	39
3.1.1. Rio Taquari-Antas e a navegação comercial	39
3.1.2. Clima da região e sua dinâmica	40
3.1.3. Hidrografia e inundações históricas	41
3.1.4. Geologia, pedologia e uso do solo na bacia	44
3.2. Estações utilizadas	46
3.3. Análise dos dados	47
3.3.1. Consistência de dados e período utilizado	47
3.3.2. Avaliação das transformações de variáveis	48
3.3.3. Repartição dos registros	50
3.3.4. Extremos da série	51
3.4. Desenvolvimento dos modelos	51
3.5. Avaliação do desempenho do modelo	54
4. RESULTADOS	55
4.1. Análises das transformações das variáveis	55
4.2. Pesquisa do número de registros de treinamento	58
4.3. Pesquisa da Complexidade Interna	61
4.4. Modelos completos de RNAs	62
4.4.1. Encantado	63
4.4.2. Estrela	64
4.4.3. Porto Mariante	65
4.4.4. Taquari	66
4.4.5. Síntese dos modelos completos	67
4.5. Modelos parciais de RNAs	67

4.6. Análise por eventos com as Árvores de RNAs	70
4.7. Discussão dos resultados com referência à literatura	81
5. CONCLUSÕES	84
6. RECOMENDAÇÕES	86
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	87
8. APÊNDICE.....	97

1. INTRODUÇÃO

1.1. Contextualização

Inundações são os desastres naturais mais danosos no Brasil. No período de 1964-2018, cerca de 82% da perda agropecuária média anual brasileira* decorrente de desastres naturais foi causada por inundações (GUHA-SAPIR *et al.*, 2019). Tipicamente, há pouco tempo para reação quando um evento extremo está em curso e vidas podem estar em jogo; assim, respostas rápidas e efetivas são imprescindíveis. Sistemas de alerta podem auxiliar na mitigação de perdas e seu uso é comum em vários países (PAGANO *et al.*, 2014). Parte dos sistemas de alerta, previsões hidrológicas servem para antever a ocorrência de um evento danoso, aumentando o tempo de reação disponível a ele.

São exemplos de abordagens para modelos de previsão hidrológicas: regressões múltiplas lineares (ECKHARDT, 2008; FINCK, 2018); modelos autorregressivos (VALIPOUR *et al.*, 2013); modelos conceituais chuva-vazão, como o MGB-IPH (SIQUEIRA, 2015a; GUIMARÃES, 2018). Alternativamente, as redes neurais artificiais (RNAs) apresentam benefícios devido ao ótimo desempenho e a simples implementação, uma vez desenvolvidas (DORNELLES, 2007; BRAVO *et al.*, 2008; PEDROLLO e PEDROLLO, 2013; MATOS *et al.*, 2014; PEDROLLO, 2017). Atribuído às suas características inerentes de aprendizado, classificação e generalização de informação, o ponto forte das RNAs é reconhecido na literatura científica e vem sendo aperfeiçoado há mais de 3 décadas (ZHANG *et al.*, 1998; ASCE, 2000a-b; OLIVEIRA *et al.*, 2015c; PEDROLLO, 2017).

As previsões hidrológicas mais recentes com vistas a Sistemas de Alerta de Eventos Críticos (SACE) estimam o estado futuro em localidades críticas a partir da condição hidrometeorológica recente e/ou atual (inferida por sensoriamento remoto, dados telemétricos *in situ* de níveis nos rios, precipitações, temperatura, umidade do solo e uso do solo). A emissão de bons boletins de previsão de um SACE está relacionada a uma instalação de estações bem planejada, à operação dessas, e à frequente manutenção da rede de monitoramento telemétricas. No Brasil, o Serviço Geológico Brasileiro instala sistemas de alerta desde 1989, operando hoje em 16 bacias (CPRM/SGB, 2020). Um desses sistemas é o SACE-Taquari, na bacia do rio Taquari-Antas, Rio Grande do Sul, Brasil.

*Atualizado a cada três meses, o banco de dados da Universidade de Louvain contempla desastres que cumprem no mínimo um dos quatro critérios: 10 ou mais pessoas mortas, 100 ou mais pessoas afetadas, declaração de estado de emergência, ou pedida por ajuda internacional. A perda calculada refere-se ao total de danos a propriedades, culturas agrícolas e pecuária, sendo que para cada desastre o valor do dano é referente ao momento do ano, *i.e.* os valores são verdadeiros para o ano do evento, são perdas diretas (emdat.be/explanatory-notes) [Dados disponíveis nos períodos de 1964 – 1967, 1969 – 1971, 1973 – 1975, 1983 – 1985, 1987 – 1988, 1991 – 1992, 1995, 1998, 2001, 2003 – 2004, 2007 – 2018]

1.2. Motivação

Há uma região crítica na bacia do rio Taquari-Antas que sofre com inundações frequentes entre Encantado e Bom Retiro do Sul (DRH/SEMA, 2011). Com previsões baseadas em um modelo conceitual chuva-vazão, um sistema de alerta nesta bacia vem sendo operado desde 2014 pela CPRM/SGB — Companhia de Pesquisa em Recursos Minerais / Serviço Geológico Brasileiro (CHAGAS *et al.*, 2014). Esta pesquisa pretende contribuir positivamente para o atual sistema de alerta, podendo implicar num benefício econômico e social.

Nesta pesquisa, as RNAs são opções lógicas, candidatas naturais, uma vez que são exemplos de aproximadores universais de funções conforme o teorema de Hornik (1991). Em outras palavras, elas apresentam acuidade na descrição de processos não-lineares, como o processo chuva-vazão. Além disso, segundo Hsu *et al.* (1995) e Nourani (2017), pesquisadores vêm optando por RNAs quando (1) a qualidade da previsão é mais importante do que a descrição física subjacente, (2) há longos registros históricos, e (3) a modelagem física de um processo não-linear implicaria em excessiva complexidade dos modelos físicos.

RNAs são capazes de aprendizado, podem generalizar o aprendido. Após serem treinadas (ajustadas) para um conjunto de dados disponíveis representativos, podem aproximar a descrição de eventos jamais vistos (dado que estejam entre os extremos dos dados de treinamento). Assim, relações bastante complexas podem ser descritas, como, por exemplo, associar estados atuais e futuros de um sistema. Isso as torna particularmente úteis para previsões (ZHANG *et al.*, 1998). Quanto maior a qualidade de uma previsão, mais valiosa a informação para direcionar ações corretivas.

Outra razão que motiva essa pesquisa é que em bacia vizinha, aquela do rio Caí, Pedrollo e Pedrollo (2013) desenvolveram com sucesso um modelo de previsão com RNAs para o SACE-Caí, utilizando dados de cota e precipitação. Além disso, para o município de Estrela, na bacia do rio Taquari-Antas, recentemente foi realizado um estudo com resultados preliminares satisfatórios para previsão com RNAs utilizando dados convencionais (7 h e 17 h) (PEDROLLO, 2017).

Em trabalho de Siqueira *et al.* (2015a), os autores comentam que os hidrogramas de cheia na bacia do Taquari-Antas podem ocorrer de maneira rápida e por isso pode não ser adequado representar esses eventos em intervalo de tempo diário. Corroborando, Kurek (2016) estimou em 6 horas a propagação da onda de cheia entre Encantado e Estrela. Trata-se, portanto, de uma bacia com respostas sub-diárias.

Na bacia do Taquari-Antas, há dados telemétricos horários de precipitação e níveis coletados pela CPRM/SGB no período 26/11/2015–24/04/2019. Esses dados jamais foram utilizados para modelagem com RNAs no contexto de previsão. Qual será o desempenho em se utilizar exclusivamente esses dados para previsão de níveis no rio Taquari-Antas com RNAs? Além disso, será que é aplicável o uso de dados horários para descrever eventos de inundação na bacia?

Às vezes filtros são utilizados para pré-processar informações presentes em dados antes de fornecer às RNAs. Os filtros agregam informações no tempo ou espaço. Um novo filtro para agregação temporal de informações, chamado Filtro de Média Móvel Gama, é aplicado. Será que o FMMG é vantajoso poderia apresentar vantagens em relação a outros presentes na literatura científica?

Não raramente algumas estações do sistema de alerta deixam de transmitir dados. Às vezes isso ocorre durante eventos críticos, eventos de inundação. Um modelo de previsão resiliente deveria poder lidar com esses acontecimentos. Adicionalmente, uma proposta, chamada Árvore de RNAs, e os meios para a empregar são contemplados. É possível realizar previsões em todos os momentos utilizando as Árvores de RNAs?

1.3. Objetivo

É objetivo determinar a aplicabilidade de RNAs para previsão em tempo atual de níveis fluviais com vistas ao alerta de cheias para os postos do sistema de alerta da bacia do rio Taquari-Antas/RS (SACE-Taquari), com uso exclusivo de dados da rede telemétrica existente.

2. REVISÃO BIBLIOGRÁFICA

A revisão está organizada em sete partes: (2.1) descreve-se o contexto de inserção da pesquisa, isto é, sua problemática, origem e possíveis medidas mitigadoras; (2.2) apresenta-se conceitualmente um sistema de alerta de eventos críticos; (2.3) define-se e discute-se a previsão; (2.4) conceituam-se modelos hidrológicos, apresentando uma classificação; (2.5) apresenta-se a modelagem com redes neurais artificiais; (2.6) descrevem-se algumas métricas para avaliação de modelos; (2.7) apresentam-se exemplos de sua aplicação, trazendo comparações feitas por outros autores, e, por fim, descrevem-se trabalhos recentes de previsão hidrológica na bacia do Taquari-Antas.

2.1. Ocupação de áreas suscetíveis no Brasil

Para tratar desse tema, certas definições são bem-vindas. A terminologia utilizada pelo escritório de Estratégia Internacional das Nações Unidas para Redução de Desastres (UNDRR, 2020) é um guia para universalização de conceitos importantes do campo da redução de risco de desastres. Os seguintes são particularmente importantes nessa pesquisa: (1) desastre; (2) perda econômica (direta ou indireta); (3) perigo, ou ameaça – são as traduções usuais utilizadas no Brasil para “*hazard*” (KOBAYAMA *et al.*, 2006) – e risco; e (4) suscetibilidade e vulnerabilidade.

(1) Desastre é uma perturbação séria na atividade de uma comunidade devido à interação entre um evento perigoso e condições de exposição e vulnerabilidade, que culmina em perdas ou impactos humanos, materiais, econômicos ou ambientais (UNDRR, 2020).

(2) A perda econômica é classificada em direta ou indireta. Mais simples de ser medida, a primeira é a estimativa, quase instantânea, do valor monetário para recuperação do patrimônio físico perdido ou destruído. Isto é, não leva em conta flutuações econômicas no valor da recuperação devido a efeito indiretos. A segunda é o declínio no valor econômico, consequente da perda econômica direta e do impacto socioambiental. Essa pode ocorrer dentro e fora da região, envolvendo um certo atraso, e ocorre em diferentes escalas: impactos microeconômicos (declínio na receita devido a perturbações do regime de trabalho), impactos mesoeconômicos (declínio na receita devido ao desemprego temporário ou a perturbações nas cadeias de produção e fornecimento; vias interdidas) e impactos macroeconômicos (aumento dos preços, aumento da dívida pública, impacto no mercado de ações e no PIB). O impacto econômico é soma da perda econômica direta e indireta (UNDRR, 2020).

(3) Os termos perigo e risco são frequentemente utilizados como sinônimos, mas eles não são. Há perigo quando é sabido que um fenômeno natural, ou uma atividade humana, pode causar danos (KOBAYAMA *et al.*, 2006). Risco é a probabilidade de perda esperada para uma área antropizada, devido à presença iminente de um perigo. O risco é função do perigo, da exposição, da vulnerabilidade e, também, da capacidade de resistir ao fenômeno, e, por isso, pode-se falar em manejar o risco.

(4) Suscetibilidade é a propensão, ou predisposição natural, de ocorrência de um evento adverso. Vulnerabilidade é a condição de um indivíduo, uma comunidade, ou um patrimônio estar sujeito a impactos de perigos. Essa é determinada por fatores físicos, socioeconômicos e ambientais (UNDRR, 2020).

A descrição hipotética de Ogura e Macedo (2002) é valiosa e didática: um fenômeno extremo, como uma inundação, que costuma ocorrer numa região (susceptibilidade) e época conhecida, pode propiciar uma situação de perigo. Esse fenômeno será denominado de desastre natural se impactar uma população, que estava vulnerável. Não havendo impactos para população, será considerado um evento natural (OGURA e MACEDO, 2002).

Por que uma população ocupa uma zona de suscetível? Historicamente, habitações instalaram-se à beira de corpos hídricos por razões logísticas e bélicas. O fácil acesso à água para irrigação de lavouras e também o potencial de transporte foram marcantes para a comercialização de bens e a exploração das terras brasileiras (ROBAINA, 2013). Já em períodos posteriores, núcleos urbanos maiores formaram-se devido, entre outros fatores, à crescente oferta de trabalho de fábricas.

Principalmente durante o processo de industrialização acelerado, pós-década de 1960, o crescimento ocorreu de maneira mais intensa e descontrolada. Na década de 1980, instaurou-se uma nova população urbana, a qual, por reflexo de investimentos reduzidos, careceu de infraestrutura (TUCCI *et al.*, 1995). Apesar da existência de políticas públicas mais recentes como a lei nº 6.776 de 1979, cujo Artigo 3º proíbe a ocupação em terrenos alagadiços e sujeitos a inundações onde não haja medidas de controle (BRASIL, 1979), hoje ainda se identifica a ocupação dessas zonas ribeiras.

O planejamento urbano em uma cidade, representado pelo plano diretor (lei), é o eixo direcionador do crescimento urbano, estipulando-se os tipos de ocupações de cada zona (SILVA e ZAIAM, 2004). Um uso para zonas inundáveis é a recreação, como o futebol de várzea: usa-se a planície de inundação para jogar futebol. Apesar da existência desse instrumento legal, existem ainda zonas suscetíveis ocupadas em diversas cidades. De fato, no contexto contemporâneo, Robaina (2013) descreve a existência de duas cidades: a legal, pautada por preceitos urbanísticos, como o saneamento básico; e a ilegal, cujos lotes são de menor valor e costumam abrigar classes sociais de baixa renda. Nem sempre é a população de baixa renda que ocupa zonas de risco, como planícies de inundação. Certos bairros nobres de cidades históricas continuam à beira de rios, e a opção escolhida é conviver com enchentes, como em Blumenau/SC ou em Veneza/Itália.

O desenvolvimento econômico descontrolado em bairros pode também dirigir à ocupação de áreas suscetíveis em um município — *e.g.* classes de menor poder aquisitivo sendo forçadas indireta ou diretamente a migrar para zonas de periferias devido ao aumento repentino de impostos sobre o terreno e o preço dos produtos na região que moravam, e a falta de um plano diretor adaptativo (FURTADO, 2014). A esse processo dá-se o nome de gentrificação.

Por outro lado, mesmo a população local esquece do perigo. A frequência irregular de eventos de inundação propicia uma falsa sensação de segurança às vezes, levando à fixação de novas moradias cada vez mais próximas aos rios (ECKHARDT, 2008; GUASSELLI *et al.*, 2013; TAPIA *et al.*, 2014).

Almejando mitigar alguns desses riscos associados a essa realidade ocupacional, são adotadas medidas de controle. Essas podem ser mais custosas, de caráter estrutural ao alterar o meio — *e.g.* barramentos e diques —, ou menos custosas, de caráter não-estrutural ao manter o corpo hídrico — *e.g.* zoneamento, construções resilientes, palafitas, e sistemas de alerta (PEDROLLO *et al.*, 2011).

As medidas estruturais podem implicar em uma falsa sensação de proteção aos beneficiados, e incentivar indiretamente a densificação da ocupação da zona de risco. As medidas não-estruturais não são projetadas para oferecer diretamente proteção, mas para reduzir os prejuízos econômicos e, especialmente, pessoais (PEDROLLO *et al.*, 2011).

Pela perspectiva institucional, o Governo do Estado do Rio Grande do Sul objetiva integrar diferentes entidades que trabalham com o monitoramento, previsão, prevenção e gestão de desastres pelo SEGIRD – Sistema Estadual de Gestão Integrada de Riscos de Desastres. Unir bancos de dados e catalogar estudos já realizados pode auxiliar na prevenção, preparação, mitigação, e em ações mais eficientes de resposta, contribuindo para uma cultura de conscientização de desastres (SEMA, 2019).

Enquanto o desenvolvimento de sistemas de alerta de inundação nacionais e continentais são uma realidade na África, América do Norte e Europa, os sistemas da América do Sul operam até os limites da bacia. O sistema de alerta para o estado do Rio Grande do Sul emite desde 2018 previsões diárias na Sala de Situação no Departamento de Recursos Hídricos do Estado, utilizando um modelo chuva-vazão com entradas de previsão meteorológica do WRF, *Weather Research and Forecasting model* (FAN *et al.*, 2019).

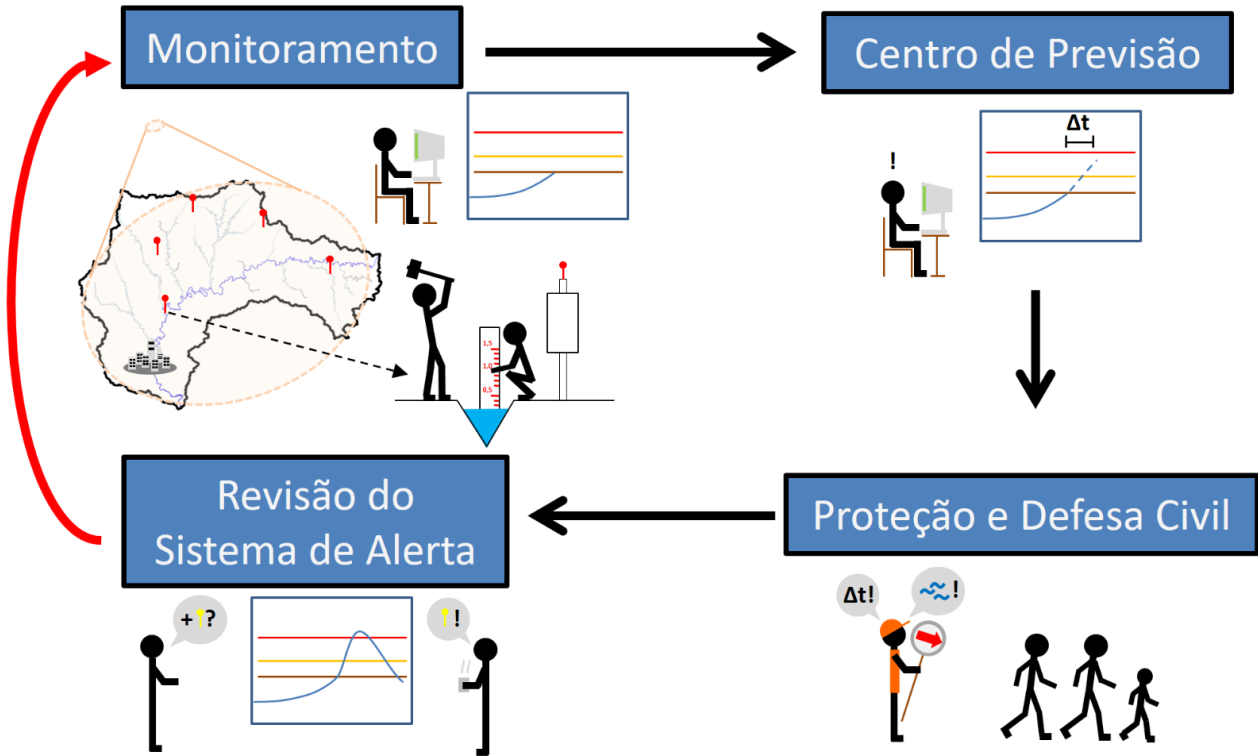
2.2. Sistemas de alerta

Sistemas de alerta objetivam providenciar informações quanto à iminência de algum evento potencialmente danoso em tempo oportuno para que outras medidas de segurança possam ser tomadas. Um sistema de alerta de cheias é essencial para uma tomada de decisão rápida em zonas de risco quanto à evacuação preventiva de bens materiais e de pessoas (PEDROLLO *et al.*, 2011).

Um sistema de alerta de cheias pode ser visualizado em quatro instância (PEDROLLO *et al.*, 2011; GERMANO *et al.*, 2015; SIQUEIRA, 2015b): monitoramento, centro de previsão, proteção e defesa civil e revisão do sistema de alerta (representado na Figura 1). A primeira instância consiste em monitorar o estado da bacia. Se limiares de referência pré-definidos são atingidos, previsões são feitas e, podendo haver a emissão de boletins de previsão. Esses servem à Proteção e Defesa Civil para ações de evacuação (e estimar posterior retorno) de pessoas e bens. A análise do desempenho do sistema como um todo visa à adaptação à novas necessidades do sistema — *e.g.* densificar a área de

controle, alterar o modelo de previsão, ou mesmo refazer a curva-chave, que implica em hidrotécnicos irem a campo para realizar medições (GERMANO *et al.*, 2015). Dificilmente um Sistema de Alerta será dado como pronto, finalizado, jamais revisto.

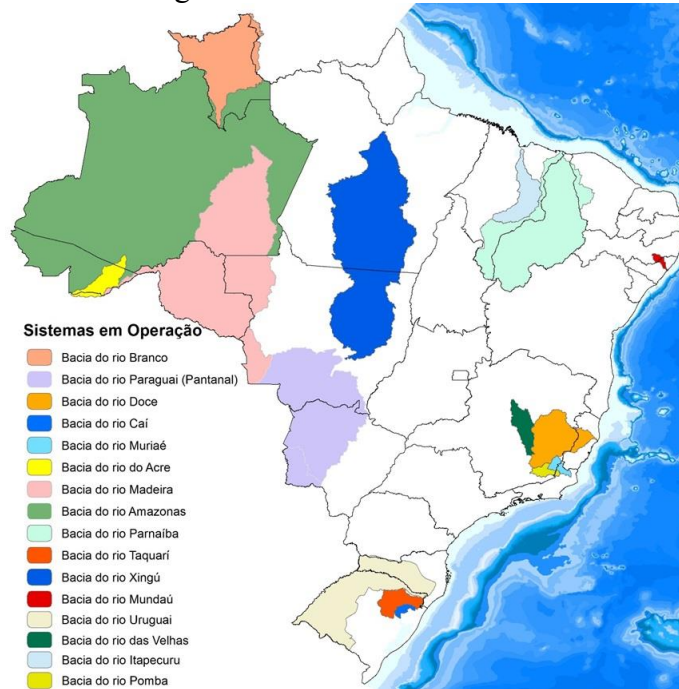
Figura 1. Componente de um sistema de alerta de eventos críticos.



A CPRM/SGB opera seus sistemas segundo limiares de referência. As primeiras previsões são realizadas em tempo atual a partir do momento em que o **nível de atenção** é atingido (representado pela linha horizontal marrom na Figura 1). Nesse momento, os níveis previstos são informados à Proteção e Defesa Civil, e os operadores da rede preparam-se para eventuais manutenções. O **nível de alerta** (linha horizontal amarela na Figura 1) indica grande risco de inundação e a frequência de emissão dos boletins de previsão aumenta. O **nível de inundação** (vermelho na Figura 1) indica que haverá prejuízos humanos ou materiais (SIQUEIRA, 2015b; GERMANO e MATOS, 2017; FINCK, 2018).

Como exemplo de sistemas de alertas, pode-se citar aqueles que vêm sendo instalados e operados pela Companhia de Pesquisa de Recursos Minerais (CPRM/SGB). Pode-se obter mais informações sobre esses sistemas na página do SACE – Sistema de Alerta de Eventos Críticos. Já foram instalados dezesseis sistemas, como apresentado na Figura 2.

Figura 2. SACE CPRM/SGB.



Fonte: CPRM (2020).

Chagas *et al.* (2014) descrevem a operação do SACE-Taquari. O monitoramento consiste em acompanhar tanto o estado da rede telemétrica que opera em tempo atual por GPRS (Serviços Gerais de Pacote de Rádio) como também previsões meteorológicas realizadas pelo INMET e o INPE/CPTEC – *e.g.* previsões do ETA (11 dias) 15 x 15 km. Os sistemas telemétricos podem transmitir dados de 15 em 15 minutos. Dados horários de outras instituições — *e.g.* AHSUL, INMET, e barramentos medem nível e vazão a montante e jusante — têm caráter complementar. Os dados são consistidos, checando variações abruptas e persistência longa de valores, comparando dados entre diferentes sensores, e averiguando limites mínimos e máximos. Depois, comparam-se os dados de precipitação com registros históricos de referência entre regime de precipitação antecedente e inundação consequente. Dois exemplos para Muçum são: se precipitem 50 mm nos últimos 3 dias, 70 mm nos últimos 7 dias e 14 dias, então é esperado um pico de cota maior que 1000 cm; se precipitem 90 mm nos últimos 3 dias, 105 mm nos últimos 7 dias e 125 mm nos últimos 14 dias, é esperado um pico de cota maior que 1500 cm. Outro indicativo são os limiares de referência para cotas (atenção, alerta e inundação) que foram definidos para o porto de Estrela. Se o nível de alerta é atingido, a CPRM/SGB contata a Proteção e Defesa Civil e emite boletins com previsões. As de 2014 advinham de modelos de regressão linear múltipla (RLM) com dados de entrada de vazão de estações de montante e com dado de saída sendo a vazão na estação de jusante para a qual se realizava a previsão. Os alcances para Estrela eram feitos com até 6 horas de antecedência.

Mundialmente, agências de previsão contatam usuários e mídia durante eventos, atualizando continuamente sobre condições na bacia (PAGANO *et al.*, 2014). Semelhantemente, a CPRM/SGB

emite boletins. Há mais de 60 boletins no portal do SACE-Taquari de 27/03/2016 até 03/10/2018 (CPRM, 2019).

Além dos SACEs da CPRM/SGB, a Sala de Situação/RS, vinculada à SEMA-RS, está inserida no futuro SEGIRD, monitora e prevê as condições meteorológicas no estado, utilizando uma rede de monitoramento em solo, radares meteorológicos e imagens de satélite. A Sala de Situação elabora boletins e emite avisos quanto à ocorrência de tempo severo e extremos hidrológicos para a Proteção e Defesa Civil (SEMA, 2019).

Previsões são importantes para gerir os recursos hídricos e para as ações emergenciais de proteção. Pedrollo e Pedrollo (2013) e Pagano *et al.* (2014) relatam dificuldades enfrentadas no desenvolvimento e operação desses sistemas, relatando também avanços desejáveis na área da previsão hidrológica. O SACE do rio Caí conta com previsões de níveis por RNAs (PEDROLLO e PEDROLLO, 2013). Neste trabalho, três aspectos desejáveis aos modelos desenvolvidos são mencionados: (1) previsões associadas à incerteza; (2) modelos adaptativos tanto as mudanças na bacia quanto modificações na densidade da rede de aquisição (3) resiliência frente à degradação da rede de aquisição de dados. Trabalhadores de agências de previsão de eventos extremos de 24 países desenvolvidos e em desenvolvimento relataram quatro principais dificuldades que são enfrentadas: (1) obter o máximo dos dados disponíveis (lidar com incertezas e informações incompletas); (2) realizar previsões precisas utilizando modelos; (3) transformar previsões hidrometeorológicas em avisos eficazes; (4) gerir o serviço operacional (PAGANO *et al.*, 2014).

Não é incomum que postos de monitoramento danifiquem durante eventos extremos, justo quando a aquisição contínua de dados é mais necessária para acompanhar a subida e a recessão dos níveis. Interpolar dados de precipitação para compor uma entrada única de chuva média na bacia pode ser uma abordagem, aos moldes da vazão simulada no posto fluviométrico de Santo Ângelo, na bacia do rio Ijuí/RS em Oliveira *et al.* (2014). Alternativamente, ou estudo utilizou um modelo completo, CM — utilizando todas as informações relevantes disponíveis — e uma série de modelos parciais — PM para cenários em que ocorra degradação do sistema de aquisição de dados (PEDROLLO, 2017).

2.3. Previsão

A previsão, prática de estimar qual será o estado futuro de um sistema, é uma instância de um SACE. Em muitos trabalhos, as previsões são classificadas em previsões de curto prazo — algumas horas até alguns dias — ou de longo prazo — várias semanas ou mesmo meses (TUCCI, 2005; PEDROLLO, 2017).

Esta variabilidade na definição das classes de previsão hidrológica é consequência das diferentes escalas de bacias para as quais se faz previsão. Em uma bacia grande — *e.g.* Uruguai

(FINCK, 2018) ou Paraná — o nível varia mais vagarosamente do que em bacias com tempo de concentração menor. Assim, técnicas de previsão mais simples podem ser suficientes.

As previsões hidrológicas de longo prazo costumam estar associadas à estimativa da demanda pelo recurso hídrico ou à ocorrência de estiagens. Essa pode ser realizada ao investigar a correlação do fenômeno climático ENOS (*El Niño/Oscilações do Sul*) e séries de vazões afluentes a reservatórios (CHIEW *et al.*, 2003; KARAMOUZ *et al.*, 2008; JAFARZADEGAN *et al.*, 2014; PAGANO *et al.*, 2014; WANG *et al.*, 2019). Já as previsões hidrológicas de curto prazo estão geralmente relacionadas aos alertas de inundação (PAGANO *et al.*, 2014; SIQUEIRA, 2015b; PEDROLLO, 2017).

Previsão em tempo atual, *real-time forecasting*, refere-se a prever com os dados disponíveis mais recentes. Alguns autores (FAN, 2015; SIQUEIRA, 2015b) adotam “tempo real” para o mesmo tipo de previsão. Citam-se alguns exemplos de trabalhos no contexto de previsão em tempo atual:

Sistemas hidrológicos mudam constantemente e seu estado é fonte de incerteza para modelos. Existem métodos para atualizar modelos a essas alterações. Esses mecanismos adicionam um viés a previsão, baseado no erro recente entre as últimas realizações observadas e a previsão que havia sido feita. Liu *et al.* (2016) testaram três mecanismos de atualização (AR, autorregressivo; RLS, mínimos quadrados recursivos; e HUP, *hydrologic uncertainty processor*) para ajustar os dados de saída do modelo conceitual chuva-vazão Xinanjiang na bacia do reservatório de Baiyunshan (464 km²), China. Esse modelo prevê com uma, duas e três horas de antecedência. As métricas para três horas de antecedência melhoraram, com o coeficiente de determinação (r^2) e o erro relativo de 0,85 e 5,61% (sem atualização) para 0,90 e 5,17% (AR); 0,91 e 4,80% (RLS); e 0,91 e 3,73% (HUP).

Matos *et al.* (2014) referem-se à previsão em tempo atual como uma atualização contínua das variáveis de estado do sistema para prever o estado futuro. O efeito da área de controle em bacias embutidas foi avaliado para a previsão, utilizando dados de precipitação de uma grade (gerada com 65 postos da região) e séries de vazões para seis estações. O período estudado foi o de 22/08/1989 até 01/06/1994. Foi desenvolvida uma RNA com seis entradas de vazão (estação Ponte Mística e cinco outras a montante), uma entrada de precipitação média, e a vazão em Ponte Mística um dia à frente como saída. Concluiu-se maior a área controlada, melhores os resultados para previsão.

O tempo de propagação da onda de cheia entre dois postos fluviométricos afeta o maior alcance possível par previsão, quando os volumes propagados são relevantes. Indicativos desse tempo pode ser obtido com o coeficiente de correlação de Pearson entre as séries históricas de estações fluviométricas a jusante e a montante, considerando atrasos temporais de diferentes horas (chamados defasagens) como em Pedrollo *et al.* (2014). A defasagem de maior correlação corresponde ao tempo de deslocamento entre os postos, e, assim, teoricamente, ao alcance máximo da previsão com os postos fluviométricos.

É notável a importância da previsão hidrológica não só no contexto dos sistemas de alerta, mas também para a gestão e o planejamento de recursos hídricos, geralmente fazendo uso de previsões de longo prazo (CHIEW *et al.*, 2003; WANG *et al.*, 2019). Prever o regime geral de vazões afluentes durante a estação de irrigação para um reservatório pode permitir decisões da gestão hídrica melhor embasadas no risco (CHIEW *et al.*, 2003). Em Wang *et al.* (2019) é feita a previsão da vazão média do mês vindouro com modelos empíricos, indicando vantagens em processar as séries de dados antes de aplicar métodos de inteligência artificial, como RNAs. Processar é buscar atributos, ou índices, na série, podendo ser feito por análise espectral da série. Analogamente, o uso de filtros de média móvel, como em Bravo *et al.* (2008), também é um processamento das séries de dados.

No Brasil há diversos órgãos que realizam previsões e que possuem sua própria rede de monitoramento de dados. Segundo Fan *et al.* (2016), isso é decorrência principalmente da ausência de um sistema de previsão continental utilizado por diferentes agências, como é o caso do EFAS que é utilizado na Europa. Integrar as previsões e as redes de monitoramento pode não só complementar modelos já existentes, como também possibilitar previsões em locais onde ela ainda não é feita.

2.4. Modelos hidrológicos

Modelos tentam representar o funcionamento de um sistema. Há uma citação geralmente atribuída a Box (1976) que diz que “todos os modelos estão errados, mas alguns são úteis”. A utilidade está em auxiliar na tomada de decisão.

Os processos do ciclo hidrológico apresentam variabilidade espacial e temporal. Alguns desses processos podem ser abordados satisfatoriamente de maneira determinística. Outros requerem estatística para lidar com incertezas, devida tanto ao baixo grau de disponibilidade de informação quanto à limitada capacidade humana em mensurar o estado exato do material de estudo. Uma maneira de classificar modelos hidrológicos é aquela proposta por Chow *et al.* (1988). Os pesquisadores sugerem o uso de três fatores para descrever o grau de simplificação de um modelo:

- (1) há consideração da aleatoriedade? (determinístico/estocástico);
- (2) há consideração da variabilidade espacial? (concentrado/distribuído);
- (3) há consideração da variabilidade temporal?

Além dessa classificação, Tucci (2005) também difere modelos de base física e empírica. Assim, adicionando outro fator: (4) há consideração de processos físicos? O Quadro 1 une essas diretrizes de classificação.

Quadro 1. Classificação de modelos hidrológicos

Fator	Grau de consideração		
	Pouco		Bastante
Aleatoriedade	Determinísticos		Estocásticos
Variabilidade Espacial	Concentrado	Semidistribuído	Distribuído
Variabilidade Temporal	Independente do tempo		Correlacionado com tempo
Processos físicos	Empírico	Semiconceitual	Conceitual (ou físico)

Fonte: adaptado de Chow *et al.* (1988) e Tucci (2005).

Modelos determinísticos retornam a mesma saída para uma mesma entrada. Modelos estocásticos retornam uma distribuição de saídas possíveis para um mesmo conjunto de entrada, podendo-se estimar a probabilidade de ocorrência de eventos por faixas (TUCCI, 2005). Utilizar a estatística é uma maneira de lidar com incertezas, porém não é a única (ZADEH, 1965). Tucci (2005) descreve o caos determinístico: às vezes, há tantas equações descrevendo um sistema que um pequeno abalo do estado inicial pode desencadear grandes mudanças. Edward Lorenz foi um meteorologista que trabalhou com previsão meteorológica e caos. Dois estados iniciais pouco diferentes com um pequeno “erro observacional” podem evoluir para estados futuros tão diferentes quanto se fossem sido selecionados ao acaso entre possíveis estados do sistema (LORENZ, 1969).

Modelos distribuídos são aqueles que consideram o espaço uma dimensão contínua, como em grade. Os concentrados representam um espaço fechado com uma variável única representativa. Há casos de modelos semidistribuído, como o MGB-IPH, que particiona o espaço em subespaços concentrados e concatenados, as minibacias (COLLISCHONN, 2001).

Em um modelo independente do tempo, representa-se uma sequência de eventos hidrológicos que não interferem um com o outro. Já um modelo correlacionado com o tempo representa uma sequência em que o próximo evento é parcialmente influenciado pelo evento atual e/ou possivelmente também por outros eventos anteriores (TUCCI, 2005).

Ao passo que modelos conceituais utilizam somente equações físicas, modelos semiconceituais utilizam também equações empíricas, como as leis de Darcy (DARCY, 1856) e de Horton (HORTON, 1933). Modelos empíricos, por sua vez, relacionam um conjunto de variáveis de entrada, com um conjunto de variáveis de saída, sem preconceber explicitamente uma relação física, e são por isso nomeados também de “modelos caixa-preta”. Modelos conceituais costumam ser mais robustos e necessitar de uma descrição mais fidedigna e espacialmente variada da bacia hidrográfica, possibilitando maior capacidade de extrapolação para eventos não registrados, enquanto que aos modelos empíricos não se aconselha extrapolação (LANNA e INDURSKY, 1983). Por outro lado, modelos empíricos, como os de RNAs, têm o potencial de expor novas relações entre dados que até

então não são conhecidas pela modelagem conceitual (OLIVEIRA *et al.*, 2014). São abordagens diferentes, com diferentes qualidades.

Uma das dificuldades de modelagem físico-conceitual emerge ao aplicar equações que foram concebidas para pequena escala — *e.g.*, a dinâmica da coluna de infiltração do solo, medível na ordem de 0,1 m — para simular o comportamento de um rio em uma bacia hidrográfica — *e.g.* da ordem de 1-100 km (PAGANO *et al.*, 2014). Lidar com essas diferenças de escala é frequentemente visto como um “grande impasse” (BEVEN, 2007).

Além dessa dificuldade relacionada à escala, mesmo o melhor modelo disponível para descrever o sistema natural pode encontrar obstáculos na previsão fluvial. Humanos interferem no ciclo hidrológico com barramentos, irrigações, drenagem subsuperficial, transferências interbacias e bombeamentos de água subterrânea (PAGANO *et al.*, 2014). Na prática, essas interferências costumam ser contabilizadas no ajuste de parâmetros dos modelos conceituais, resultando, às vezes, em uma descrição certa do comportamento do sistema, mas baseada em parâmetros físicos errados (ou irreais). Embora alguns pesquisadores relutem usar a abordagem empírica por acreditar que modelos físico-conceituais, ao apresentar bons resultados, estão “acertando por motivos certos” (equações físicas), não necessariamente os coeficientes físicos ajustados a partir da série de dados são aqueles do sistema real (KIRCHNER, 2006).

Diversos trabalhos comentam sobre a necessidade de extensa série de dados e de atentar aos limites dos eventos utilizados durante o desenvolvimento do modelo empírico (BRUN e TUCCI, 2001; DORNELLES, 2007; SIQUEIRA, 2015b). No contexto do curto prazo, quando a bacia não é grande, modelos empíricos podem ser preferíveis (LANNA e INDURSKY, 1983).

RNAs são modelos empíricos por não representar *a priori* de maneira explícita relações físicas entre as variáveis. Justamente por não estarem sujeitas a uma relação fixa de equações hidrológicas específicas, elas podem reafirmar o que é cientificamente sabido e indicar espaço para avanço científico ao que já se sabe dos processos físicos. Além disso, existem técnicas para “abrir a caixa-preta” e realizar sua interpretação (JAIN e KUMAR, 2009).

Alguns princípios importantes da prática da modelagem são apresentados em Box (1976). Um cientista não deve se apaixonar por seu modelo. O conhecimento nasce de teorias que precisam ser reformuladas frente a novas discrepâncias factuais encontradas. Geralmente atribuído a Occam, filósofo do séc. XIV, um dos princípios, conhecido como a “navalha de Occam”, que guiam a modelagem é a parcimônia: já que nenhum modelo pode estar perfeitamente correto, é de se buscar sempre o modelo mais econômico. Box (1976) sugere que se deve desenvolver modelos simples e valiosos, enquanto o excesso de parametrizações e a superelaboração devem ser evitadas. Outro princípio é identificar e diferenciar grandes e pequenos erros (BOX, 1976).

2.5. Redes neurais artificiais

Redes neurais artificiais (RNA) são estruturas de processamento paralelo e distribuído de informação, compostas por nós de entrada, elementos de processamento e nós de saída, os quais são interconectados por conexões na forma de um grafo direcionado (HECHT-NIELSEN, 1990). Com intuito de apresentar a origem, o desenvolvimento e algumas peculiaridades das redes neurais artificiais do tipo *Perceptron* de Múltiplas Camadas (RNA MLP) utilizadas neste trabalho, apresentam-se os itens:

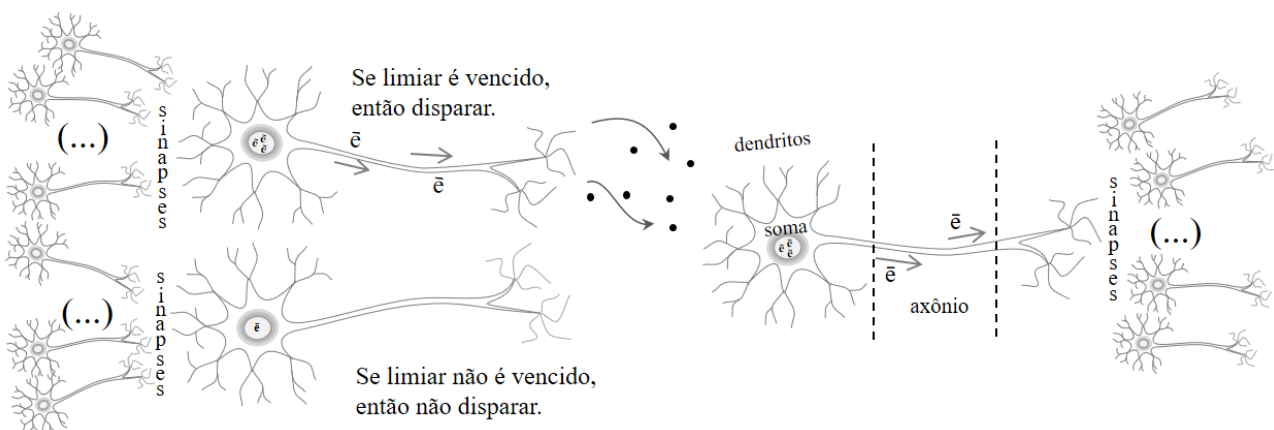
- Histórico
- Aproximador Universal
- Treinamento, Validação e Verificação
- Filtros

Salienta-se que a área das RNAs é vasta e arquiteturas além da MLP não são abordadas aqui. Outras obras que podem ser consultadas para tal: Haykin (1994) e Braga *et al.* (2007).

2.5.1. Histórico

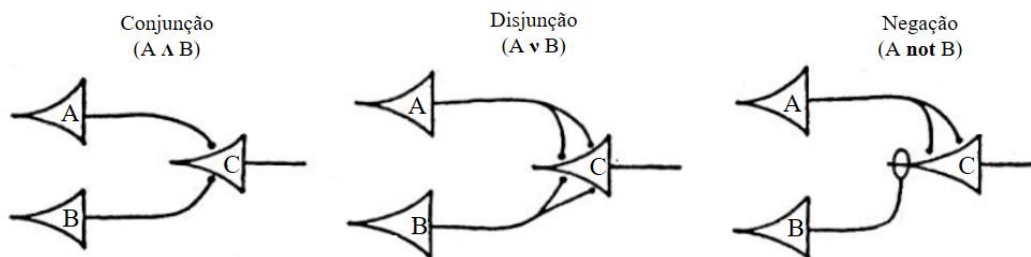
O início dos estudos das RNAs é atribuído ao psicólogo e neurofisiologista Warren McCulloch e ao matemático Walter Pitts (1943). Segundo as bases da neurofisiologia da época, o sistema nervoso seria constituído por uma rede de blocos fundamentais, chamados de neurônios, cada qual possuindo dendritos, soma e axônio. O sistema nervoso trataria de interpretar e responder aos estímulos chegados do mundo externo ao sistema, o que ocorreria através da transmissão de pulsos elétricos. Esses pulsos chegariam a um neurônio pelos dendritos, e seriam concentrados na soma. Se durante o período de adição latente ($<0,25$ ms) se instaurasse uma diferença de potencial maior que um determinado limiar, particular de cada neurônio, então o pulso elétrico seria disparado adiante. O pulso elétrico seria, então, disparado da soma, para o axônio, atravessaria a sinapse, onde encontraria os dendritos do próximo neurônio, como retratado na Figura 3.

Figura 3. Rede neural biológica e alguns componentes.



Identificou-se uma característica binária associada ao funcionamento dos neurônios: ou o neurônio dispara, ou não dispara. Devido a esse “tudo ou nada”, McCulloch e Pitts (1943) propuseram que se poderia tratar a atividade dos neurônios biológicos, em termos práticos, por lógica proposicional (negação, disjunção, conjunção – Figura 4). Cada espécie de proposição deferiria na natureza das conexões entre os neurônios. Outra contribuição importante que seguiu é que diferentes redes poderiam reproduzir o mesmo padrão de proposição, porém algumas demorariam mais que outras.

Figura 4. Neurônios e proposições lógicas: conjunção, disjunção e negação.



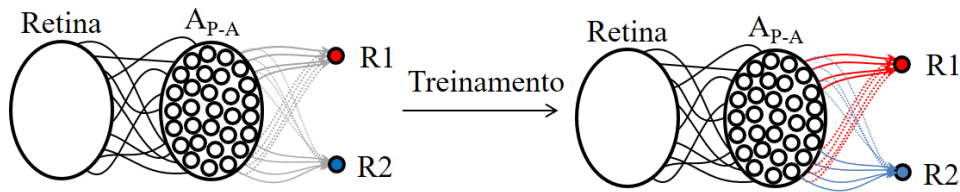
Fonte: adaptado de McCulloch e Pitts (1943).

Facilitação-inibição e aprendizado são alterações que ocorrem em redes neurais biológicas e foram duas dificuldades que emergiram durante a formulação desse modelo artificial pioneiro. Ambas são fenômenos de modificações da rede devido à atividade passada. A primeira é governada por potenciais residuais e concentrações iônicas, e a segunda por mudanças duradouras dos limiares biológicos. Os autores apontam que essas podem ser representadas pela substituição fictícia, porém equivalente, com mais redes de neurônios, mantendo ainda assim os limiares não alterados. Posteriormente, essas questões sobre o aprendizado foram aprofundadas por outros trabalhos.

Em *The Organization of Behaviour* (HEBB, 1949), é proposto que o aprendizado de redes neurais biológicas ocorre através de reestruturação celular de sinapses, ou, em outras palavras, através da alteração das conexões entre neurônios. Assim, um estímulo outrora inadequado torna-se adequado. Hebb cunhou o conceito de plasticidade sináptica. Essa teoria veio a ser conhecida como regra de Hebb.

Como organismos superiores tem a capacidade de reconhecimento perceptivo, generalização, relembrar e pensar? Com esse questionamento, Rosenblatt (1958) introduz o *perceptron*. Inspirado no olho humano, o autor apresenta o *perceptron* óptico, composto pelas seguintes etapas: retina, área de projeção-associação (A_{P-A}) e resposta (R1 e R2) (Figura 5).

Figura 5. O perceptron óptico.



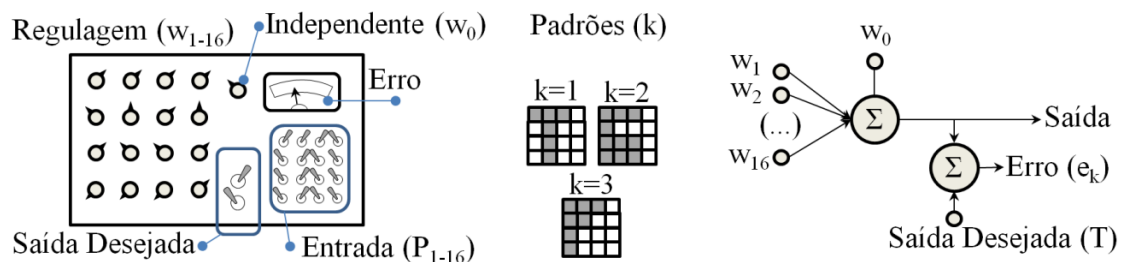
Fonte: adaptado de Rosenblatt (1958).

Um dos exemplos de regras de treinamento descritas é aquele para sistemas univalentes: as redes são criadas aleatoriamente (linhas cinzas), e cada estímulo exterior captado pela retina provoca aos neurônios ativos reforços positivos em caminhos similares (linhas contínuas) e reforços negativos em caminhos não utilizados (linhas pontilhadas). Probabilisticamente, após estímulos suficientes, a estrutura do *perceptron* converge para caminhos que levam marcadamente para R1 ou R2. No jargão, diz-se que o *perceptron* é capaz de aprendizado heurístico. Assim, o autor descreve o potencial de formação espontânea de conceito (mais tarde nomeado de treinamento não-supervisionado).

Por fim é tratado o caso dos sistemas bivalentes: independente de ativo ou não ativo, cada caminho pode ser reforçado positivamente ou negativamente segundo um agente externo. Esse sistema é semelhante ao que seria mais tarde denominado de treinamento supervisionado, que é utilizado na presente pesquisa. Desta maneira um ser vivo é capaz de aprender conceitos.

Até então as funções de ativação que descreviam o comportamento dos neurônios eram discretas: ou o neurônio estava ativo, ou inativo. Utilizando funções de ativação contínuas e lineares, Widrow e Hoff (1960) desenvolveram o ADALINE, máquina de Adaptação Linear, um dispositivo com um termo independente e 16 entradas reguláveis (nós de entrada), as quais, ao serem expostas a um padrão k (estímulo), retornam uma saída com um erro associado, como retratado na Figura 6.

Figura 6. A máquina ADALINE.



Fonte: adaptado de Widrow e Hoff (1960).

Para o dispositivo funcionar com eficácia, é necessário que se regulem os pesos de entrada (w_i). Partindo da equação do erro quadrático em função dos pesos dos nós de entrada e do termo independente,

$$e_k^2 = \left(T - f \left(w_{0,k} - \sum_{i=1}^{16} (w_{i,k} * P_{i,k}) \right) \right)^2 \quad (1)$$

Em que:

- e_k^2 – Erro médio quadrático associado ao padrão k
- T – Reposta-objetivo, *target*
- w_0 – Peso sináptico do termo independente, *weight*
- $w_{i,k}$ – Peso sináptico da entrada i, *weight*
- $P_{i,k}$ – Sinal da entrada i, aplicando o padrão k, *pattern*
- f – Função de ativação utilizada

os autores diferenciaram parcialmente o erro quadrático em relação a cada peso, aplicando a regra da cadeia,

$$\begin{aligned} \frac{\partial e_k^2}{\partial w_{i,k}} &= \left(\frac{\partial e_k^2}{\partial f} \right) * \left(\frac{\partial f}{\partial (w_{i,k} * P_{i,k})} \right) * \left(\frac{\partial (w_{i,k} * P_{i,k})}{\partial w_{i,k}} \right) \\ \frac{\partial e_k^2}{\partial w_{i,k}} &= (2 * e_k^2) * (f'(w_{i,k} * P_{i,k})) * (P_{i,k}) \end{aligned} \quad (2)$$

O erro quadrático é uma hypersuperfície de grau dois, e os autores apresentaram as derivadas parciais de forma analítica, as quais apontam o caminho de maior variação dessa superfície. Dada a entrada de um padrão k, pode-se estimar qual a mudança que deve ser empregue a cada peso sináptico (w_i) para minimizar o erro quadrático referente a esse padrão k.

$$\Delta w_i = -\tau * e_k^2 * f'(w_{i,k} * P_{i,k}) * P_{i,k} = -\tau * \delta_k * P_{i,k} \quad (3)$$

Em que:

- τ – Taxa de aprendizado a ser definida
- δ_k – Delta associado ao treinamento para o padrão k, ($e_k^2 * f'(w_{i,k} * P_{i,k})$)

Esse método de procura de um ponto de mínimo se chama regra delta, ou regra do gradiente descendente. O algoritmo demonstrou capacidade de aprendizado de padrões, e mesmo com a adição de ruídos os padrões puderam ser generalizados até certo grau. Sobre limitações, um ADALINE único talvez não reconheça uma gravata verde e uma camisa xadrez em fotos, mas dois ADALINE em paralelo podem reconhecer uma gravata verde e/ou uma camisa xadrez (WIDROW e HOFF, 1960).

Minsky e Papert publicaram *Perceptrons* (1969), no qual discutiram rigorosamente os limites dos *perceptrons*. O objetivo de uma rede é mapear padrões de entrada para desejados padrões de saída. Um *perceptron* não pode mapear certos padrões não-lineares (simetria, paridade e conectividade), como o operador lógico de disjunção exclusiva (XOR). A falta de um procedimento simples e genérico para treinar redes de três ou mais camadas, análogo àquele da regra delta para redes sem

camadas internas, instaurou uma situação de congelamento de recursos. Na seção “13.2 *Other Multilayer Machines*”, ao discutir sobre *perceptrons* com múltiplas camadas, Minsky e Papert (1969) afirmaram haver um problema estratégico em usar múltiplas camadas:

O perceptron já se provou valioso apesar de suas severas limitações. Ele tem várias características que atraem a atenção: sua linearidade, seu teorema de aprendizado intrigante; sua simplicidade paradigmática como um tipo de computação paralela. Não há nenhuma razão para supor que qualquer uma dessas virtudes se manteria nas versões de múltiplas camadas. No entanto, nós consideramos um problema de pesquisa importante elucidar (ou rejeitar) nosso julgamento intuitivo de que a ‘extensão seria estéril’. Talvez algum teorema de convergência poderoso será descoberto, ou alguma profunda razão para a falha em produzir um ‘teorema de aprendizado’ interessante para a máquina multicamadas será encontrado (MINSKY e PAPERT, 1969, p. 231).

As redes neurais profundas (3 ou mais camadas) voltaram à tona duas décadas depois, quando procedimentos para seu treinamento foram desenvolvidos. Rumelhardt *et al.* (1986) apresentaram um algoritmo de retropropagação que generaliza a regra delta, retropropagando os erros da última camada para penúltima camada. O erro na penúltima camada é a variação do erro quadrático de saída em função dos estados da penúltima camada, que, matematicamente, é expresso como:

$$e_{int}^2 = \frac{\partial e_{sai}^2}{\partial s_{int}} = \frac{\partial e_{sai}^2}{\partial f} * \frac{\partial f}{\partial (w_{sai} * s_{int})} * \frac{\partial (w_{sai} * s_{int})}{\partial s_{int}} = e_{sai}^2 * f'(w_{sai} * s_{int}) * w_{sai} \quad (4)$$

Os autores comentam que esse método não converge tão rápido quanto aqueles que utilizam derivada de segunda ordem, mas são bem mais simples e de fácil implementação. Também, significativo ganho pode ser obtido adicionando um termo que depende da última alteração feita nos pesos o qual, por sua vez, acelera o treinamento em superfícies de erro muito planas e evitar a oscilação de pesos. Esse termo ficou conhecido como o termo do momento de inércia:

$$\Delta w(t) = -\tau * \delta * P + \alpha * \Delta w(t - 1) \quad (5)$$

Em que:

- $\Delta w(t - 1)$ – Passo aplicado no ciclo anterior
- α – Coeficiente do momento de inércia a ser definido

Vogl *et al.* (1988) demonstram que três modificações podem acelerar muito a taxa de convergência do algoritmo retropropagação de Rumelhardt *et al.* (1986). A primeira trata de atualizar os pesos sinápticos apenas depois que todo conjunto de treinamento é apresentado (por batelada) ao invés de a cada registro apresentado. A segunda consiste em definir regras que variam a taxa de aprendizado dinamicamente segundo o desempenho do ciclo anterior, respondendo melhor a topografia do problema ao evitar mínimos locais (nomeada de taxa heurísticamente variada). A terceira consiste em anular o termo de momento se o último ciclo piorou o desempenho. Anular ele, implica em realizar uma pesquisa mais local, desconsiderando a direção que foi tomada anteriormente que levou a um aumento no erro. O método demonstrado em Vogl *et al.* (1988) pode ser chamado de

retropropagação conjugada, pois utiliza a taxa heurísticamente variada e o termo de momentos apresentado por Rumelhardt *et al.* (1986).

2.5.2. Aproximador universal

A partir do teorema de Kolmogorov (1957), que trata da representação de funções contínuas arbitrárias do cubo n -dimensional para os números reais em termo de funções contínuas unidimensionais, Hecht-Nielsen (1987) apresentou uma reinterpretação para RNAs, que ficou conhecida como “Teorema de Existência”, ou “Teorema de Kolmogorov-Nielsen”. Em livre tradução, segue o enunciado por Hecht-Nielsen (1987): “Dada qualquer função contínua $\Phi : I^n \rightarrow \mathcal{R}^m$, $\Phi(x) = y$, em que I é o intervalo-unidade fechado $[0,1]$ (I^n , portanto, o cubo n -dimensional), Φ pode ser implementada exatamente por uma RNA de três camadas, possuindo n elementos de processamento na camada de entrada (x), $2n+1$ elementos de processamento na camada interna e m elementos de processamento na camada de saída” .

Hornik (1991) apresentou o “Teorema da aproximação universal”, dando um passo em direção à aplicação das RNAs. No teorema é abordada a capacidade de aproximação das RNAs, ao invés de uma representação exata, como no Teorema de Kolmogorov-Nielsen. Em livre tradução segue o enunciado por Hornik (1991): “Dada uma relação mensurável $r: \mathbf{R}^n \rightarrow \mathbf{R}^m$, existe uma RNA de uma única camada interna que pode aproximar a relação com qualquer critério de aproximação $L^p(\mu)$, desde que um número suficiente de neurônios internos estejam disponíveis e adequadamente treinados e suas funções de ativação sejam contínuas, limitadas e não-constantas” .

Uma função de ativação popularmente utilizada é a sigmoide unipolar. Para realizar o treinamento de uma RNA, é preciso calcular a regra delta e retropropagar o erro novamente para cada ciclo. Nestes cálculos se faz necessário utilizar a função de ativação e também sua derivada. Vantajosamente, a derivada da sigmoide unipolar pode ser escrita analiticamente.

$$f(x) = \frac{1}{(1 + e^{-x})}$$

$$f'(x) = f(a) = a * (1 - a) \mid a = \frac{1}{(1 + e^{-x})}$$

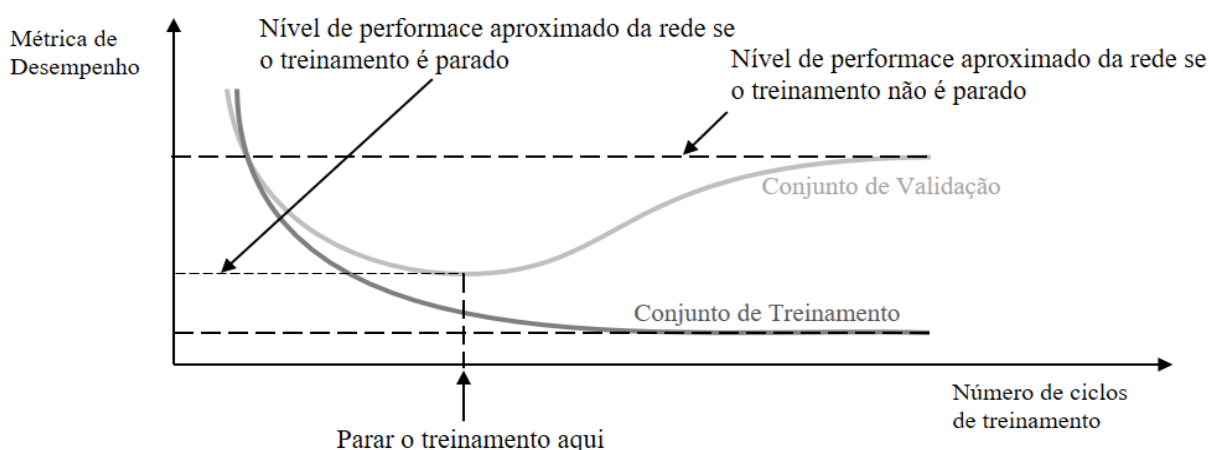
Ser um aproximador universal torna as RNAs valiosas para previsão. Já que RNAs MLP podem mapear qualquer relação de entrada e saída e esse mapeamento depender do bom ajuste dos pesos sinápticos, não estando restrito a parâmetros de equações pré-concebidas da modelagem físico-conceitual, as RNAs MLP podem apresentar uma das quatro características desejadas em agências de previsões em 24 países desenvolvidos e em desenvolvimento: obter o máximo possível dos dados disponíveis (PAGANO *et al.*, 2014). As outras características são descritas no item 4.2.

2.5.3. Treinamento, validação e verificação

Um impasse das RNAs é o sobreajustamento dos dados (*overfitting*). RNAs iniciam o treinamento tipicamente mapeando as entradas para uma superfície plana. Posteriormente, o mapeamento fica mais detalhado, mostrando vales e picos, enquanto ainda mantendo uma boa capacidade de interpolação dentro do domínio. Treinamento excessivo conduz a mapeamentos exatos e perda da capacidade de generalização, comprometendo a aplicação do modelo para outros valores não usados no treinamento, mesmo que dentro do domínio desse (HECHT-NIELSEN, 1990).

O método da validação cruzada pode contornar esse efeito indesejado (HECHT-NIELSEN, 1990), dividindo os dados em três conjuntos: treinamento, validação e verificação. O treinamento é realizado com o primeiro conjunto, e, concomitantemente, testa-se a rede a cada ciclo com o conjunto de validação. Se a métrica de desempenho deixa evidentemente de melhorar para o conjunto de validação, o treinamento é interrompido (Figura 7).

Figura 7. Sobreajustamento de uma rede.



Fonte: adaptado de Hecht-Nielsen (1990).

Quanto à divisão desses conjuntos, em alguns trabalhos, os registros foram divididos numa razão de 2:1:1 (DORNELLES, 2007; OLIVEIRA *et al.*, 2013; PEDROLLO, 2017). Porém, há casos em que a falta de dados forçou os autores a adotar outras proporções. Bravo *et al.* (2008) desenvolveram uma RNA para prever vazões afluentes ao reservatório de Furnas com entradas de previsão de precipitação do ETA. Já que os 260 dados disponíveis de previsão de precipitação do ETA limitariam a quantidade de registros entrada-saída, utilizou-se precipitações observadas como previsão de precipitação para o conjunto de treinamento e validação, e previsões reais de precipitação para verificação, culminando na seguinte razão de conjuntos de 2461:1548:260.

De qualquer maneira, a divisão nos três conjuntos é um meio para não comprometer a capacidade de generalização. A etapa de treinamento e validação pode ser vista, assim, como a

calibração do modelo. Avalia-se o desempenho final apenas com o conjunto de verificação, que não faz parte do treinamento e validação nem da definição da arquitetura (HECHT-NIELSEN, 1990).

O conjunto de treinamento precisa dos valores extremos das entradas e saídas, pois RNAs não são indicadas para extrapolações fora do domínio de treinamento (PEDROLLO, 2017). O conjunto de verificação precisa de alguns eventos extremos para verificar que o treinamento garantiu uma boa generalização destes eventos de menor frequência (como exemplo, inundações), os quais, no caso da previsão hidrológica, são tão onerosos (ECKHARDT, 2008).

Já que o algoritmo de treinamento depende do conjunto de pesos iniciais, diferentes abordagens podem ser utilizadas para evitar mínimos locais: variar a taxa de aprendizado, adicionar um termo de momento, ou ruído aos dados (BRAGA *et al.*, 2007). Outra abordagem é repetir o treinamento novos pesos aleatórios e escolher a RNA que mais se adequou ao conjunto de validação, como em Dornelles (2007) e Oliveira *et al.* (2014).

2.5.4. Filtros

O estado de uma variável pode depender de uma combinação dos estados passados de uma outra variável — *e.g.* nível do rio e regime recente de precipitação. Para descrever estados passados de uma variável sem usar várias entradas em tempo subsequente, filtros temporais sintetizam informação em uma única entrada, visando parcimônia. Filtros são operadores matemáticos, como, por exemplo, chuva acumulada nas últimas 48 horas. Exemplos de filtros temporais são encontrados em Bravo *et al.* (2008), Oliveira *et al.* (2015b) e Oliveira *et al.* (2017).

Uma maneira é utilizar a média dos dados de precipitação mais recentes, isto é, aplica-se um Filtro de Média Móvel Uniforme (FMMU) para a série de precipitação de cada estação. A média de precipitações mais recentes é calculada segundo a janela móvel (j_n). Se “ j_n ” fosse 12, então, com a passagem do filtro, seria substituído cada precipitação no tempo “ t ” pela média dos últimos 11 dados e da precipitação no tempo “ t ”, como realizado no trabalho de Bravo *et al.* (2008) em que um dos nós de entrada foi a precipitação acumulada nos últimos doze dias, anteriores ao dia da previsão.

Moore (1982) propõe utilizar um Filtro de Média Móvel com decaimento Exponencial (FMME) das precipitações antecedentes para estimar a umidade do solo como uma alternativa a outros métodos que a estimam pelo balanço hídrico, utilizando vazão e evaporação. A hipótese do FMME é que precipitação mais recentes são exponencialmente mais importantes para descrever a umidade do solo. O parâmetro de meia-vida (mv) do FMME define até qual momento do passado as precipitações passadas mais recentes somarão metade da importância para a série filtrada.

$$FMME_t = (1 - \lambda) * FMME_{t-1} + \lambda * P_t \quad (6)$$

Em que:

FMME – Filtro da média móvel com decaimento exponencial no instante t

- P – Precipitação no instante t
 λ – Coeficiente de decaimento relacionado ao tempo de meia vida, tal que

$$\lambda = 1 - \left(\frac{1}{2}\right)^{\Delta t/mv} \quad (7)$$

Em que:

- Δt – Passo de tempo da série de precipitação
 mv – Tempo de meia vida (50%)

Oliveira *et al.* (2015c) analisaram a importância de certos dados de entrada para simular a vazão média diária. Aplicar o FMME para precipitação obteve destaque por extrair informações mais proximamente relacionadas com a vazão do que utilizar o próprio dado bruto de precipitação.

2.6. Métricas para avaliação

Há diversas métricas sendo sugeridas na literatura científica para analisar modelos, algumas *ad hoc* (KITANIDIS e BRÁS, 1978; MORIASI *et al.*, 2007). Cada métrica avalia aspectos ligeiramente diferentes: algumas avaliam em função de momentos de primeira ordem; outras avaliam a variância e covariância, ou comparam o modelo desenvolvido com modelos hipotéticos. Não há critérios claros que indiquem quando o desempenho do modelo é aceitável, bom e muito bom (RITTER, 2013). Assim, a avaliação de modelos acaba por vezes sujeita à subjetividade. Alguns autores, por exemplo, sugeriram metodologias com valores limítrofes para algumas das métricas (SINGH *et al.*, 2004; MORIASI *et al.*, 2007; RITTER, 2013). Neste último trabalho, por exemplo, defende-se que a raiz do quadrado do erro médio e o erro médio absoluto são apropriados para a avaliação de modelos, e que esses valores podem ser considerados baixos se forem menores que a metade do desvio padrão dos dados observados.

Alguns pesquisadores buscam um consenso do que se espera de uma boa metodologia de avaliação para modelos hidrológicos: ela deve abranger pelo menos uma análise gráfica, uma métrica adimensional e um índice de erro absoluto com informação do desvio padrão dos dados observados (LEGATES e MCCABE, 1997; RITTER, 2013). Três abordagens comumente usadas para a análise gráfica de modelos são: traçar e comparar (1) a série observada e a modelada no tempo, (2) as curvas de permanência do observado com calculado (ou frequência acumulada), ou (3) a série observada contra a modelada, examinando o quanto os pontos ficam na linha identidade 1:1 (MORIASI *et al.*, 2007; RITTER, 2013). E, destas identificar se há alguma tendência de viés.

O coeficiente de eficiência de Nash-Sutcliffe (1970) é uma métrica adimensional tradicionalmente utilizada para a avaliação de modelos hidrológicos. Análogo ao coeficiente de

determinação, compara-se o erro quadrado dos erros da série modelada com os desvios observados em relação à média histórica, expressando a proporção da variância dos dados observados que são representados pelo modelo (LANNA e INDURSKY, 1983). NS próximo a 1 indica que o erro da série modelada é mínimo perante a variância natural dos dados observados. NS zero indica que o modelo tem tanta precisão quanto um modelo hipotético que sempre adota a média observada como previsão. NS negativo indica que utilizar a média como previsão seria melhor que o modelo concebido.

$$NS = \frac{\sum(Q_{obs} - \overline{Q_{obs}})^2 - \sum(Q_{obs} - Q_{mod})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} = 1 - \frac{\sum(Q_{obs} - Q_{mod})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} \quad (6)$$

Há muitas variações dessa métrica em que os dados de entrada sofrem transformações, ou mesmo que a própria equação é modificada para, por exemplo, diminuir a sensibilidade da métrica para desvios extremos (LEGATES e MCCABE, 1997; RITTER, 2013). Essas modificações são pouco utilizadas em comparação com o NS original (MORIASI *et al.*, 2007).

As duas próximas métricas expressam o erro da série modelada. A média do erro (ME), *mean error*, avalia se o modelo costuma superestimar ou subestimar a série observada (Eq. 7). A média do erro absoluto (MEA), *mean absolute error*, serve para verificar em quanto (valor absoluto) o modelo em média se equivoca da série observada (Eq. 8). Idealmente, esperar-se-ia que um bom modelo apresentasse a média do erro e a média do erro absoluto nulas.

$$ME = \frac{1}{n} * \sum_{i=1}^n (Q_{obs,i} - Q_{mod,i}) \quad (7)$$

$$MEA = \frac{1}{n} * \sum_{i=1}^n |Q_{obs,i} - Q_{mod,i}| \quad (8)$$

A média do erro absoluto não permite perceber a relação com a unidade do dado, como a média, em percentil, do erro absoluto, ou *mean absolute percentage error*, MPEA (Eq. 9). Trata-se uma normalização, uma adequação à grandeza dos dados, que é realizada ao adicionar a $Q_{obs,i}$ como divisor do argumento do somatório na Eq. 8.

$$MPEA = \frac{1}{n} * \sum_{i=1}^n \frac{|Q_{obs,i} - Q_{mod,i}|}{Q_{obs,i}} \quad (9)$$

Outras métricas indicam o erro não ultrapassado 90, ou 95% do tempo — E90 e E95, respectivamente. Em Pedrollo e Pedrollo (2013) utilizou-se o E90 para avaliar o desempenho de redes neurais desenvolvidas para a bacia do rio Caí. Comenta-se, porém, que se deve tomar cuidado ao tirar conclusões sobre um modelo, pois, embora o E90 apresente valores aparentemente encorajadores,

isso não garante que extremos serão bem previstos — *i.e.* os 10% não considerados podem ser justo aquelas cotas extremas mais problemáticas e importantes que potencializam os danos à população. Analisar a distribuição dos erros (ou quantis dos erros) é uma maneira de inferir a qualidade de um modelo, e se esse está tendendo a superestimar ou subestimar a previsão de cotas. Em trabalho recente Pedrollo (2017) utilizou o EA95, erro absoluto não ultrapassado 95% do tempo, em sua avaliação para modelos de redes neurais artificiais para o município de Estrela-RS.

Outra estatística também mais rigorosa, utilizada no contexto de previsão de curto prazo, é aquela do coeficiente de persistência (KITANIDIS e BRÁS, 1978). Enquanto a métrica NS compara o modelo de previsão com um modelo hipotético que sempre prevê a média dos dados observados, o coeficiente de persistência compara a previsão realizada com o último dado medido. O modelo hipotético de persistência segue o processo de Wiener (a variância aumenta linearmente com o tempo) e, portanto, a melhor estimativa futura é dada pela última medição. A antecedência é j .

$$P(j) = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{mod,i})}{\sum_{i=1}^n (Q_{obs,i} - Q_{obs,i-j})} \quad (10)$$

Outro coeficiente é o coeficiente de extrapolação (KITANIDIS e BRÁS, 1978). Ele compara a previsão com a extrapolação dos últimos dados medidos. A maneira mais simples é utilizando uma extrapolação linear das duas medições mais recentes. $Q_{2,i}$ é a previsão realizada com uma reta ajustada às últimas duas medições.

$$L(j) = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{mod,i})}{\sum_{i=1}^n (Q_{obs,i} - Q_{2,i})} \quad (11)$$

Analisando publicações da área hidroambiental, Moriasi *et al.* (2007) sugeriram limites por categorias qualitativas de desempenho para algumas métricas, como o NS: muito bom, bom, satisfatório e não satisfatório. Para previsão hidrológica em SACEs, o desempenho desejado é rigoroso e está relacionado aos eventos raros, os eventos extremos. Assim, metodologias tradicionais podem indicar métricas “ótimas”, mas apresentar um erro grande, porém muito raro, de cotas extremas. Por isso, deve-se analisar justamente as situações de picos dos eventos hidrológicos extremos. Analisar também a distribuição de erros é uma maneira de averiguar isto.

2.7. Exemplos de Aplicações das RNAs

Os trabalhos de aplicações de RNA encontrados não se restringem apenas ao tipo MLP. Esse item está subdividido em aplicações (2.7.1) internacionais, (2.7.2) nacionais, (2.7.3) de previsão na bacia do Taquari-Antas. As referências estão sintetizadas na Tabela 1.

Tabela 1. Tabela de referências em aplicação de RNAs, e trabalhos de previsão na área de estudo. EMQ, Erro Médio Quadrado; MER, Média do Erro Relativo; REMQ: Raiz do Erro Médio Quadrado; r: coeficiente de correlação de Pearson; E90: Percentil 90% do erro. MEA, Erro Absoluto Médio.

INTERNACIONAL								
Autor	Data	Local	Área da Bacia	Dado Previsto	Frequência dos Dados	Período Dados	NS Verificação	Métricas Outras
Karunanithi <i>et al.</i>	1994	Rio Huron, Michigan	ca. 1.888 km ²	Vazão em local não monitorado: Estação de amostragem de Dexter	Diária	Treinamento: 1960-1972 (4748 dados) Validação: 1976-1977 (731 dados)	Não usado	EMQ: 7,272; MER: 7,6 com estação de Mill Creek EMQ:10,247 MER: 8,0 sem estação de Mill Creek
Hjelmfelt e Want	1996	Goodwater Creek, Missouri	12,2 km ²	Vazão (RNA-hidrograma unitário)	Horária*	Treinamento: 8 eventos chuva-vazão Validação: 16 eventos chuva-vazão	Não usado	Erro no pico da vazão: 21,82% (1 nni); 16.13% (10 nni); 17,28% (20 nni)
Bataineh <i>et al.</i>	1996	Jordânia	Não se aplica	Demanda de Energia Elétrica	Horária	Treinamento: abril 1993 (720 dados) Validação: 1-4 maio 1993 (96 dados)	Não usado	MER: 1%
Jayawardena <i>et al.</i>	1997	Shang Qiao e Tai Bin Chong, China	110 km ² 58.270 km ²	Cota (2 horas antes) Cota (5 horas antes)	Horária	Shang Qiao - Treinamento: 1 evento (144 dados); Validação: 5 eventos (1032 dados) Tai Bin Chong - Treinamento: 1 evento (110 dados); Validação: 4 eventos (146 dados)	Não usado	REMQ expressado como porcentagem da média observada: 7,2% (Shang Qiao); 2,6% (Tai Bin Chong)
Tawfik <i>et al.</i>	1997	Malakal e Melut, Nilo Branco	1.800.000 km ²	Cota (efeito de histerese)	Diária	Não explicitado quantos registros 1975-1977 (dois anos hidrológicos)	Não usado	MER: 3,0-4,9% dependendo do ano hidrológico e do local.
Fernando e Jayawardena	1998	Kamihonsa, Japão	3,12 km ²	Vazão em Yamaguchi, Japão	cada 10 minutos	Treinamento: 2 eventos chuva-vazão (1980-1989) Validação: Treinamento + 10 eventos chuva-vazão (1980-1989)	Não usado	REMQ expressado como porcentagem da média observada: 7,3%
Tokar e Johnson	1999	Rio Little Patuxent, Maryland	98 km ²	Vazão em Guilford	Diária	Não explicitado: redes foram treinadas com dados de 1, 2 e 3 anos. Havia dados de 1968 até 1992.	Não usado	Razão do pico previsto pelo rico observado: ca. 70%

Tabela 1 (Continuação). Tabela de referências em aplicação de RNAs, e trabalhos de previsão na área de estudo. EMQ, Erro Médio Quadrado; MER, Média do Erro Relativo; REMQ: Raiz do Erro Médio Quadrado; r: coeficiente de correlação de Pearson; E90: Percentil 90% do erro. MEA, Erro Absoluto Médio.

INTERNACIONAL								
Autor	Data	Local	Área da Bacia	Dado Previsto	Frequência dos Dados	Período Dados	NS Verificação	Métricas Outras
Jain e Kumar	2009	Rio Kentucky, Rio Alesa, e riacho Bird	10.420 km ² 885 km ²	Vazão no tempo atual (Simulação)	Diária Diária cada 6 horas	Treinamento: 1960-1972; Validação: 1977-1989 Treinamento: 01/1990-10/1993; Validação: 11/1993-01/1996 Treinamento: 11/1972-04/1974; Validação: 04/1974-12/1974	0,914 0,836 0,990	EMQ normalizado: 0,124, 0,198, 0,076
Haghverdi <i>et al.</i>	2012	Irã e Austrália	Não se aplica	Retenção de água no solo	Não se aplica	Treinamento: 80% dos dados; Validação: 20% dos dados (Método Bootstrap 50-100 vezes) 01/1929-09/1996 (813 dados)	Não usado	r entre 0,9 e 0,95
Awchi	2014	Grande e Pequeno Zab, Iraque	26.000 e 22.000 km ²	Vazões no próximo mês em Bakhma e Dokan	Mensal	Treinamento: 76%; Validação: 24%	Não usado	Bakhma: 6,22 (EMQ) 0,858 (r) Dokan: 1,417 (EMQ) 0,913 (r)
He <i>et al.</i>	2014	Pailugou, China	2,91 km ²	Vazão	Diária	Treinamento: 2001-2003 Validação: 2009-2011 (Total de 2190 dias)	0,871 (RNA) 0,887 (SVM)	0,938 (RNA) 0,947 (SVM)
Shoab <i>et al.</i>	2017	Baihe e Yanbian, China Brosna, Irlanda; Nan, Tailândia	61.780 km ² 2.350 km ² 1.207 km ² 4.609 km ²	Vazão	Diária	Treinamento: 6, 6, 8, 6 anos Validação: 2, 2, 2, 3 anos	0,78; 0,61; 0,28; 0,16 (MLPNN) 0,88; 0,84; 0,60; 0,79 (MLPNN+Wavelet)	REMQ: 655; 59,0; 9,61; 149 (MLPNN) 468; 37,6; 7,15; 74,3 (MLPNN+Wavelet)
BRASIL								
Dornelles	2007	Rio Quaraí/RS, Brasil	4.500 km ²	(1 até 5 dias antes) Níveis fluviais	Diária	(2005-2007) 1043 registros (50% treinamento; 25% validação; 25% verificação)	Não usado	r: 0,972; 0,872; 0,772; 0,720; 0,734 E90 (m): 0,369; 0,811; 0,984; 1,047; 1,029
Bravo <i>et al.</i>	2008	Reservatório de Furnas, Brasil	51.900 km ²	Vazão afluente de 1 até 12 dias à frente	Diária	Treinamento: 01/1970-07/1974 e 10/1978-12/1980 (2461 dados) Validação: 07/1974 até 10/1978 (1548 dados) Verificação: 01/1996-12/2000 (260 dados)	4d: 0,85; 12d: 0,45 (previsão ETA) 4d: 0,9; 12d: 0,85 (previsão "perfeita")	MER 4d: 14%; 12d: 35% (previsão ETA) 4d: 12%; 12d: 15% (previsão "perfeita")

Tabela 1 (Continuação). Tabela de referências em aplicação de RNAs, e trabalhos de previsão na área de estudo. EMQ, Erro Médio Quadrado; MER, Média do Erro Relativo; REMQ: Raiz do Erro Médio Quadrado; r: coeficiente de correlação de Pearson; E90: Percentil 90% do erro. MEA, Erro Absoluto Médio.

BRASIL								
Autor	Data	Local	Área da Bacia	Dado Previsto	Frequência dos Dados	Período Dados	NS Verificação	Métricas Outras
Oliveira <i>et al.</i>	2013	Rio Guaíba e bacias embutidas/RS, Brasil	967 – 71.401 km ²	Vazão no tempo atual (Simulação Vazão-Vazão)	Diária	Treinamento: 50% Validação: 15-20%; Verificação: 30-35%	NS médio de 0,75 para 35,5% de área controlada	NS>0,9 para área controlada maior que 65%
Pedrollo e Pedrollo	2013	Rio Caí/RS, Brasil	4.938 km ²	Níveis Fluviais (10 horas antes) em Barca do Caí e Passo Montenegro	Sub-diária	maior parte dos dados utilizada para treinamento e validação cruzada (01/1991-12/2007) Verificação: 01/2008-03-2011	0,962; 0,982	EM (cm): 17,6 e 11,5 E90 (cm): 39,0 e 24,4
Oliveira <i>et al.</i>	2014	Posto São Ângelo, Rio Ijuí/RS, Brasil	5.414 km ²	Vazão (Simulação)	Mensal	Treinamento: 264 (02/1966-06/1974 e 02/1978-02/1998) Validação: 122 (09/1963-01/1966, 07/1974-01/1978, e 09/2007-12/2011) Verificação: 128 (02/1961-08/1963 e 03/1998-08/2007)	0,904 (RNA 3 entradas) e 0,907 (RNA 10 entradas)	REMQ: 37,1 (RNA 3 entradas) e 36,4 (RNA 10 entradas) Erro Relativo Absoluto: 21,7% (RNA 3 entradas) e 19,1% (RNA 10 entradas)
Melo <i>et al.</i>	2015	Diferentes partes do mundo. Dados de solo do banco de dados UNSODA (v. 2.0)	Não se aplica	Curva de retenção da água no solo	Não se aplica	Treinamento: 137 dados Validação: 40 dados Verificação: 51 dados	Não usado	r da ordem de 0,65-0,69
Oliveira <i>et al.</i>	2015a	Posto São Ângelo, Rio Ijuí/RS, Brasil	5.414 km ²	Vazão (Simulação com observada ou precipitação do ETA)	Mensal	1961-1990 Não explicitado como foram divididas as subamostras.	Não usado	O maior erro médio na vazão encontrado ao utilizar os dados de precipitação do ETA foi de cerca de 36%
Oliveira <i>et al.</i>	2015b	Posto São Ângelo, Rio Ijuí/RS, Brasil	5.414 km ²	Vazão (Simulação de cenários futuros com o ETA)	Mensal	1961-1990 e 2011-2040 Não explicitado como foram divididas as subamostras.	Não usado	Modelo de Oliveira <i>et al.</i> 2015a
Oliveira <i>et al.</i>	2017	Bacia do Rio Taboão/RS, Brasil	78 km ²	Estimar conteúdo de água no solo	Não se aplica	650 dados coletados em período úmido e seco (15/01/2013-10/08/2013) Treinamento: 50%; Validação: 25%; Verificação: 25	0,77 para os melhores modelos (M69 e M71)	Distribuição dos erros com mediana próxima de 0.

Tabela 1 (Conclusão). Tabela de referências em aplicação de RNAs, e trabalhos de previsão na área de estudo. EMQ, Erro Médio Quadrado; MER, Média do Erro Relativo; REMQ: Raiz do Erro Médio Quadrado; r: coeficiente de correlação de Pearson; E90: Percentil 90% do erro. MEA, Erro Absoluto Médio.

TAQUARI-ANTAS								
Autor	Data	Local	Área da Bacia	Dado Previsto	Frequência dos Dados	Período Dados	NS Verificação	Métricas Outras
Both	2008	Bacia do Rio Taquari/RS, Brasil	19.200 e 23.300 km ²	Cota máxima de inundação em Encantado e Estrela (usando regressões)	Por evento	1956-2007 - Série de cotas máximas pareadas por eventos (Veranópolis, Encantado, Estrela)	Não usado	Veranópolis-Encantado r = 0,94; MEA = 131 cm Encantado-Estrela r=0,96; MEA = 85 cm
Eckhardt	2008	Bacia do Rio Taquari/RS, Brasil	23.300 km ²	Cota máxima de inundação em Lajeado (ca. de 6 horas antes) (usando regressões)	Por evento	1942-2007 - Série de cotas máximas pareadas por eventos (Encantado e Lajeado)	Não usado	MEA próximo de 99 cm
Chagas <i>et al.</i>	2014	Bacia do Rio Taquari/RS, Brasil	15.000 – 26.000 km ²	Muçum, Encantado, Estrela, Bom Retiro, Porto Mariante e Taquari	Não se aplica	Não explicitado	Não usado	Não mencionado
Siqueira <i>et al.</i>	2016	Bacia do Rio Taquari/RS, Brasil	19.200 km ²	Vazão em Encantado até 84 horas antes (usando MGB)	Cada 12 horas	Calibração: 06/2009 até 12/2010; Validação: 01/2012-12/2013	0,854 na simulação	MER da vazão simulada: - 2,1% Momento do pico foi acertado. Análise foi feita por dois eventos
Kurek	2016	Bacia do Rio Taquari/RS, Brasil	23.300 km ²	Previsão de níveis em Estrela/RS (usando regressões)	Por evento	Apenas ajuste: eventos entre 08/1980 - 06/1914	0,932 no ajuste	r: 0,963 no ajuste
Pedrollo	2017	Bacia do Rio Taquari/RS, Brasil	23.300 km ²	Previsão de níveis em Estrela/RS (usando regressões)	A cada 6 horas	Período não informado; Treinamento: 2660 registros; Validação: 1300 registros; Verificação: 1478 registros.	Modelos completos: 5 horas antes: 0,97 11 horas antes: 0,93	Modelos completos, MEA: 5 horas antes: 0,09 11 horas antes: 0,14
Gerhard	2017	Bacia do Rio Taquari/RS, Brasil	23.300 km ²	Previsões de níveis em Lajeado/RS (usando regressões por evento ou RNA)	Por evento ou cada 6 h	Abordagem por evento: Calibração: 1980-2009; Validação: 2010-2016 Abordagem com RNA: Não informado (alcance de 6 horas)	0,951 0,980	MEA 48,4 cm MEA 6,6 cm

2.7.1. Internacional – revisões

Algumas revisões analisaram tipos de RNAs e aplicações no campo da hidrologia antes dos anos 2000 (ZHANG *et al.*, 1998; ASCE, 2000a-b). Na época, o uso das redes ainda era incipiente e recém ganhava popularidade no campo da hidrologia.

As principais vantagens identificadas das RNAs foram: (1) reconhecer relações multidimensionais entre variáveis de entrada e saída sem explícita consideração física; (2) implementar relações mesmo com ruído e erros de medição; (3) adaptação para compensar mudanças do sistema; (4) ser prática, pois depois de treinada é de fácil uso (ZHANG *et al.*, 1998; ASCE, 2000a). Algumas das limitações levantadas foram: depender da qualidade e da grande quantidade de dados, e carecer de relações e conceitos físicos durante seu desenvolvimento. Esta carência é uma das principais razões para o ceticismo contrário às RNAs (ASCE, 2000a).

Algumas aplicações nas subáreas da hidrologia foram observadas (ASCE, 2000b). Categorizam-se elas nos grupos como: chuva-vazão, vazão-vazão, qualidade da água, água subterrânea, estimar precipitação por dados de sensoriamento remoto. Dado o contexto dessa pesquisa, apresentam-se alguns dos trabalhos das duas primeiras categorias com objetivo de firmar um panorama geral dos trabalhos internacionais até o fim do último milênio.

Uma boa seleção de variáveis de entrada para o treinamento pode ter muito impacto na posterior precisão de previsão. Fernando e Jayawardena (1998), em sua pesquisa sobre previsão de vazões em tempo atual, analisaram a correlação e a autocorrelação entre possíveis variáveis de entrada e de saída, identificando por meio destas o alcance a ser adotado e a memória da vazão no posto de previsão. Foi avaliada a diferença entre as redes MLP, RBF (*Radial Basis Function*, outro tipo de RNA que operam com as chamadas funções de base radial, que determinam os espaços do domínio das entradas para os quais a rede pode responder) e o modelo ARMAX: as duas primeiras apresentam resultados semelhantes e superiores aos da terceira abordagem (JAYAWARDENA *et al.*, 1997; FERNANDO e JAYAWARDENA, 1998). Nesses últimos dois trabalhos os modelos foram calibrados com eventos, o que não é sugerido no contexto da previsão de curto prazo, pois não se sabe *a priori* que um evento vai acontecer (PEDROLLO, 2017). Tokar e Johnson (1999), por exemplo, com as variáveis de entrada já escolhidas, testaram diferentes seleções de período para o treinamento (período seco, úmido ou seco e úmido) e notaram que as melhores previsões foram justamente aquelas treinadas com registros dos períodos seco e úmido. Hoje essa necessidade já se torna mais intuitiva à luz dos fundamentos das RNA, pois uma vez que o objetivo é realizar previsões em qualquer momento do ano, é imprescindível que a rede aprenda a partir de padrões de ambos os períodos.

Hjelmfelt e Wang (1996) demonstraram, num estudo peculiar, que mudanças na estrutura das redes podem promover ganhos. Os autores desenvolveram RNAs de arquitetura diferentes, que

dialogavam melhor com a teoria do hidrograma unitário e da precipitação efetiva. Isto é, os nós de entrada recebem uma sequência de precipitação, e, ao invés conectarem com todos os neurônios internos, cada nó de entrada conecta apenas com um único neurônio interno (uma estrutura paralela, porém não distribuída). Essa camada interna tem o papel de extrair apenas a precipitação efetiva. Por fim, a camada de saída trata de ponderar uma soma para estimar o hidrograma unitário (estrutura distribuída). Os autores concluem que, para esse fim específico, a arquitetura utilizada foi melhor do que usar uma rede MLP de três camadas, ou o método do hidrograma unitário.

Karunanithi *et al.* (1994) avaliaram duas abordagens (regressão exponencial e RNA) para preencher dados de vazão no rio Huron, Michigan, a partir de outras três estações fluviométricas nas redondezas. Sua motivação foi estudar como a variação de vazão em diferentes trechos do rio Huron afetava o habitat de uma espécie de robalo, *Micropterus dolomieu*. A RNA apresentou melhor desempenho e foi mais robusta: ao adicionar mais um nó de entrada com dados de uma estação que prejudicou o modelo exponencial, o desempenho da RNA melhorou. Segundo os autores, apesar de investigações preliminares indicarem certas variáveis como espúrias/ruidosas, a RNA pode ainda assim obter informações dessas.

RNAs apresentaram desempenho satisfatório para diferentes aplicações no meio da hidrologia, inclusive para previsão de vazões. Na maioria dos estudos apresentados, as RNAs tiveram desempenho superior quando comparadas com outros modelos empíricos e conceituais. Na opinião do comitê do ASCE, porém, elas não foram particularmente úteis em ampliar o conhecimento sobre os processos hidrológicos em uma bacia. Segundo os autores, é como se a física ficasse trancada num conjunto de pesos e não fosse revelável após o treinamento (ASCE, 2000b). Trabalhos mais recentes apresentaram maneiras de revelar essas relações entrada-saída:

Conscientes da importância em analisar o funcionamento subjacente das redes, Jain e Kumar (2009) apresentaram uma metodologia para “dissecar” as regras ocultas de redes neurais treinadas para previsão hidrológica. Segundo os autores, as regras podem ser buscadas em dois locais: ou na estrutura paralela massiva das camadas internas da rede, ou nos pesos sinápticos da rede. Jain e Kumar (2009) investigaram no primeiro local ao comparar as saídas parciais das redes, considerando tanto cada neurônio interno por si só como também a combinação deles, com certos componentes conceituais. Os componentes conceituais de análise escolhidos foram escoamento de base e escoamento superficial, que foram calculados utilizando métodos clássicos. A metodologia envolveu o uso de análise visual do gráfico de dispersão (saídas parciais x componentes conceituais), correlação e regressão. Os autores associaram a tendência de comportamento de partes da rede com os componentes conceituais.

Zhang *et al.* (1998) publicaram uma revisão sobre o estado da arte das RNAs para realizar previsões num âmbito geral (não apenas hidrológico). Uma das barreiras identificadas para o uso das

redes é a carência de uma metodologia para definir a estrutura ideal, o que acaba sendo realizado heurísticamente (em inglês, *trial and error*) de maneira um tanto exaustiva. Alguns trabalhos sugeriram o uso de ferramentas auxiliares para a seleção de variáveis de entrada: análise espectral dos dados, conhecida também como *wavelet* ou ondaleta, e sistemas especialistas difusos, que fazem uso de lógica difusa (ou *fuzzy*). Bataineh *et al.* (1996), a partir de um sistema especialista difuso (SED) associado com RNAs, identificou a demanda por energia elétrica na Jordânia a partir de temperatura, localização geográfica, grau de industrialização e outros fatores, para depois definir regras de operação para geradores.

Uma perspectiva dos trabalhos a partir de 2013 sobre previsão hidrológica com RNAs é apresentada, citando artigos buscados nas seguintes revistas: *Journal of Hydrology* (Amsterdam), *Water Resources Management*, *Hydrological Sciences Journal* e *Journal of Hydrologic Engineering*.

Awchi (2014) realizou estudo de previsão mensal nas bacias dos rios Grande Zab e Pequeno Zab no Iraque setentrional, cujos comprimentos e áreas de drenagem são, respectivamente, de 400 km e 26.000 km², 370 km e 22.000 km². Os estudos de previsão foram de caráter autorregressivo (realizar previsão baseando-se em vazões dos meses passados), utilizando FFNN, GRNN, RBF e RLM. Os autores concluíram que todas as redes apresentaram valores de MSE, RE e R superiores àqueles para RLM. Embora as redes tenham apresentado métricas bastante semelhantes, a RNA-MLP foi a que se destacou. Ao adicionar um neurônio de entrada referente ao mês de previsão (1-12) notou-se significativa melhora das métricas. Pôde-se preencher falhas de vazões mensais de um rio, baseando-se nas vazões passadas do outro. Por fim, análises das vazões extremas expuseram que vazões máximas foram um pouco subestimadas, e mínimas, notavelmente superestimadas.

Valipour *et al.* (2013) compararam modelos ARMA, ARIMA, e MLP autorregressiva para prever vazões mensais afluentes ao reservatório Dez no Irã. Essa MLP modifica os pesos da camada interna segundo a saída. Por transformar a série ser modelada em estacionária, o modelo ARIMA foi melhor que o ARMA, prevendo com até doze meses de antecedência. A MLP autorregressiva pôde realizar previsões com até 60 meses de antecedência. Não se notou diferenças marcantes nas MLP estáticas e dinâmicas para diferentes números de neurônios internos.

Segundo Nourani *et al.* (2014), houve um aumento considerável de trabalhos no meio hidroclimatológico inteligência artificial e ondaletas a partir de 2009, as quais são usadas principalmente no pré-processamento de dados de entrada para decompor o ruído do sinal principal. Trata-se de uma filtragem, uma decomposição da série, em que se busca eliminar erros sistemáticos.

Em estudo de previsão envolvendo diferentes RNAs MLP com e sem ondaleta para quatro bacias de climas diferentes (China, Irlanda, Tailândia), vantagem foi averiguada com NS melhorando de 0,75 para 0,9 para bacia de Baihe, China (SHOAIB *et al.*, 2017). O uso da ondaleta pode melhor descrever a “assinatura hidrológica” da bacia.

Foram desenvolvidos modelos de previsão de vazão para diferentes alcances baseados em RNAs comuns e RNAs emocionais (*Emotional Artificial Neural Networks*) em duas zonas climáticas diferentes (bacia hidrográfica de Moselle, cerca de 5.158 km², sub-bacia do rio Reno; e Lobbs Hole Creek, cerca de 28.800 km², sub-bacia do rio Murrumbidgee na Austrália). As redes emocionais possuem unidades de processamento hormonal que auxiliam o procedimento de treinamento. Três estratégias (75% e 25%, 50% e 50%, 40% e 60% para treinamento e verificação) de divisão de dados foram utilizadas. Os autores concluíram que menor o número de registros de treinamento, pior o desempenho na verificação, e que as RNAs emocionais precisam de menos dados de treinamento do que as RNAs comuns, apresentando desempenho na verificação até 34% melhores, que foi atribuído a saberem reconhecer e distinguir dias úmidos e secos (NOURANI, 2017). O período de verificação foi diferente daquele utilizado para testar o tamanho do conjunto de treinamento.

Comparou-se o desempenho de redes MLP, SVM (máquinas de vetores de suporte) e ANFIS (RNA acopladas com SED) para realizar previsão de vazões na bacia semiárida e montanhosa de Pailugou na China (HE *et al.*, 2014). As métricas NS, R e RMSE apresentaram valores similares. Há uma pequena vantagem das SVM segundo os autores. Apesar das boas métricas (NS em torno de 0,87 e R de 0,94) indicarem que esses modelos podem ser usados até em regiões de topografia acidentada, houve subestimação significativa de vazões de inundação, o que indica que no quadro da previsão de vazões de inundação uma outra métrica mais sensível para vazões máximas deveria ser utilizada.

Esses trabalhos internacionais reafirmam a vantagem da técnica das RNAs do tipo MLP sobre outras técnicas, como a dos modelos ARIMA e suas variações. Além disso, o uso associado com a técnica de pré-processamento, como ondaletas, pode favorecer o desempenho das RNAs.

2.7.2. Trabalhos no Brasil

Aqui são apresentados trabalhos do grupo de Aplicações Hidroambientais com RNAs (AHRNA) que fizeram uso de RNAs MLP, demonstrando a aplicabilidade desse tipo de modelagem em região brasileira. Esses também servem para embasar a metodologia a ser proposta nesta pesquisa.

Há alguma relação entre área controlada e capacidade de simulação? Oliveira *et al.* (2013) realizaram 83 simulações em quinze bacias embutidas na região hidrográfica do rio Guaíba. Concluiu-se que há maior capacidade conforme aumenta a área controlada. Simulações com pouco mais de 35% de área controlada apresentaram um NS superior a 0,8, o que não só é um indicativo para o monitoramento do recurso hídrico, como também um indício para a aplicabilidade das RNAs.

Desenvolveu-se e avaliou-se um modelo com RNA MLP para simular vazões médias mensais para a bacia do rio Ijuí (OLIVEIRA *et al.*, 2014). As variáveis de entrada foram derivadas de evapotranspiração estimada por Penman-Monteith e a precipitação mensal foi aplicada com diferentes atrasos. Inicialmente, a rede tinha dez variáveis de entrada, mas essas foram reduzidas para três com

um método de poda chamado análise de sensibilidade, adaptado de Lek *et al.* (1996). Nem sempre o dado puro é a melhor opção para a RNA generalizar um processo. Das três variáveis de entrada, duas utilizaram FMME, e todas continham dados médios do mês atual e anterior. Assim, cada variável de entrada representou mais do que apenas um processo hidrológico isolado. Comparou-se o desempenho da RNA com o modelo conceitual SMAP (LOPES *et al.*, 1982), utilizando o mesmo período de dados. O NS da RNA foi de 0,898, enquanto do SMAP foi de 0,672.

Melo *et al.* (2015) também utilizaram o método de análise de sensibilidade. Para estimar a curva de retenção da água no solo, o método mais utilizado é o das funções de pedotransferência. Foram desenvolvidas três RNAs: uma de referência baseada no trabalho de Haghverdi *et al.* (2012) com medições brutas como dados de entrada e outras duas com dados ajustados. A rede de Haghverdi utiliza potencial mátrico como entrada, enquanto as de Melo *et al.* (2015) não. Ambas abordagens apresentaram desempenho semelhante. O método da análise de sensibilidade trouxe contribuições para compreensão da física do processo: nas RNAs de Melo *et al.* a densidade aparente não desempenhou um papel relevante em condições secas. Outra contribuição foi a de que, ao ajustar os dados com uma metodologia já bastante usada na literatura, é possível extrapolar para além dos extremos das medições brutas com maior segurança.

Modelos climáticos globais não são indicados para uso regional devido à sua baixa resolução. O modelo climático regional ETA CPTEC/HadCM3 possui quatro membros de sensibilidade climática, que tratam de reproduzir as incertezas das condições de contorno — *i.e.* a incerteza quanto ao estado da atmosfera. Foram avaliadas as incertezas em utilizar os dados de precipitação dos quatro membros do ETA CPTEC/HadCM3 como entrada para simular as vazões médias mensais na bacia do rio Ijuí (OLIVEIRA *et al.*, 2015a). Por não ser sugerido utilizar os dados brutos do ETA, foram testados quatro métodos de correção para cada membro. O cenário climático advindo do ETA-Low, aplicando a correção *Direct Approach*, destacou-se. Este estudo serviu como base para simular cenários futuros de disponibilidade de água utilizando as RNAs.

A partir da simulação hidrológica realizada com uma RNA para o período 2010-2040, utilizando o ETA-Low com correção *direct approach*, decompôs-se a série histórica em quatro componentes: tendência de longo período, componente cíclico/sazonal, componente dependente do tempo e componente estocástico. Com esses componentes identificados, geraram-se 1000 possíveis séries estocásticas. A média da vazão tende a aumentar durante esse período, assim como a variabilidade, o que implica em mais ocorrências de secas e inundações (OLIVEIRA *et al.*, 2015b), o que dentre outros pontos sugere, indiretamente, a importância de mais sistemas de alertas para prevenção de desastres.

Bravo *et al.* (2008) desenvolveram uma RNA com entradas de previsão de precipitação para prever a vazão afluente ao reservatório de Furnas com até 12 dias de antecedência, o que auxilia na

definição de regras de gestão desse. A inclusão de previsões de precipitações permite estender o alcance da previsão para além do tempo de concentração da bacia. Há quatro sub-bacias e uma sub-bacia incremental com dados de vazão que contribuem para vazão afluente em Furnas. Para definir a defasagem dos eventos medidos nessas e em Furnas os autores investigaram correlogramas. O mesmo foi feito para definir o número de dias de precipitação acumulada que contribuem para a vazão afluente em Furnas, resultando em 12 dias acumulados. Foram realizadas previsões com precipitações previstas pelo ETA e utilizando precipitações observadas, o que de certa maneira demonstrou uma espécie de patamar máximo que poderia ser atingido nas redes utilizadas, caso fossem as previsões de precipitação extremamente certas.

Dornelles (2007) propôs o uso da previsão do ETA para aumentar a antecedência de previsão em tempo atual para a bacia do Rio Quaraí, de 4.500 km², pois alcance da bacia é curto apenas com dados de níveis de montante. Nesse trabalho foram utilizados, entre outras métricas, o E90 e o coeficiente de persistência para analisar a qualidade dos modelos de RNAs desenvolvidos. Entre os modelos desenvolvidos, aqueles com previsão de precipitação mostraram ligeira vantagem sobre aqueles que não utilizam a previsão de precipitação. Segundo o autor, o ETA tem uma baixa qualidade em acertar a magnitude da precipitação futura.

RNAs MLP foram aplicadas para realizar previsões na bacia do rio Caí no quadro do SACE da CPRM/SGB (PEDROLLO e PEDROLLO, 2013). A bacia tem 4.938 km² e suas inundações são rápidas, ocorrendo em qualquer época do ano e podendo durar até alguns dias. Desenvolveram-se modelos para o posto Barca do Caí, em São Sebastião do Caí, e Passo Montenegro. Os desempenhos das redes foram satisfatórios: para o posto de Barca do Caí, obteve-se NS de 0,962, erro médio de 17,6 cm e E90 de 39 cm; para o posto de Passo Montenegro, obteve-se NS de 0,982, erro médio de 11,5 cm e E90 de 24,4 cm. Apresentam-se três aspectos desejáveis no trabalho: incorporar incertezas e desenvolver modelos adaptáveis às modificações da rede e à eventual falta de sinal.

2.7.3. *Alguns trabalhos envolvendo previsão no Taquari*

Alguns estudos envolvendo previsão foram realizados na bacia do Taquari-Antas (CHAGAS *et al.*, 2014; SIQUEIRA, 2015; SIQUEIRA *et al.*, 2016; PEDROLLO, 2017), sendo alguns no quadro da UNIVATES à região de Lajeado/Estrela no vale do Taquari (BOTH, 2008; ECKHARDT, 2008; KUREK, 2016; GERHARD, 2017).

Os trabalhos aqui descritos estão em três categorias: (1) Previsão empírica discreta (para Estrela/RS), realizada por meio de regressões entre cotas máximas pareadas que foram atingidas em montante e jusante para cada evento (vantagem: utilizar série com mais informação de dados extremos desde 1940; desvantagem: horizonte de previsão indefinido, já que a correlação é baseada em cotas máximas pareadas que ocorreram com defasagens diferentes); (2) Previsão empírica

contínua com até 11 horas de antecedência (para Estrela/RS), realizada por meio de regressões ou RNAs entre informações hidrológicas de montante para explicar fenômenos em jusante (vantagem: horizonte de previsão claro e bem definido; desvantagem: séries históricas disponíveis relativamente curtas); (3) Previsão conceitual contínua do tipo chuva-vazão com até 84 horas de antecedência (para Encantado/RS), *i.e.* através da transformação dos dados de precipitação prevista para descrever a vazão futura em Encantado.

Both *et al.* (2008) elaboraram modelos de previsão de inundação baseado em regressão linear simples, correlacionando cotas máximas de montante e jusante. Uma vez constatada a elevação do nível em Veranópolis, pode-se estimar os níveis a serem atingidos em Encantado e Estrela. Interpreta-se que, segundo os autores, pode-se prever com até 10 horas de antecedência a proporção da enchente a ocorrer no Vale do Taquari utilizando a regressão Veranópolis-Estrela, cujo R^2 é de 0,78.

Eckhardt (2008) desenvolveu um modelo cartográfico de previsão para prever a cota de inundação atingida em Lajeado, e a respectiva área urbana inundada. A equação cota-cota foi desenvolvida com base na série de cotas máximas pareadas atingidas em Lajeado e Encantado por eventos entre 1942 e 2007. O modelo prevê com cerca de 6 horas de antecedência, apresentando um erro médio absoluto próximo de 1 m.

Houve um Sistema de Previsão e Alerta de Enchentes (SPAEE) da UNIVATES em operação entre 2001 e 2007 no vale do Taquari. Em um segundo período uma rede foi mantida por esforços conjuntos da UNIVATES e da CPRM (GERHARD, 2017). A bacia do rio Taquari-Antas conta hoje com o SACE-Taquari, operado pela CPRM/SGB. O sistema começou a ser instalado no segundo semestre de 2013 (CHAGAS *et al.*, 2014). Os primeiros modelos desenvolvidos foram regressões múltiplas baseadas em propagação da onda de cheia, e, segundo os autores, pretendia-se o desenvolvimento de um modelo chuva-vazão. O artigo não traz métricas para esses modelos, porém, sim, alcances dos modelos baseados em regressões múltiplas — *e.g.* para o porto de Estrela é possível previsão com 6 horas de antecedência.

Um modelo conceitual chuva-vazão está em desenvolvimento pela CPRM/SGB para previsões do SACE-Taquari. O *software* de modelagem utilizado, HEC-HMS, foi desenvolvido em 1992 pelo *Hydrologic Engineering Center* do Corpo de Engenheiros do Exército dos Estados Unidos (2019) (*United States Army Corps of Engineers*). O modelo utilizado pela CPRM/SGB está sendo aprimorado tanto no quesito da modelagem hidrológica como também na distribuição das estações em campo. Uma das dificuldades é lidar com a variabilidade na distribuição das chuvas que distorcem os resultados.

Devido ao estágio atual do modelo, não se tem acesso à documentação ou às métricas resultantes da aplicação do modelo. Ainda assim, no portal SACE-Taquari foram disponibilizados 67

boletins de previsões, sendo o primeiro do dia 27/03/2016 e o mais recente do dia 03/20/2018 (CPRM, 2019). Nesses boletins constam previsões com limiares de máximo e mínimo.

Houve esforços para realizar previsões de inundações com previsões de precipitação, como aquelas do ETA determinístico 15 km e do ETA de 5 membros (SIQUEIRA, 2015b). Versão do ETA operacional de 40 km, o modelo regional ETA determinístico 15 km utiliza um esquema de parametrização para precipitação convectiva (Betts-Miller-Janjic) e um esquema de parametrização para microfísica de nuvens e precipitação estratiforme (Ferrier). Esse modelo é bastante utilizado no *software* SISMADEN (Sistema de Monitoramento de Desastres) que foi desenvolvido pelo INPE. O modelo regional ETA de 5 membros, não mais disponível, possuía cinco diferentes membros, os quais tratavam de combinar diferentes suposições de condições iniciais, diferentes maneiras de parametrização para descrever o comportamento da atmosfera (*e.g.* a microfísica de nuvens e a precipitação convectiva) para tentar lidar com o caos determinístico envolvido em previsões meteorológicas. O ETA de 5 membros trata de prover diferentes cenários baseados em suposições do estado e comportamento da atmosfera (SIQUEIRA, 2015b). Siqueira *et al.* (2016) utilizaram o modelo MGB-IPH acoplado com duas versões diferentes do ETA/CPTEC (ETA de 5 membros, e ETA 15 km determinístico) para realizar previsões para dois eventos hidrológicos extremos na bacia do rio Taquari-Antas para o município de Encantado, que é tido como uma localidade crítica e de difícil previsão. O MGB-IPH é um modelo conceitual chuva-vazão semidistribuído, desenvolvido inicialmente por Walter Collischonn em sua tese (2001) e posteriormente aprimorado por outros pesquisadores. Os eventos analisados foram a inundação do dia 21/07/2011 às 12 h e a inundação do dia 6/6/2014 às 12 h. A metodologia utilizada trata de primeiro “aquecer” o modelo MGB-IPH pelos seis meses anteriores ao momento em que se realiza previsões em Encantado com dados *in situ*. Depois, são utilizados os dados de previsão de precipitação como sendo dados de entrada futuros. Ao analisar a previsão com até 84 horas de antecedência (de 12 em 12 h) para Encantado durante os dois eventos de inundação, nota-se que os picos previstos e observados ocorrem em momentos semelhantes, porém ainda há bastante incertezas nesta abordagem, em grande parte envolvendo aquelas advindas da previsão de precipitação. O trabalho apresenta métricas para a calibração (Jun/2009-Dez/2010) e a validação (Jan/2012-Dez/2013) do modelo MGB-IPH: para a estação fluviométrica de Encantado (86720000), o NS foi 0,923 e 0,854 e o erro de volume foi -9,2% e -2,1%.

Kurek (2016) desenvolveu modelos de previsão de níveis para Estrela/RS, baseados em RLM com eventos discretos. Foram testadas 59 combinações entre dados diários de cota convencionais do Hidroweb (Encantado, Estrela, Muçum no dia do evento e no dia anterior), dados de cota máxima de inundação advindos de estações telemétricas da rede Univates (Encantado e Estrela) e da CPRM (Muçum), e precipitação média na bacia do rio Taquari-Antas — obtida por polígonos de Thiesen com dados de 15 estações de precipitação do Hidroweb, da rede da Univates e da rede da CPRM.

Filtros foram aplicados nas precipitações, acumulando as séries de dados em 1, 2, 3, 4 e 5 dias. Os melhores modelos tiveram apenas a cota máxima atingida na inundação, resultando em duas regressões (uma com todos os registros, outra com um *outlier* removido). Os modelos foram aplicados para dois eventos de inundação ocorridos em Estrela. Entre meados julho de 2015, a cota máxima registrada em Estrela foi de 21,35m e a diferença entre picos em Encantado e Estrela foi de cerca de 6 horas. Os modelos superestimaram a cota máxima em cerca de 80 cm, e o hidrograma de previsão antecipou a ascensão da inundação com superestimativas de até 3 m. Para o evento de inundação de cota máxima 24,51, ocorrido em Estrela entre 8 e 10 de outubro de 2015, a cota máxima foi subestimada em 22 cm e a diferença dos picos de Encantado e Estrela foi de 12 horas. Semelhante ao primeiro evento, a ascensão prevista é antecipada com erro de até 1 m (KUREK, 2016). Os dados de precipitação não contribuíram para o melhor modelo, o que foi atribuído a grande parte da informação de precipitação de montante já constar nos dados de níveis de Muçum e Encantado. O excesso de informação pode ser prejudicial ao modelo, devendo-se utilizar apenas a informação da chuva ocorrida na bacia complementar (KUREK, 2016).

Gerhard (2017) desenvolveu modelos empíricos para prever níveis em Lajeado/RS com duas abordagens. Utilizaram-se dados de cotas em Encantado e Lajeado, e precipitação na bacia incremental. A abordagem discreta por eventos utiliza RLM de cotas máximas pareadas, semelhante a Eckhardt (2008). O período de dados da primeira abordagem foi de 01/07/1940 até 31/12/2016. Na abordagem contínua (aplicada para RLM e RNA), empregam-se séries em intervalo sub-diário (6h, 12h, 18h, 0h) de níveis fluviais em Lajeado e Estrela. Precipitação foi agregada com filtros de soma móvel e decaimento exponencial (meia-vida de 0,46 e 0,87 para RLM e RNA, respectivamente). O período utilizado foi de 15/08/1980 até 31/12/2007 e de 01/08/2015 até 31/12/2016, respectivamente. A abordagem por eventos pôde ser melhorada utilizando a precipitação média da bacia incremental, como em Kurek (2016). O evento de inundação ocorrido em 05/01/2010 teve grande contribuição do rio Forqueta, que está na bacia incremental entre Encantado e Lajeado. O erro para esse evento é de -3,26 m em Kurek (2016) e 0,84 m em Gerhard (2017). Na abordagem por eventos, dois modelos foram desenvolvidos com diferentes períodos de dados. O de maior período (calibração: 1941-2005; validação: 2006-2016) apresentou piores métricas (*e.g.* NS de 0,899 e MEA de 73,9 cm) do que o de menor período (calibração: 1980-2009; validação: 2010-2016) (*e.g.* NS de 0,951 e MEA de 48,4 cm). Embora seja dito que o modelo para o período mais recente seja melhor (GERHARD, 2017), o período de validação ser diferente deixa a análise subjetiva. Na abordagem contínua, o modelo baseado em RLM, utilizando o nível em Encantado no “ t_0 ” e o nível em Lajeado no “ t_0 ” para prever o nível em Encantado no “ $t_0 + 6$ horas”, apresentou as seguintes métricas de desempenho para o horizonte de previsão de 6 horas: NS de 0,967, MEA de 8,7 cm e EA95 de 33,6 cm. O modelo de RNA para o mesmo alcance possui 4 nós de entrada (nível em Encantado no t_0 ; nível em Lajeado no

t0; precipitação da bacia complementar acumulada em 3 dias e precipitação com FMME com meia-vida de 6 horas), 7 neurônios na camada interna e 1 neurônio de saída que é o nível em Lajeado em t+6. O modelo apresentou o seguinte desempenho: NS de 0,980, MEA de 6,6 cm e EA95 de 26,6 cm. Na visão da autora, apesar das métricas serem muito semelhantes em ambas abordagens, há grande vantagem em trabalhar com séries contínuas: o horizonte de previsão é definido (GERHARD, 2017).

Pedrollo (2017) desenvolveu, em caráter preliminar, um modelo baseado em um conjunto RNAs MLP com vistas a realizar previsões para um sistema de alerta para Estrela/RS. Utilizando dados convencionais do monitoramento da bacia do Taquari e AHSUL, pôde-se desenvolver dois modelos para efetuar previsões. Um com antecedência de 5 horas, o outro, 11 horas. O CM (modelo completo) de 5 horas de antecedência apresentou NS de 0,97, EA95 de 32 cm e MEA de 9 cm. O CM de 11 horas de antecedência apresentou NS de 0,93, EA95 de 53 cm e MEA de 14 cm. O autor reforça que se desejam modelos resilientes, previsões associadas à incerteza, e adaptabilidade caso ocorram mudanças no sistema físico e também de aquisição de dados, isto é, número e cobertura de postos. O conjunto de redes neurais é uma maneira de tornar a previsão mais resiliente e adaptativa frente a eventuais degradações do sistema de monitoramento, que não são raras de acontecer como no exemplo de acompanhamento de um evento na bacia do Caí entre os dias de 16 e 19/09/2012 (PEDROLLO e PEDROLLO, 2013). Trata-se de uma árvore de modelos que opta por diferentes RNAs, dependendo de quais estações estão de fato transmitindo dados. Pedrollo (2017) discute brevemente a seleção de períodos para treinamento e verificação dos modelos: previsões confiáveis são importantes não só durante a ascensão do evento, mas também em sua recessão. Na prática não se sabe, inicialmente, quando um evento está ocorrendo. Portanto, a previsão deve ser de caráter contínuo e não “por eventos”. Os modelos de Pedrollo (2017) foram ajustados com dados convencionais da Rede Hidrometeorológica Nacional da ANA: postos Passo Tainhas, Encantado, Muçum, Santa Lúcia e Estrela. Dos dois primeiros postos, séries históricas apenas de precipitação foram utilizadas. Do posto de Estrela usaram-se dados monitorados às 6, 12, 18 e 24 horas (dados esses da AHSUL). No total, havia 5.458 registros (após a exclusão de falhas), dos quais 2.660 serviram para o treinamento, 1.300 para a validação, e o resto para a verificação.

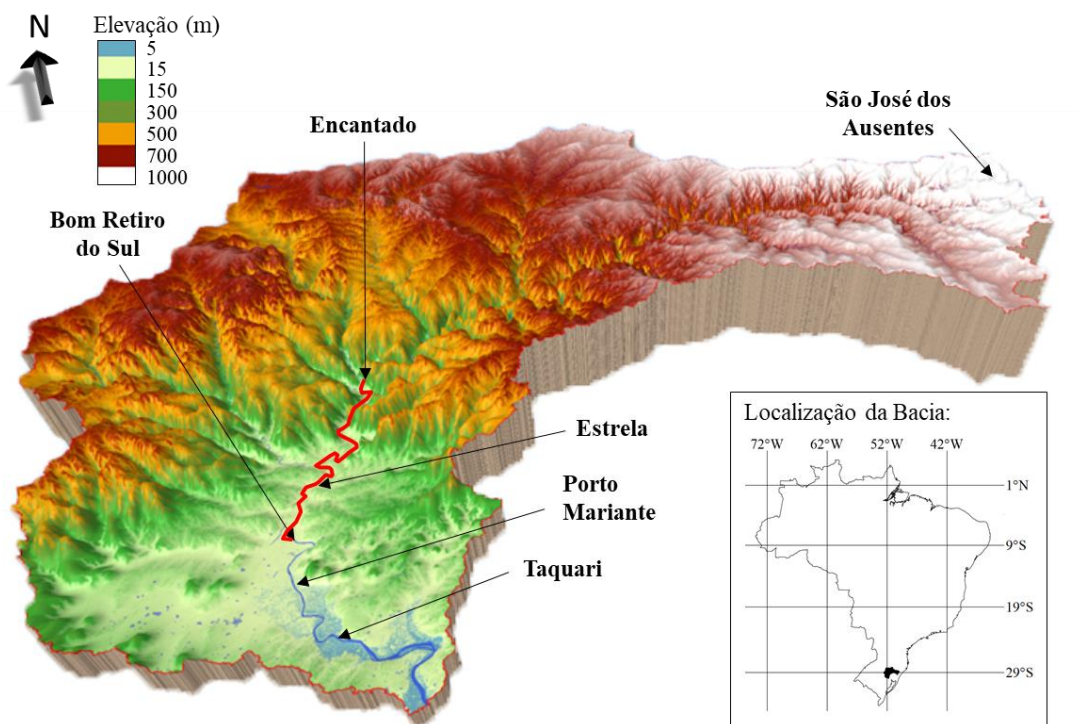
3. MATERIAIS E MÉTODOS

Este item está dividido em (3.1) Área de estudo; (3.2) Estações utilizadas (3.3) Análise dos dados; (3.4) Desenvolvimento dos modelos; (3.5) Avaliação do desempenho do modelos.

3.1. Área de estudo

A área de estudo é a bacia do rio Taquari-Antas com seus 26.600 km² e rio principal de 530 km de comprimento. A topografia e alguns detalhes relevantes da bacia, como o trecho crítico a inundações entre Encantado e Bom Retiro do Sul, são apresentados na Figura 8.

Figura 8. Representação da topografia da bacia do rio Taquari-Antas. A linha vermelha apresenta o trecho mais crítico a inundações.



Fonte: produção do autor do trabalho no software QGIS 3.10.4, utilizando dados do SRTM30 (NASA/NGA) e o complemento “qgis2threejs”.

3.1.1. Rio Taquari-Antas e a navegação comercial

O rio Taquari-Antas inicia no município de São José dos Ausentes na Serra Geral em uma altitude superior a 1100 metros, onde é chamado de rio das Antas. A partir da junção com o rio Carreiro, cerca de 55 km a montante de Encantado, seu nome é rio Taquari (CHAGAS *et al.*, 2014). Nessa pesquisa, o rio principal será chamado de rio Taquari-Antas, como já em outros trabalhos (KICH *et al.*, 2015; SIQUEIRA *et al.*, 2016).

Realizada da cidade de Muçum (13 km a montante de Encantado) até a foz do rio Taquari-Antas no rio Jacuí, a navegação comercial conta com três portos públicos de montante a jusante: Estrela, Mariante e Taquari. Entre eles, o porto de Estrela é o principal meio para importação e

exportação com o porto de Rio Grande/RS. Os produtos comercializados são grãos, farelos, fertilizantes, carvão, óleos vegetais, e areia e brita para construção (FEPAM, 2012).

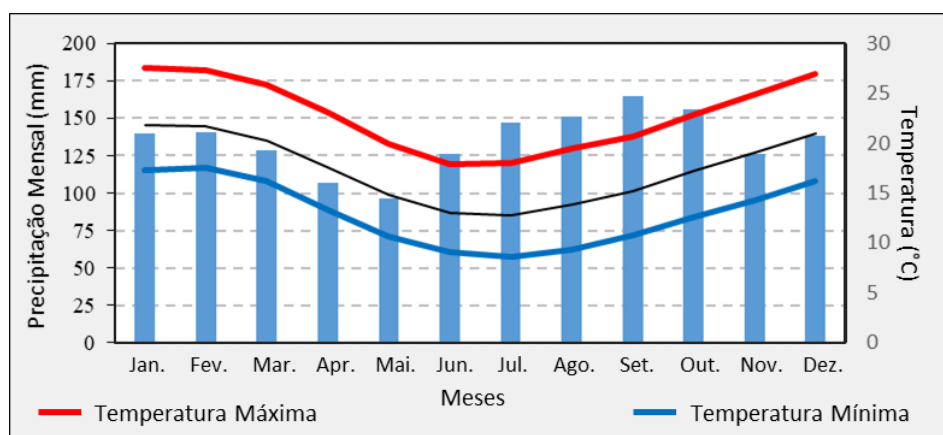
Há uma barragem com eclusa entre Porto Mariante e Estrela, que desde a década de 1970 busca manter a cota de 13 metros em Estrela para garantir a navegação em qualquer época do ano. Pretende-se instalar uma Central Hidrelétrica junto a esta barragem segundo inventário da ANEEL de 31/08/2018. O empreendimento encontra-se no estágio de “UHE com Outorga para Implantação”.

3.1.2. Clima da região e sua dinâmica

Há dois tipos de climas na área de estudo segundo a classificação climática de Köppen-Geiger (1933): subtropical (Cfa), nas altitudes baixas e médias; e temperado (Cfb), nas altitudes mais altas. Cfa faz referência ao clima com chuvas durante todos os meses do ano, temperatura do mês mais quente superior a 22 °C, e temperatura do mês mais frio superior a 3 °C. Cfb faz referência ao clima com chuvas durante todos meses do ano, temperatura do mês mais quente inferior a 22 °C, e temperatura do mês mais frio superior a 3 °C (RIO GRANDE DO SUL, 2017).

Na Figura 9, são apresentadas as normais de precipitação e temperatura para cada mês da estação de Bento Gonçalves do INMET (período de 1961-2013). Situada na bacia do rio Taquari-Antas, a cidade de Bento Gonçalves se encontra a leste de Encantado, em uma altitude de 690 metros. A temperatura média do mês mais quente (janeiro) é de quase 22 °C, o que corresponde ao extremo da categoria Cfb (temperado).

Figura 9. Precipitação mensal média, temperatura mensal máxima, temperatura mensal média e temperatura mensal mínima da série da estação Bento Gonçalves do INMET entre 01/01/1961 e 31/12/2013.



Fonte: produção própria do autor do trabalho utilizando dados da estação Bento Gonçalves, de código 83941, do INMET.

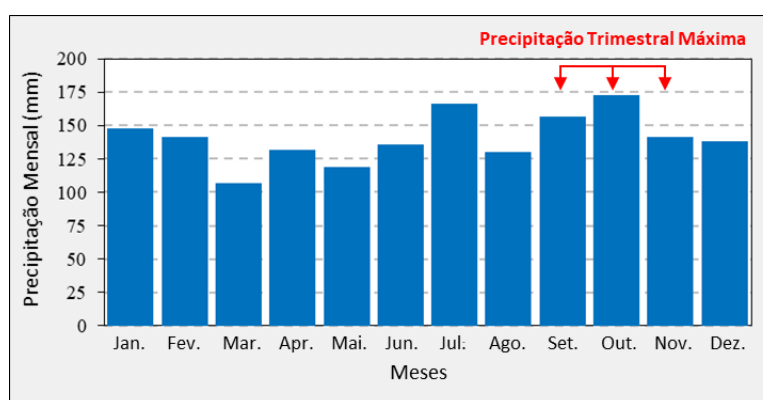
A dinâmica climática que define o clima da região Sul (RS, SC e PR) é governada sobretudo pela posição do centro de alta pressão do Atlântico Sul, por um sistema de baixa pressão continental intermitente (Baixa do Chaco) localizado na região nordeste da Argentina, e pela circulação anticiclônica associada. No verão, o centro de alta pressão continental está deslocado para sudeste, e

os ventos do Atlântico não penetram muito no continente. Nesse cenário de mais calma e maior entrada de radiação solar, é geralmente no final do dia que se formam complexos convectivos de mesoescala — CCM (CAVALCANTI *et al.*, 2009). Esses contribuem bastante para precipitação total desencadeada entre outubro e abril, o semestre quente. Essas chuvas costumam durar menos em comparação com aquelas formadas por sistemas frontais de precipitação. No inverno, as precipitações resultam de penetrações frontais (frentes frias), associadas a ciclones extratropicais migratórios e maior proximidade do centro de alta pressão do Atlântico (CAVALCANTI *et al.*, 2009). A inundação de 15/07/2015 em Lajeado, por exemplo, foi atribuída à ocorrência da precipitação gerada do encontro de um centro de baixa pressão, uma frente fria e uma frente quente. A rede UNIVATES apontou chuvas concentradas sobre o rio Forqueta da ordem de 65 mm no dia 13, e chuvas da ordem de 40 mm até 10 mm em um gradiente nordeste-sudoeste no dia 14 (KUREK, 2016).

O estado gaúcho está numa zona climática de transição muito sensível aos eventos de *El Niño* e *La Niña*, com a máxima precipitação ocorrendo em diferentes épocas do ano. Por esse motivo, a região Sul apresenta um regime pluviométrico aproximadamente uniforme, com precipitações relativamente mais volumosas no inverno, de junho a setembro (CAVALCANTI *et al.*, 2009).

Segundo Grimm *et al.* (1988), os três meses consecutivos de maior precipitação na bacia do rio Taquari-Antas são agosto-setembro-outubro. Kich *et al.* (2015) utilizaram dados de 15 estações pluviométricas para estudar a distribuição das chuvas na bacia do rio Taquari-Antas para o período de 1977 até 2006 (Figura 10), indicando que os meses consecutivos mais chuvosos costumam ser setembro-outubro-novembro.

Figura 10. Precipitação mensal média na bacia do rio Taquari-Antas/RS no período de 1977-2006, calculadas com 15 estações pluviométricas.



Fonte: Adaptado de Kich *et al.* (2015).

3.1.3. Hidrografia e inundações históricas

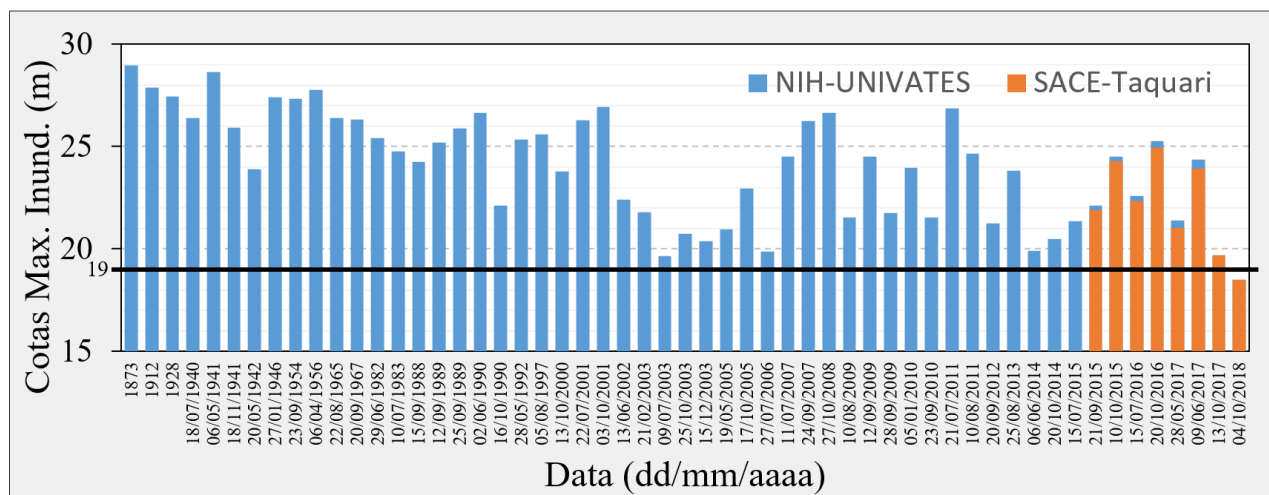
O rio Taquari-Antas faz parte da bacia do rio Guaíba (IBGE, 2013). Os principais afluentes do Taquari-Antas organizam-se de maneira radial (ver Figura 8), favorecendo a rápida concentração do escoamento, ocasionando frequentes inundações e enxurradas (CHAGAS *et al.*, 2014). Quatro dos

seus principais afluentes (rio Forqueta, rio Guaporé, rio Carreiro e rio Turvo) apresentam um tempo de concentração da ordem de 2 dias. Enquanto isso, as bacias hidrográficas delimitadas a partir de Estrela e Encantado apresentam um tempo de concentração da ordem de 3 dias. Ocorrendo uma precipitação esparsa sobre a bacia, a água tende a acumular no rio principal no mesmo dia, e a propagação da onda de cheia entre Encantado e Estrela é da ordem de 6 horas (KUREK, 2016).

A junção dos afluentes ocorre próximo de Encantado, e, de fato, é a partir daí até Bom Retiro do Sul o trecho crítico a inundações — em vermelho na Figura 8 — onde também há a eclusa já mencionada (DRH/SEMA, 2011). Neste trecho se encontram os seguintes municípios, nos quais se abrigam as respectivas populações segundo estimativa publicada no Diário Oficial da União de 01/07/2018: Encantado (22.520 pessoas); Roca Sales (11.300 pessoas); Arroio do Meio (20.637 pessoas); Colinas (2.442 pessoas); Lajeado (82.951 pessoas); Estrela (33.820 pessoas); e Cruzeiro do Sul (12.290 pessoas). Esses números somam 185.960 pessoas (BRASIL, 2018).

Lajeado e Estrela são municípios vizinhos e estão na margem direita e na margem esquerda do rio Taquari-Antas, respectivamente. Na Figura 11, é apresentada, em azul, a série histórica de cotas máximas de inundações em Lajeado no período de 1873 até 9/06/2017 fornecida pelo Núcleo de Informações Hidrometeorológicas (NIH) da Universidade do Vale do Taquari (UNIVATES). Também estão representadas, em laranja, as cotas máximas monitoradas no SACE-Taquari durante os eventos de inundação no período de 21/09/2015 até 08/11/2018.

Figura 11. Cotas máximas de inundação em Lajeado/RS, período de 1873-2018.



Fonte: adaptado dos dados do NIH-UNIVATES (2019) e do SACE-TAQUARI (CPRM, 2019).

Segundo o NIH da UNIVATES, cotas superiores a 19 metros são sinais de inundações. A série NIH-UNIVATES advêm de diferentes fontes, apresentando registros de inundações históricas anteriores a 2013, como a de 1873, que foram obtidos pela Proteção e Defesa Civil através de meios de comunicação da época e registros realizados por membros da comunidade lajeadense. As cotas

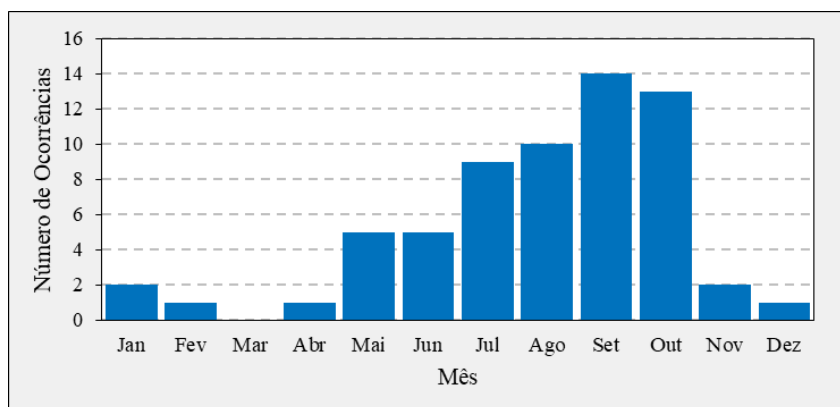
máximas de 21/09/2015 até 09/06/2017 da série NIH-UNIVATES foram obtidas de leituras dos linígrafos da UNIVATES. As últimas duas cotas máximas de 13/10/2017 e 04/10/2018 foram obtidas da rede do SACE-Taquari.

A discrepância entre medições no linígrafo da série NIH-UNIVATES e aquelas do SACE-Taquari no período entre 21/09/2015 e 9/06/2017 pode ser devido ao fato do SACE-Taquari monitorar dados em uma frequência horária, enquanto as medições de cota máxima por linígrafo em tempo atual tem maior resolução temporal. São dados com incertezas diferentes.

Na Figura 12, é apresentado o número de ocorrências de inundações em cada mês, obtido da série histórica de cotas máximas registradas durante eventos de inundação em Lajeado no período de 1940 até 2018 (NIH/UNIVATES, 2019). Essa série conta com 63 dados de cota de inundação máxima em 78 anos, com o primeiro dado datando de 18/7/1940 e o último, de 04/10/2010. O NIH considera que eventos que atingiram cota superior a 19 metros configuram inundações em Lajeado. Analisando as cotas máximas de Lajeado em Gerhard (2017), a série do posto de Estrela do SACE-Taquari e a série de Estrela do NIH-UNIVATES, identificou-se que a cota medida foi 85 vezes superior a 19 metros no período de 18/07/1940 até 05/11/2019.

Embora a maior inundação dessa série, que atingiu a cota de 29,92 m (ver Figura 11), tenha ocorrido num mês de maio (06/05/1941), as inundações tendem a ocorrer em agosto, setembro e outubro, que é no fim da estação das precipitações com grande duração, *i.e.* aquelas devido ao encontro de sistemas frontais de precipitação, que estacionam-na região.

Figura 12. Número de ocorrências de inundação por mês em Lajeado/RS, período de 1940-2018.



Fonte: Adaptado da série histórica do Núcleo de Informações Hidrometeorológicas da UNIVATES (2019).

Os dados em Eckhardt (2008) evidenciam o impacto econômico desses eventos. Duas inundações, ocorridas em julho e outubro de 2001, atingindo cotas de 26,30 m e 26,95 na cidade de Lajeado, trouxeram, respectivamente, perdas diretas (definição em 2.1) da ordem de R\$ 196.000,00 e R\$ 396.000,00 segundo à Proteção e Defesa Civil e a Prefeitura Municipal de Lajeado.

No quadro do SACE-Taquari, alguns limiares de referência para a operação da rede instalada foram identificados. Os valores para cota de atenção, de alerta e de inundação, presentes no portal do SACE no dia 28 de dezembro de 2018, são apresentados no Quadro 2 (CPRM/SGB, 2019).

Quadro 2. Limiares de referência para estações de Muçum, Encantado, Estrela e Taquari, retirados do portal SACE-Taquari no dia 05 de novembro de 2019.

Estação	Código	Cotas de Referência (cm)		
		Atenção	Alerta	Inundação
Muçum	MU	500	900	1.600
Encantado	ENC	500	900	1.100
Estrela	EST	1.400	1.500	1.700
Taquari	TA	350	650	850

Fonte: adaptado de SACE (cprm.gov.br/sace/index_bacias_monitoradas.php).

3.1.4. Geologia, pedologia e uso do solo na bacia

O mapa de unidades geológicas regional do RS na escala 1:750.000 foi utilizado para gerar a Figura 13.A (CPRM, 2020). Nele, 90,8% das rochas da bacia são da formação Serra Geral e estão nas zonas mais elevadas da bacia. Nas zonas mais baixas, a jusante de Estrela, estão outras unidades geológicas (2,6% de lamito, 1,8% de areia, e os restantes 4,8% sendo arenito e siltito).

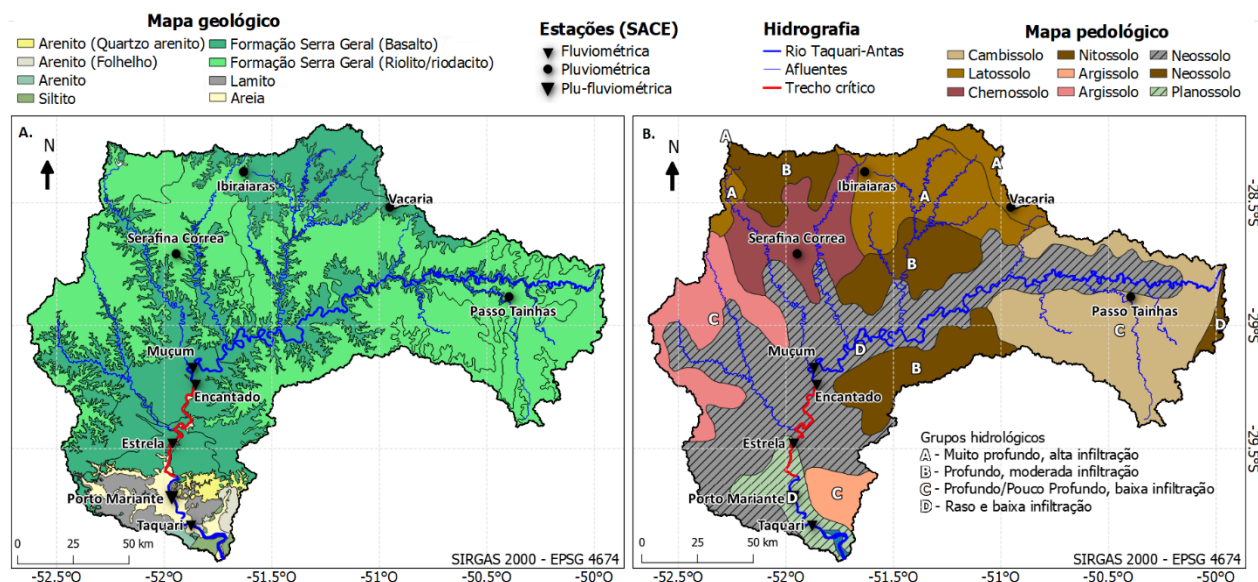
O mapa de solos do Brasil na escala 1:5.000.000 está apresentado na Figura 13.B. Há 7 ordens de solo: nas zonas de cabeceira, cambissolos, chernossolos, latossolos e nitossolos; nas zonas médias, argissolos; nos vales encaixados, neossolos e planossolos (Figura 13.B.). Esses solos foram classificados segundo os grupos hidrológicos A, B, C e D, que indicam a resposta hidrológica desses solos quanto a capacidade de armazenamento, infiltração e formação de escoamento de acordo com Sartori *et al.* (2005). Essa classificação é análoga àquela do CN do método SCS, porém adaptada para os solos brasileiros.

Os neossolos da região (classificados com D na Figura 13.B.) são rasos e pouco permeáveis, somam 34% dos solos da região e estão localizados justamente nas planícies de inundação dos rios. Os planossolos, que também são classificados com D e somam 4%, estão na semelhança em regiões de planície de inundação. A presença de solos rasos e pouco permeáveis são duas características que contribuem para os recorrentes cenários de inundações e enxurradas na bacia (CHAGAS *et al.*, 2014).

Baseado em mosaicos obtidos do LANDSAT, a organização MapBiomas mapeia anualmente a cobertura e uso do solo brasileiro (MapBiomas Project). O MapBiomas nasceu de uma pergunta: “É possível gerar anualmente mapas da cobertura e uso do solo para o Brasil inteiro de uma maneira mais barata, rápida e atualizável em comparação com as práticas e métodos já empregados, permitindo recuperar o histórico de décadas passadas?”. Desde então, o propósito da iniciativa, com origem em 2015, é contribuir para o conhecimento da dinâmica de uso do solo no Brasil e em países

tropicais, desenvolvendo uma metodologia rápida e barata para realizar esse mapeamento do ano de 1985 até o presente (atualizando anualmente). Essa iniciativa aproxima especialistas dos biomas brasileiros, formando uma rede de pesquisadores que estudam mudanças na dinâmica e uso do solo.

Figura 13. Representações gráficas do (A.) Mapa geológico regional das unidades geológico-ambientais obtido no GeoSGB da CPRM/SGB (2020) e do (B.) Mapa pedológico do Brasil seguindo o Sistema Brasileiro de Classificação de Solos (SiBCS, 2006) obtido no GeoINFO da EMBRAPA (2020).



Fonte: produzido no QGIS 3.10.4 adaptado das bases de dados nacionais.

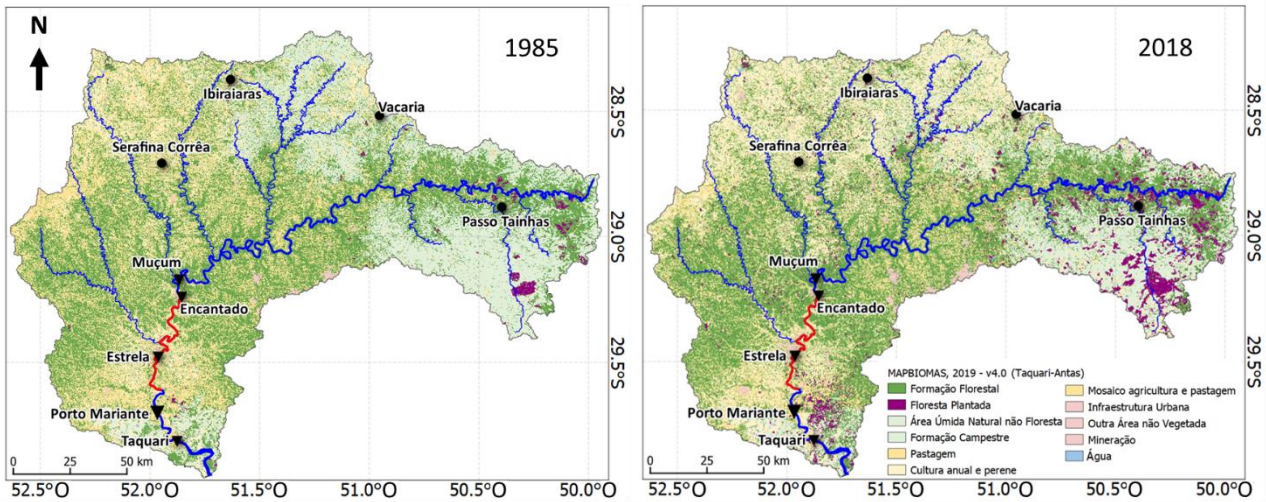
Na Tabela 2, referente à Figura 14, está apresentada a frequência relativa de classes do uso do solo na bacia para os anos de 1985, 2015 e 2018. Entre os anos de 2015 e 2018 houve uma mudança do uso do solo da bacia, em que aumentou a área de floresta plantada e a área de cultura anual e perene cerca de 5,6% e 10,2%, respectivamente.

Tabela 2. Uso do solo por categorias na bacia do rio Taquari-Antas nos anos de 1985, 2015 e 2018. ID é o número de identificação da classe utilizado no MapBiomas v4.0.

Classe (%)	ID	Ano		
		1985	2015	2018
Formação Florestal	3	40,58	41,20	40,81
Floresta Plantada	9	0,73	4,24	4,48
Área Úmida Natural não Florestal	11	0,00	0,01	0,00
Formação Campestre	12	20,97	13,64	12,88
Pastagem	15	13,88	8,53	8,55
Cultura Anual e Perene	19	9,30	19,25	21,21
Mosaico de Agricultura e Pastagem	21	13,75	11,58	10,44
Infraestrutura Urbana	24	0,44	1,01	1,09
Outra Área não vegetada	25	0,05	0,12	0,13
Mineração	30	0,00	0,00	0,00
Rio, Lago e Oceano	33	0,28	0,43	0,42

Fonte: adaptado do produto do MapBiomas v4.0 (2019).

Figura 14. Evolução no uso do solo na bacia do rio Taquari-Antas de 1985 para 2018.



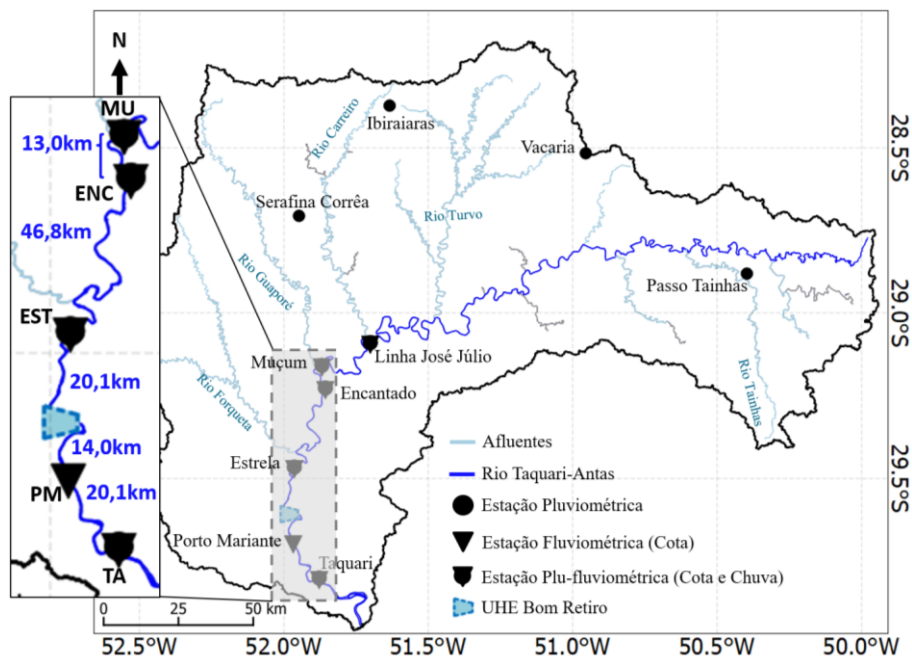
Fonte: adaptado do produto do MapBiomias v4.0 (2019).

Notam-se núcleos de floresta plantada aparecendo principalmente na parte leste da bacia e também a leste de Porto Mariante. Além disso, cultura anual e perene vem tomando espaço de formação campestre entre Ibiraiaras e Vacaria (ver Figura 14).

3.2. Estações utilizadas

Foram utilizados os dados horários de precipitação e nível das estações telemétricas instaladas a partir de 2015 no âmbito do SACE para bacia do rio Taquari-Antas. Esses dados foram fornecidos pelo Serviço Geológico Brasileiro (CPRM/SGB). A distribuição das estações e a hidrografia da região são ilustradas na Figura 15. Dados mais específicos das estações são apresentados na Tabela 3.

Figura 15. Distribuição das estações utilizadas.



Fonte: produção do autor do trabalho no software QGIS 3.10.4.

Tabela 3. Informações das estações do SACE-Taquari. Fluviométrica – “f”. Pluviométrica – “p”.

Estação	Código	Tipo	Coordenadas (Graus)		Altitude (m)	Área de Dren. (km ²)
			Lat.	Long.	<i>srtm30</i>	<i>srtm30</i>
Passo Tainhas	PT	p	-28,88269	-50,39594	872	
Vacaria	PV	p	-28,51750	-50,95361	972	
Ibiraiaras	PI	p	-28,37277	-51,63277	797	
Serafina Corrêa	SC	p	-28,70769	-51,94646	530	
Linha José Júlio	LJJ	f,p	-29,09807	-51,69956	91	
Muçum	UM	f,p	-29,16694	-51,86722	56	16.090
Encantado	ENC	f,p	-29,23519	-51,85507	30	19.200
Estrela	EST	f,p	-29,47349	-51,96281	17	23.290
Porto Mariante	PM	f	-29,69958	-51,96650	16	24.720
Taquari	TA	f, p	-29,80688	-51,87659	3	25.940

3.3. Análise dos dados

A análise de dados está dividida nas seguintes etapas: (3.3.1) Consistência dos dados e período utilizado; (3.3.2) Avaliação das transformações de variáveis; (3.3.3) Repartição dos registros; (3.3.4) Extremos da série.

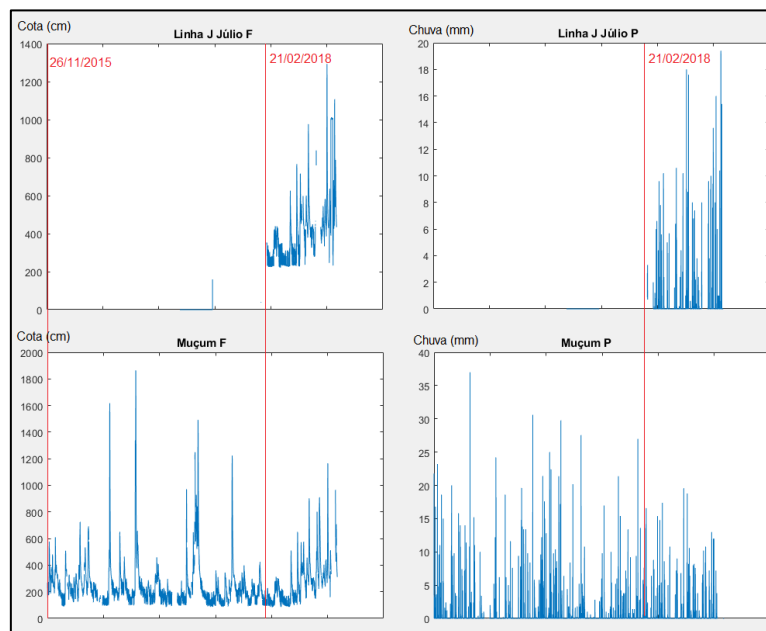
3.3.1. Consistência de dados e período utilizado

Já que algumas das estações têm resolução de 15 minutos, transformou-se os dados para frequência horária. Para a precipitação, os dados foram acumulados; para os níveis, os dados foram tomados de hora em hora, simplesmente.

Apesar dos dados de telemetria da CPRM serem automaticamente consistidos, segundo Chagas *et al.* (2014), os dados foram adicionalmente consistidos para garantir a ausência de informações falsas para as RNAs. Vistos alguns dados duvidosos, adotou-se uma regra de consistência de dados: precipitações horárias > 200 mm foram substituídas por *NaN* (*Not a Number*). Cada série tem cerca de 29.000 dados. O número de dados alterados de cada série pluviométrica é enunciado entre parênteses: Vacaria (1); Ibiraiaras (0); Serafina Correa (1); Muçum (12); Encantado (9); Estrela (1); Taquari (0).

O período utilizado de dados abrange desde 26/11/2015 às 19h até 24/04/2019 às 6h. A série de Linha José Júlio é consideravelmente menor do que as de outras estações, iniciando somente a partir de 21/02/2018 às 17h. Na Figura 16, são apresentadas as séries das estações de Linha José Júlio e Muçum. A série de Linha José Júlio não será, portanto, utilizada nesta pesquisa.

Figura 16. Comparação entre dados de Muçum e Linha José Júlio.



3.3.2. Avaliação das transformações de variáveis

Os níveis em Encantado, Estrela, Porto Mariante e Taquari são modelados a partir de variáveis explicativas que descrevam o estado hidrológico na bacia de contribuição respectiva. Correlações “nível-nível” costumam ser altas ($>0,8$) quando as estações estão espacialmente próximas e não há barreiras entre elas. Correlações “precipitação-nível” costumam ser quase nulas quando trabalhando em regime horário ou sub-horário, porém podem apresentar maior correlação quando não se compara a precipitação pontual com o nível, mas o regime de precipitação recente.

Ao invés de fornecer várias entradas horárias subsequente em diferentes neurônios de entrada, agregou-se temporalmente informações com filtros para obter uma assinatura temporal da série. Teoricamente, uma RNA poderia trabalhar com vários nós de entrada para cada intervalo horário, porém o modelo seria demasiadamente grande.

Os seguintes filtros foram empregados:

- Filtro de média móvel uniforme (FMMU) (descrito em 2.5.4)
- Filtro de média móvel exponencial (FMME) (descrito em 2.5.4)
- Filtro de diferença móvel (FDM)
- Filtro de média móvel gama (FMMG)

Cada filtro pode representar assinaturas temporais sob aspectos diferentes. Segue-se com a descrição do FDM e do FMMG, visto que os outros dois foram descritos anteriormente (2.5.4).

O FDM baseia-se na hipótese de que um pulso ocorrido em uma série de dados ocorre também em outra série de dados, porém com um atraso e uma possível dispersão. Com ele, é testado se a variação do nível em uma dada estação de jusante do momento “ t ” até “ $t + \text{alcance}$ ” pode ser explicada

por uma variação do nível em uma dada estação de montante do momento “ $t - \text{diferença}$ ” até “ t ”. Inspecciona-se a diferença que implica na maior correlação com o pulso a jusante. Apesar do tempo de propagação da onda não ser uniforme por depender de diversos fatores (altura da onda, vento, vazão, estação da seção fluvial durante o evento), um valor médio pode ser adotado.

O FMMG funciona a partir de uma convolução entre a série original e pesos ponderados segundo a função de densidade de probabilidade Gama 2. Este filtro requer os parâmetros de forma (F) e escala (E). A função Gama 2 pode ser moldada para dar pesos pequenos a dados recentes, dar pesos maiores a dados não tão recentes, e dar pesos pequenos a dados antigos, priorizando, em teoria, a precipitação de maior importância com maior flexibilidade do que o FMME. Já que a precipitação mais atual pode não ser importante para descrever uma cota, talvez o FMMG possa prescindir do parâmetro extra de defasagem.

O parâmetro extra de defasagem busca permitir que o instante “ t ” de montante possa ser menor ou igual ao instante “ t ” de jusante considerado para todos filtros. Os parâmetros dos filtros foram ajustados de modo a maximizar o coeficiente de correlação linear de Pearson entre a série filtrada e a diferença da cota de jusante entre o tempo atual “ t ” e o horizonte de previsão “ $t+\text{alc}$ ”.

Não necessariamente uma correlação alta implica na existência de uma relação física causa-efeito associada, mas há duas hipóteses físicas subentendidas aqui: (1) o nível de montante afetará, com certo atraso, o nível de jusante (*e.g.* ocorrida uma elevação do nível de montante, é esperada que a propagação desse sinal eleve o nível de jusante) e (2) o regime de precipitação afetará, com certo atraso, o nível de jusante.

Para as relações “nível-nível” foram inspeccionadas, inicialmente, quais as defasagens dos dados que resultavam maiores correlações, o que foi utilizado como um indicativo do tempo médio de propagação do escoamento. Posteriormente, analisou-se também com o uso dos filtros FDM, FMME e FMMU. Para as relações “precipitação-nível”, analisou-se o uso do FMMU, FMME e FMMG após notar que, realmente, a não agregação dos dados de precipitação retornavam correlações praticamente nulas.

Nesta pesquisa tentou-se limitar o número de registros necessários para a aplicação dos filtros para um período mais recente. No máximo poderia ser utilizado 30 registros. Os alcances inspeccionados foram de 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 horas. Cada coeficiente de correlação de Pearson (r) foi obtido com mais de 29.000 pares de dados. A escolha dos filtros é realizada conforme o coeficiente de correlação de Pearson e o número de horas necessárias para montar o filtro. Para o FMMG e FMME, por se tratar de duas funções com assíntotas horizontais, apresentou-se o número de registros necessários que somam 90% para o filtro (ao invés de número de registros total, como para o FMMU e FDM) para guiar a escolha do filtro.

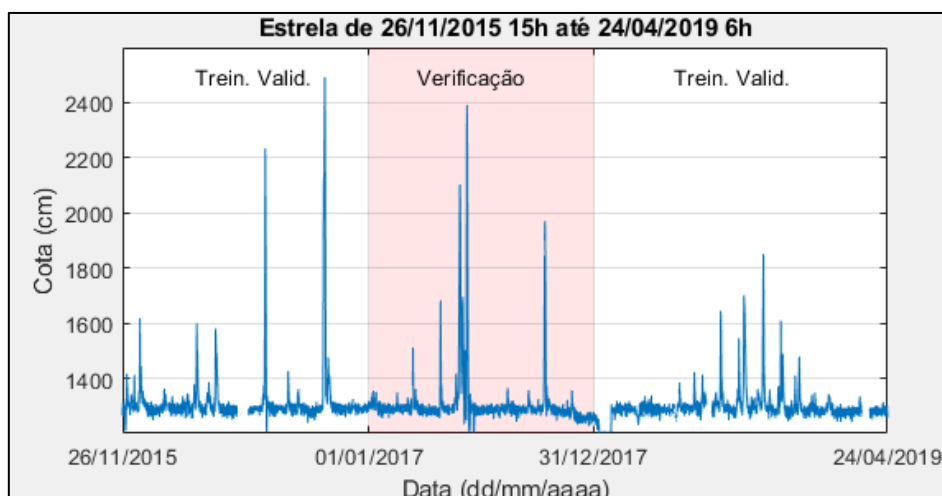
3.3.3. Repartição dos registros

Cada série explicativa, depois de transformada pelos filtros com os respectivos parâmetros ajustados para cada alcance, representa uma informação própria quanto ao estado hidrológico da variável representada. Ao associar as informações do estado da bacia (séries filtradas) com o estado futuro no alcance desejado, montam-se os “registros”. A RNA associará cada estado da bacia com a repercussão futura nos níveis. A metodologia aplicada para esse aprendizado necessita da repartição dos registros em conjunto de treinamento, validação e verificação (2.5.3).

O ano de 2017 foi selecionado como conjunto de verificação, pois visualmente não havia os eventos mais extremos das séries disponíveis, os quais devem estar presentes no conjunto de treinamento. Além disso, é um período com eventos representativos, isto é, com valores suficientemente extremos (e contínuos) para indicar como os modelos funcionariam em uma situação real de alerta. Os registros restantes foram repartidos entre treinamento e validação utilizando amostragem sistemática.

O procedimento da amostragem sistemática é determinístico e ordena os registros crescentemente em função dos valores de saída (nível em “t+alc”). Os registros que possuem pelo menos em uma das variáveis um dado de extremo são selecionados automaticamente para o treinamento. Do restante são selecionados os outros registros de treinamento de maneira igualmente espaçada. O conjunto de treinamento tem 2.000 registros. Todos registros restantes pertencem à validação. O número de registros de validação é diferente para cada alcance e número de neurônios de entrada porque basta um elemento do registro em branco para ele ser descartado. Na Figura 17, está ilustrada a repartição sob a perspectiva dos dados de Estrela. A região vermelha indica o período separado para a verificação.

Figura 17. Visualização gráfica da repartição dos períodos de treinamento e validação do período de verificação (2017) para a estação de Estrela.



3.3.4. Extremos da série

A série utilizada abrange desde 26/11/2015 às 14 h até 24/04/2019 às 6 h. No caso da modelagem por RNAs e também de outras modelagens empíricas, não é recomendável que se realizem extrapolações fora dos limites dos dados de treinamento. Os valores de mínimo e máximo das séries de cota para ENC, EST, PM e TA estão na Tabela 4.

Tabela 4. Valores de mínimo e máximo nas séries de cota para ENC, EST, PM e TA.

Estação	Código	Extremos (cm)	
		Mínimo	Máximo
Encantado	ENC	155	1.735
Estrela	EST	1.197	2.493
Porto Mariante	PM	87	1.467
Taquari	TA	169	1.053

Embora as réguas estejam possivelmente zeradas em altitudes diferentes, serve como indicativo a variação de 1495 cm entre extremos históricos registrados na estação Taquari (de código 86950000 no Hidroweb e dados de 01/03/1931 até 31/01/2019): o mínimo registrado é de 147 e o máximo é de 1.642 cm. A série de cotas máximas de inundação em Lajeado apresentada na Figura 11 traz registros de 2.897 cm para uma inundação em 1.873 e de 2.870 cm para a inundação de 6/5/1941. Essas informações expõem que há certas cotas não representadas ainda nas séries do SACE-Taquari.

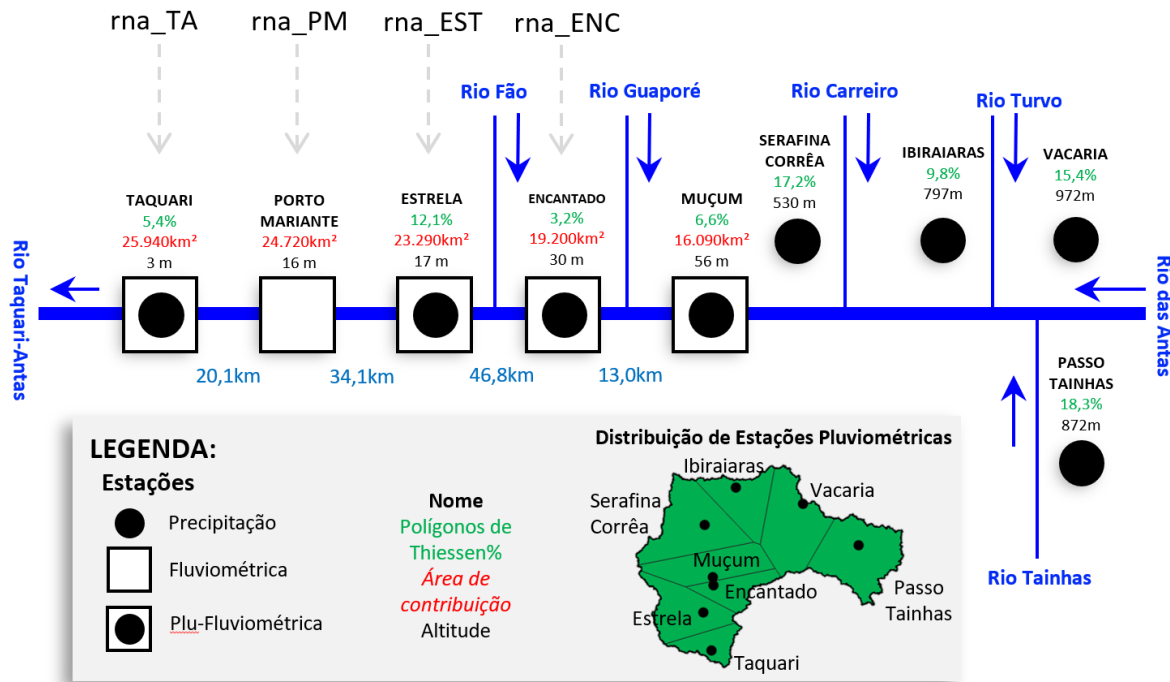
3.4. Desenvolvimento dos modelos

Já que o trecho crítico a inundações é a partir de Encantado (DRH/SEMA, 2011), decidiu-se prever para quatro localidades partindo de Encantado: Encantado (rna_ENC), Estrela (rna_EST), Porto Mariante (rna_PM) e Taquari (rna_TA), respectivamente de montante para jusante. Um diagrama unifilar foi confeccionado para sistematizar informações e auxiliar na montagem dos modelos (Figura 18).

Nesse trabalho, adotou-se o uso de Árvores de RNAs — modelos completos, CM (com dados de todas estações relevantes) e parciais, PM (quando há falha de transmissão em alguns equipamentos), como em Pedrollo (2017) — por causa da necessidade de modelos lidarem com cenários de eventual indisponibilidade de dados (PEDROLLO e PEDROLLO, 2013).

Os CMs possuem entradas com informações do estado da bacia a montante até o instante atual (precipitação e cota). A saída do modelo é a cota que será atingida na estação de jusante para qual se deseja uma previsão para diferentes alcances: 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 horas. As etapas do desenvolvimento dos modelos são descritas a seguir:

Figura 18. Diagrama unifilar das estações utilizadas, sintetizando informações sobre a área controlada e com as redes neurais pretendidas.



A primeira etapa consiste em identificar as estações de precipitação e níveis que caracterizam a bacia hidrográfica de previsão. Na segunda etapa, realiza-se o pré-processamento de dados com as defasagens e filtros ajustados em 3.3. Depois, os dados são sincronizados — *i.e.* todos sofrem translação temporal, considerando defasagens e alcances, para compor conjuntos ordenados de entrada-saída, chamados de registros, que representam o estado hidrológico da bacia (entrada) e o nível futuro (saída).

Na terceira etapa, segue-se com a identificação dos parâmetros de escalonamento das variáveis de entrada e saída para que fiquem com valores dentro do intervalo de sensibilidade das funções de ativação utilizadas. Salienta-se que esse é um passo importante, no caso das entradas, para evitar que o algoritmo de retropropagação dê vantagem às grandezas de maior valor, e, no caso das saídas, para adequação à imagem da função de ativação sigmoide unipolar, utilizada, neste trabalho, para o neurônio de saída.

Na quarta etapa, são eliminados registros ordenados com pelo menos uma falha — perda de sinal telemétrico. Os registros são separados nos conjuntos de treinamento, validação, e verificação. No treinamento, usam-se os dois primeiros conjuntos; o conjunto de verificação é o ano de 2017.

Na quinta etapa, pesquisa-se o número de registros de treinamento. O conjunto de validação é mantido fixo enquanto testa-se usar 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800 e 2000 registros de treinamento, a fim de verificar possíveis limites para o número de registros de treinamento. Esses registros são obtidos por amostragem sistemática a partir dos 2000 registros totais. Esta pesquisa busca verificar se há um limite de registros para o treinamento devido a redundâncias (que são comuns em dados experimentais e podem prejudicar o treinamento). Por sua vez, pode

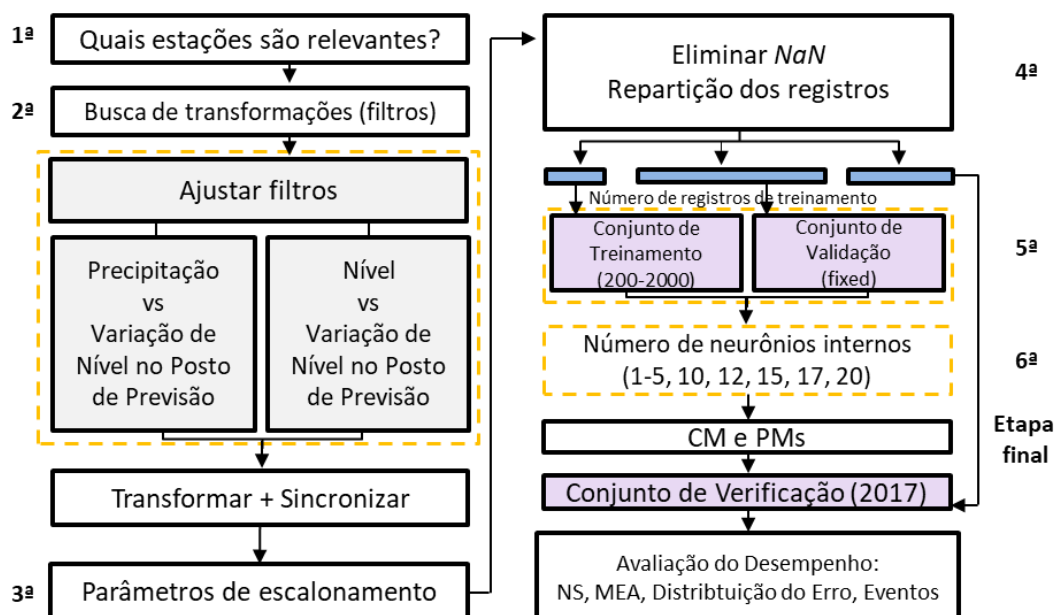
indicar de que talvez o padrão de proporções 2:1:1 recomendado na literatura nem sempre seja necessário (HECHT-NIELSEN, 1990).

Na sexta etapa, busca-se a complexidade interna necessária ao identificar a RNA com menor número de neurônios internos que apresente desempenho de validação semelhante ao de uma RNA superdimensionada, não sobreajustada, com excessivo número de neurônios internos (20 neurônios). Já que métricas de avaliação tem comportamento de variável aleatória, apresentando uma variabilidade, esta análise é realizada visualmente. Isso pode facilitar no discernimento, evitando equívocos de procedimentos automáticos.

Nas quinta e sexta etapas, são realizados treinamentos. Cada treinamento é realizado no mínimo 10 vezes, escolhendo-se a rede de menor erro quadrático na validação. As repetições visam evitar que se escolha pesos sinápticos advindos de um mau começo oriundo do sorteio dos pesos sinápticos iniciais. É comum que diferentes repetições tenham diferentes pesos sinápticos e ainda assim possuam capacidade de generalização. Quanto maior o número de repetições, maior é o investimento em tempo de processamento, pois a operação mais dispendiosa computacionalmente dessa pesquisa é justamente a propagação do sinal de entrada na rede e a retropropagação do erro durante o treinamento (multiplicar matrizes).

Por fim, os CMs são treinados com os parâmetros pesquisados durante a Análise dos dados (3.3), na quinta etapa (registros de treinamento) e na sexta etapa (neurônios internos). Os CMs e PMs são desenvolvidos com 10 repetições de treinamento e os mesmos parâmetros obtidos nas análises anteriores. A diferença é que os PMs tem menor número de variáveis de entrada que os CMs para que a Árvore de RNAs possa lidar com cenários de falta de informação em tempo atual. Um fluxograma da metodologia é apresentado na Figura 19.

Figura 19. Fluxograma da Metodologia. *NaN* – *Not a number*; CM – Modelo Completo; PM – Modelo Parcial.



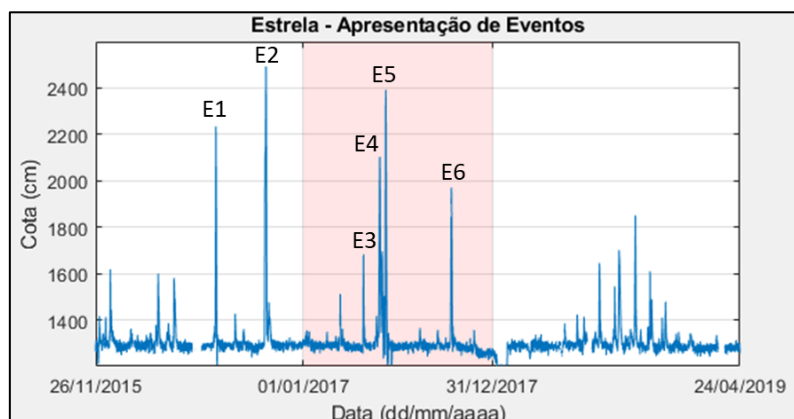
3.5. Avaliação do desempenho do modelo

O desempenho dos modelos é verificado com estatísticas obtidas no conjunto de verificação, que não é usado para nenhum procedimento de escolha da arquitetura, funções internas, complexidade, determinação de pesos sinápticos, ou quaisquer outros parâmetros (HECHT-NIELSEN, 1990). Foram adotados as seguintes métricas de avaliação descritas no item 2.6: NS, E10, E25, E50, E75, E90, MEA, MPEA, e os coeficientes de eficiência relativa em relação aos modelos hipotéticos de persistência e de extrapolação, sendo estes últimos calculados conforme Kitanidis e Brás (1978).

De maneira adicional, realizou-se uma análise visual dos hidrogramas observados e daqueles resultantes dos modelos para seis eventos para ter uma visão mais detalhada do comportamento dos modelos. Os picos dos eventos escolhidos e seu momento de ocorrência estão na lista seguinte e estão representados também na Figura 20:

- Evento 1 (E1) – 2234 cm em 15/07/2016 às 22 h
- Evento 2 (E2) – 2493 cm em 20/10/2016 às 16 h
- Evento 3 (E3) – 1682 cm em 27/04/2017 às 2 h
- Evento 4 (E4) – 2103 cm em 28/05/2017 às 12 h
- Evento 5 (E5) – 2392 cm em 09/06/2017 às 1 h
- Evento 6 (E6) – 1970 cm em 13/10/2017 às 23 h

Figura 20. Visualização dos seis eventos selecionados para demonstrar o desempenho dos modelos. E1 e E2 fazem parte dos conjuntos de treinamento e validação. E3, E4, E5 e E6 fazem parte do conjunto de verificação.



Os eventos 1 e 2 foram selecionados do conjunto de treinamento e validação justamente para averiguar se ocorreram diferenças sistemáticas de desempenho entre esses eventos e os outros do período de verificação. Se os eventos do ano 2017 tiverem desempenho da mesma ordem dos demais, a hipótese da capacidade de generalização dos modelos, a nível de eventos, fica comprovada. Assim, a verificação por eventos é complementar com vistas ao uso prático. A análise por eventos foi puramente gráfica para uma interpretação qualitativa. A abordagem foi de fazer previsões para todos os alcances estudados, em cada tempo atual, ao longo de cada evento, para cada estação.

4. RESULTADOS

Essa seção está dividida em (4.1) Análises das transformações das variáveis, (4.2) Pesquisa do número de registros de treinamento, (4.3) Pesquisa da Complexidade Interna, (4.4) Modelos completos de RNAs, (4.5) Modelos parciais de RNAs, (4.6) Análise por eventos com as Árvores de RNAs e (4.7) Discussão dos resultados com referência à literatura.

Apesar de algumas siglas já terem sido mencionadas anteriormente, os seguintes códigos são lembrados: as letras minúsculas antecedentes aos códigos de localidade “f”, “p” e “d” significam, respectivamente, fluviométrico (indica que se trata de uma variável de cota), pluviométrico (indica que se trata de uma variável de precipitação) e diferença de níveis (entre o tempo atual e o horizonte de previsão). O leitor pode recorrer a LISTA DE ABREVIATURAS E SIGLAS.

4.1. Análises das transformações das variáveis

Os resultados das análises preliminares indicam a defasagem horária entre os dados de nível (Tabela 5), a correlação dos filtros já ajustados para precipitação (Figura 21), a forma dos pesos ajustados do FMMG (Figura 22), e as correlações dos filtros ajustados para níveis de montante e jusante (Figura 23) e níveis da própria estação (Figura 24).

Tabela 5. Defasagem horária referente à maior correlação entre os dados de nível: estação (defasagem em horas, coeficiente de Pearson associado, distância em km até estação jusante).

Estação Jusante	Estação Montante (defasagem em horas; r associado; distância em km)
Encantado	fMU(1; 0,990; 13,0)
Estrela	fMU(5; 0,848; 59,8), fENC(3; 0,904; 46,8)
Porto Mariante	fMU(8; 0,938; 93,9), fENC(7; 0,930; 80,9), fEST(5; 0,787; 34,1)
Taquari	fMU(15; 0,854; 114,0), fENC(15; 0,854; 101,0), fEST(14; 0,745; 54,2), fPM(6; 0,908; 20,1)

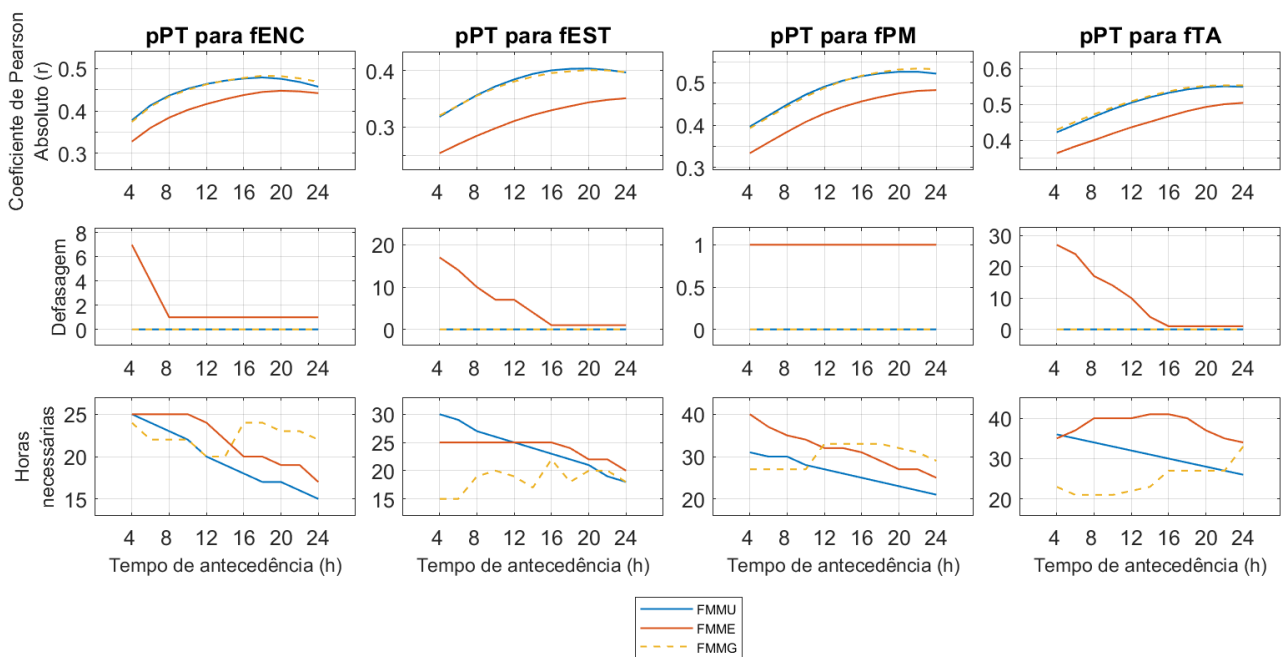
O naturalmente esperado é que quanto maior a distância entre os postos maior seja a defasagem e menor o coeficiente de correlação de Pearson devido a contribuições laterais. Na Tabela 5, os atrasos de maior correlação estão coerentes com a distância, mas as correlações não são sempre maiores quanto menor a distância entre as estações. Muçum e Encantado têm maior correlação com a cota em Porto Mariante do que em Estrela. Uma explicação para as menores correlações pode ser o barramento em Bom Retiro do Sul, que mantém a cota de 13,0 m para navegação em Estrela.

Por representar o mesmo comportamento averiguado nas outras estações investigadas, Passo Tainhas foi escolhida. Os resultados indicam a correlação entre a precipitação transformada e a variação da cota desde o tempo atual (t) até aquela no horizonte de previsão (t+alc) (Figura 21). Além

dos coeficientes de correlação, constam o número de registros necessários para o filtro e o parâmetro extra de defasagem.

As correlações “precipitações-nível” em Estrela foram menores que aquelas em Encantado, Porto Mariante e Taquari. O motivo pode advir da interferência da regularização dos 13,0 m pelo barramento de Bom Retiro do Sul. As correlações “precipitações-nível” foram decrescentemente menores em Taquari, Porto Mariante, Encantado e Estrela. O número de registros necessários do FMMU diminui conforme o alcance aumenta. O FMMG não precisou do parâmetro extra de defasagem, indicando uma possível maior flexibilidade do filtro.

Figura 21. Precipitação com FMMU, FMME e FMMG - Passo Tainhas e Localidade (ENC, EST, PM, TA).

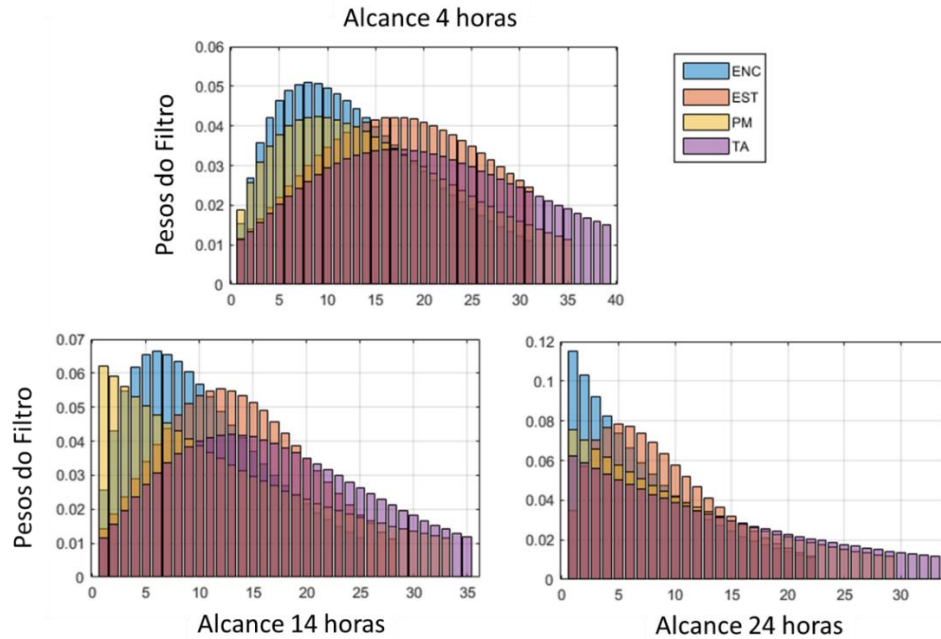


A diferença entre o r de FMMU e FMMG não é clara na Figura 21, mas o número de horas necessárias (3.3.2) para montar os filtros é menor com FMMG em alcances pequenos. Nesse critério de desempate, a função Gama foi melhor.

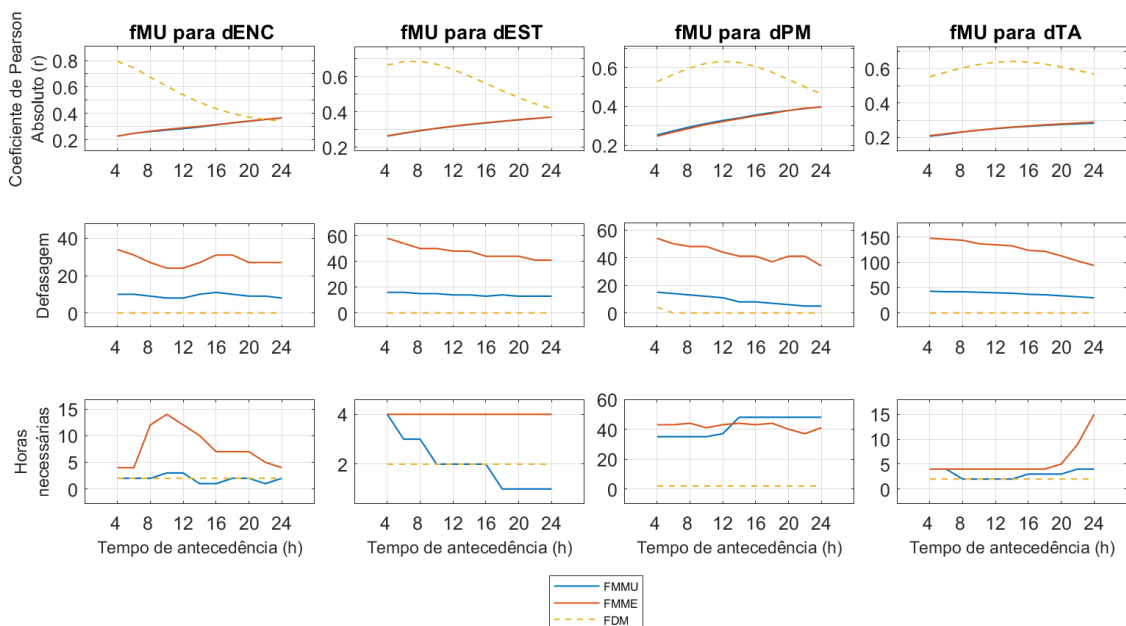
Quanto maior o alcance, a precipitação de maior peso costuma estar mais próxima do tempo atual para o FMMG (Figura 22). Uma figura semelhante para o FMMU mostraria o que pode ser visto na Figura 21: o FMMU tem menor número de registros necessários para alcances maiores em comparação com o FMMG porque em sua rotina é permitido buscar uma janela fora da qual os pesos são zero. A rotina do FMMG utiliza os pesos de uma função Gama 2 até que o peso seja ‘0,01’.

A similaridade nas correlações ao utilizar FMMU e FMMG implicou em realizar a próxima etapa testando os três filtros de precipitação: FMMU, FMME e FMMG. Na etapa seguinte (4.2), RNAs são desenvolvidas e a não-linearidade pode trazer mais indícios escolha entre FMMU e FMMG. Montar redes para as três alternativas também é uma maneira validar a etapa preliminar de selecionar variáveis com o coeficiente de correlação linear de Pearson e talvez reforçar Oliveira *et al.* (2017).

Figura 22. Pesos do FMMG para Passo Tainhas nos alcances de 4, 14 e 24 horas.



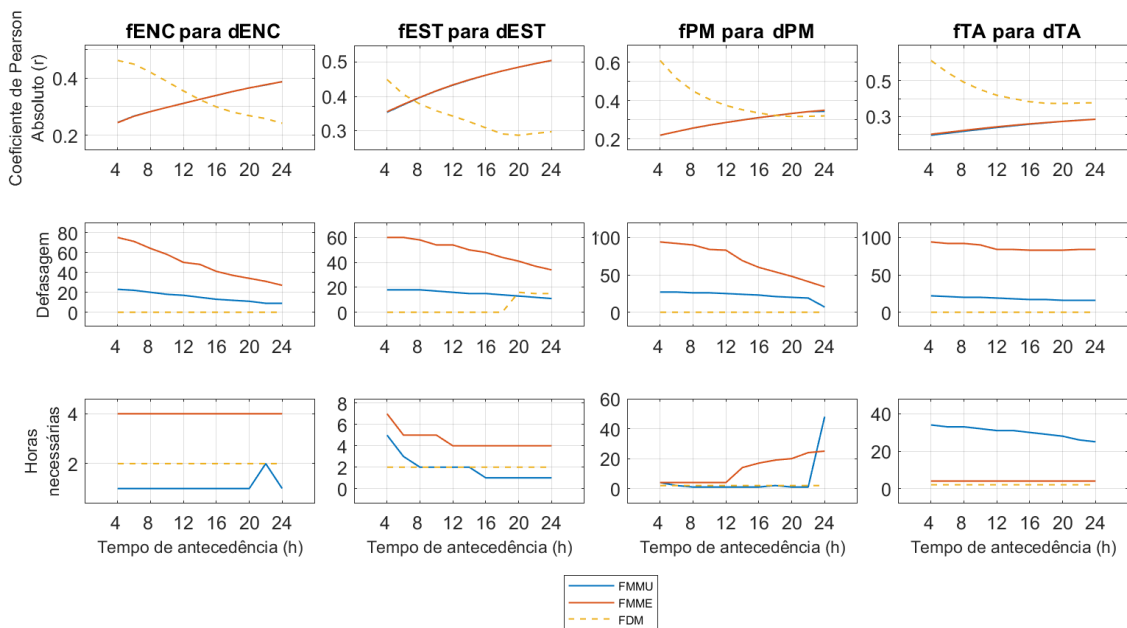
Os efeitos dos filtros para níveis são demonstrados de maneira resumida sob duas perspectivas. Na Figura 23, consta como os filtros aplicados para a estação de Muçum se correlacionam com a mudança de nível em cada localidade (Encantado, Estrela, Porto Mariante e Taquari) para os alcances estudados. O comportamento de Muçum é representativo para estações de jusante. Na Figura 24, está demonstrada a correlação que pode ser obtida quando níveis passados da mesma estação para qual se pretende a previsão são filtrados (f_{ENC} , f_{EST} , f_{PM} e f_{TA}) e comparados com as mudanças de nível até o alcance (d_{ENC} , d_{EST} , d_{PM} e d_{TA}).

Figura 23. Análise dos filtros nos níveis de Muçum e diferença, t e $t+alc$, entre níveis das localidades de previsão (ENC, EST, PM, TA) - Pesquisa dos filtros FMMU, FMME e FDM.

O parâmetro extra “def” para o FDM resultou quase sempre nulo, indicando que o nível no tempo atual é um dado bastante importante. O parâmetro “dif”, a diferença em horas entre dois níveis, que melhor explica a mudança do nível no local de previsão do tempo atual até o alcance, foi tipicamente baixo (da ordem de 1-4 horas independente do alcance) quando a correlação foi pesquisada com dados da mesma estação (caso da Figura 24). O mesmo parâmetro para as correlações entre estações diferentes indicou que, quanto mais distantes as estações estão, maior a diferença em horas encontrada, o que têm sentido físico. Houve uma leve tendência do parâmetro “dif” diminuir conforme o alcance aumenta para essas correlações de diferenças de níveis, mas não se considerou um efeito suficientemente forte para tirar maiores conclusões.

A correlação do FDM (vermelho na Figura 23 e Figura 24) é consideravelmente maior do que com FMME e FMMU, os quais podem estar indicando outro atributo presente nos dados. Nessa pesquisa, porém, manteve-se no escopo de utilizar apenas um tipo de filtro para cada tipo de dado. Portanto, o FMME e FMMU para níveis foram desconsiderados para as RNAs dessa pesquisa.

Figura 24. Nível – Análise de filtros nos níveis da localidade e diferença (t e t+alc) entre níveis das localidades de previsão (ENC, EST, PM, TA) - Pesquisa dos filtros FMMU, FMME e FDM.

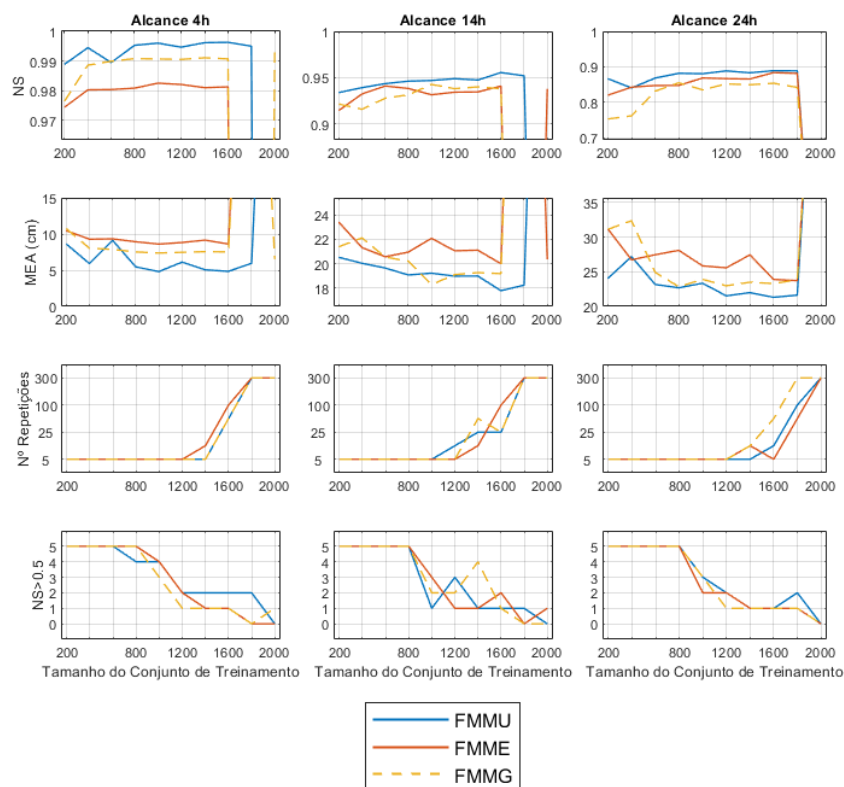


4.2. Pesquisa do número de registros de treinamento

Nem sempre a proporção recomendada na literatura (2:1:1 para treinamento, validação e verificação) vem apresentando melhor desempenho nos modelos de RNA. Portanto, foi realizada a pesquisa do número de registros de treinamento. Usou-se os três tipos de filtros para precipitação porque houve dúvidas entre FMMU e FMMG, esperando que a não-linearidade das RNAs evidencie vantagens não visualizáveis na análise linear anterior 4.1.

Nessa etapa, para os alcances de 4, 14 e 24 horas, é apresentado o efeito do número de registros de treinamento nas RNAs em Encantado (Figura 25) e Taquari (Figura 26), que demonstram o comportamento geral observado. Ressalta-se que o conjunto de validação esteve fixo para evitar comparações injustas. As métricas NS e MEA estão, respectivamente, na primeira e segunda linhas das figuras apresentadas e indicam também a diferença entre FMMU, FMME e FMMG nas RNAs. A terceira e quarta linha das figuras apresentadas são excepcionalmente importantes para escolha do número de registros de treinamento. A terceira linha é o número de repetições empregada para o treinamento. A quarta linha indica quantas repetições atingiram NS superior a 0,5.

Figura 25. Pesquisa do número de registros de treinamento para a localidade de Encantado para os alcances de 4, 14 e 24 horas.

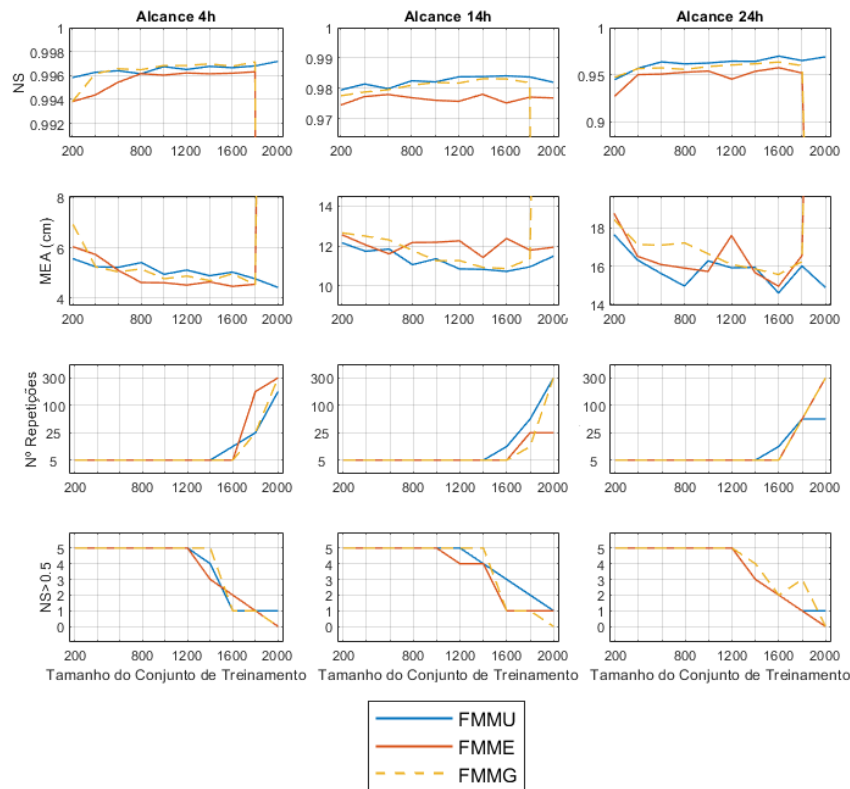


Dependendo do número de registros de treinamento, nenhuma das 5 repetições iniciais realizou mais que cerca de 20 ciclos de treinamento. O treinamento parou prematuramente pelo critério da validação cruzada. Assim, foi adicionada uma nova condição de treinamento: Caso após 5 repetições nenhuma delas retorne um treinamento satisfatório (aqui representado por um $NS > 0,5$), deve-se refazer o procedimento de treinamento com 10 repetições. Falhando-se repetidamente em encontrar um caminho, 25, 50, 100, 150 e, por fim, 300 repetições foram empregadas. Mesmo 300 repetições não bastaram, às vezes, para que com 2000 registros de treinamento fosse encontrado um “caminho de treinamento”.

Algumas repetições com maior número de registros (>1600) pararam prematuramente e têm métricas claramente piores (NS próximo de 0). Foi escolhido comprometer a visualização dos

extremos das métricas de NS e MEA, para investigar, na região de melhor desempenho (400-1600 registros), se realmente havia alguma tendência de algum dos filtros de precipitação ser melhor que os outros (FMMU, FMME e FMMG).

Figura 26. Pesquisa do número de registros de treinamento para localidade de Taquari para os alcances de 4, 14 e 24 horas.



As curvas apresentadas demonstram que o desempenho das redes neurais no conjunto de validação aumenta conforme mais dados de treinamento estão disponíveis. Porém, também, o grande número de registros de treinamento dificulta que o algoritmo de treinamento encontre um mínimo satisfatório para ajustar os pesos sinápticos. Isso se evidencia ao analisar o número de repetições aplicado e as repetições $NS > 0,50$ (terceira e quarta linhas das imagens a seguir).

Em geral, o FMMU apresentou melhor desempenho para todas as localidades (Encantado, Estrela, Porto Mariante e Taquari) e alcances (4-24 horas), mas FMMG necessitou menor número de horas, apresentando desempenho semelhante. Optou-se por seguir o desenvolvimento das redes com FMMG por uma razão operacional: um menor número de horas necessárias pode ser mais importante para lidar com o impasse de falta de informação momentânea.

Talvez o mais interessante desses resultados é que a partir de um certo número de registros de treinamento algumas das 5 repetições não conseguiram ser satisfatoriamente treinadas. Esse acontecimento levou à definição do seguinte critério para definir o número de registros de treinamento a ser utilizado: maior número de registros de treinamento em que, para todos os alcances, pelo menos

4 das 5 repetições apresentaram treinamento satisfatório. Por exemplo, Encantado (ver Figura 25) foi escolhida com 800 registros e Taquari (ver Figura 26), com 1200 (Tabela 6).

Tabela 6. Número de registros de treinamento adotado para cada localidade.

rna	nt
ENC	800
EST	1200
PM	1200
TA	1200

4.3. Pesquisa da Complexidade Interna

As RNAs são treinadas com os filtros e o número de registros de treinamento resultantes das análises anteriores. A pesquisa da complexidade interna busca o número de neurônios internos necessários. Na Figura 27 e Figura 28, apresentam-se os resultados para Encantado e Taquari, que representam o comportamento geral visualizado.

Aumentando-se o número de neurônios internos progressivamente a partir de 1, toma-se como hipótese que se chegará a um platô de desempenho semelhante ao da rede superdimensionada (20 neurônios internos) porque há um o número de neurônios internos suficiente para descrever a não-linearidade dos registros entrada-saída. Cada rede neural nas figuras subsequentes é resultado de um treinamento com 5 repetições, em que a repetição com menor erro quadrático para o conjunto de validação foi escolhida, o que pode implicar em uma incerteza: se outra repetição fosse escolhida, seria o desempenho muito diferente? Possivelmente há uma banda aleatória já que as figuras demonstram uma certa variabilidade no referido platô hipotético.

O número de neurônios internos mínimos para atingir o patamar teórico foi identificado para cada localidade: Encantado (Figura 27); Estrela; Porto Mariante; Taquari (Figura 28). O número de neurônios internos é apresentado na Tabela 7.

Figura 27. Pesquisa da complexidade interna para a RNA da localidade de Encantado para os alcances de 4, 6, 12, 16 e 24 horas.

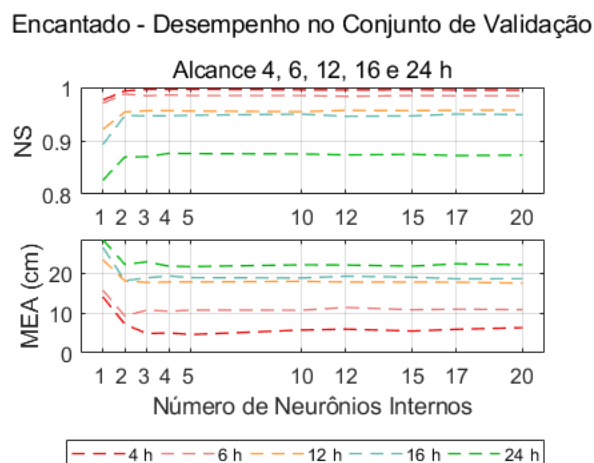


Figura 28. Pesquisa da complexidade interna para a RNA da localidade de Taquari para os alcances de 4, 6, 12, 16 e 24 horas.

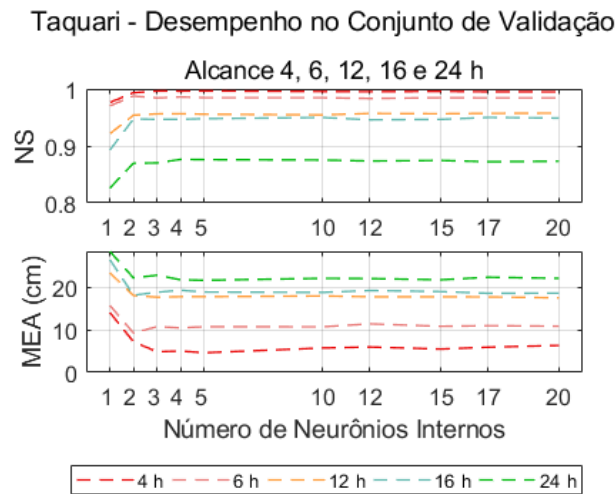


Tabela 7. Número de neurônios internos adotado para cada localidade.

rna	nni
ENC	4
EST	4
PM	4
TA	5

4.4. Modelos completos de RNAs

Os dados de entrada e saída, com as respectivas transformações realizadas, constam nas Tabela 8, 10, 12 e 14. Ao lado de cada tabela, há ilustrações da RNA (Figura 29, 30, 31 e 32). As métricas de desempenho (NS, MEA, MPEA, E10, E25, E50, E75, E90 e os coeficientes de eficiência relativa ao modelo de persistência e extrapolação) são apresentadas nas Tabela 9, 11, 13 e 15.

O número máximo de ciclos de treinamento foi mantido em 2.000.000 ciclos porque em alguns casos raros na pesquisa da complexidade interna (4.3), em alcances pequenos, a rede não obteve parada pela validação cruzada, indicando que a generalização não estava sendo comprometida. As estações com dados de nível, relevantes ao modelo como dado de entrada, são utilizadas duas vezes: (1) nível com defasagem obtida da análise com o FDM (2) nível no tempo atual. O número de neurônios internos é o mesmo para todos os alcances.

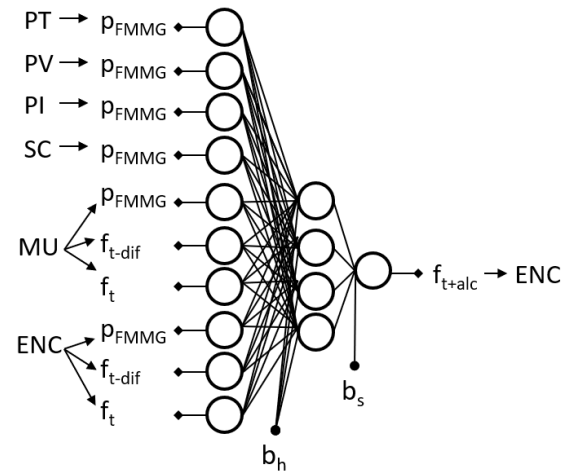
4.4.1. Encantado

As variáveis de entradas e suas transformações utilizadas para o modelo preliminar rna_ENC (Encantado) são apresentadas na Tabela 8. O modelo é ilustrado na Figura 29.

Tabela 8. Tabela de apresentação, variáveis de entrada e saída para rna_ENC (Encantado). PT – Passo Tainhas, PV – Vacaria, PI – Ibirairaras, SC – Serafina Correa, MU – Muçum, EN – Encantado, f – Nível, p – Precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.

Variável	Tipo	Filtro	Instante
pPT	1	Entrada	FMMG
pPV	2	Entrada	FMMG
pPI	3	Entrada	FMMG
pSC	4	Entrada	FMMG
pMU	5	Entrada	FMMG
pEN	6	Entrada	FMMG
fMU	7	Entrada	t - dif
fMU	8	Entrada	t
fENC	9	Entrada	t - dif
fENC	10	Entrada	t
fENC	11	Saída	t + alc

Figura 29. Ilustração da rna_ENC , referente à Tabela 8. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).



As métricas de avaliação para o CM rna_ENC são apresentadas na Tabela 9.

Tabela 9. Métricas de avaliação da verificação para a rna_ENC . Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.

Modelo	Alc	E10	E25	E50	E75	E90	MEA	MPEA	NS	CEr Persist.	CEr Ext. Linear
rna_ENC	4 h	-5,8	-2,8	0,8	4,5	8,2	5,2	0,020	0,992	0,720	0,979
rna_ENC	8 h	-17,9	-9,9	-0,3	12,6	24,0	14,6	0,058	0,958	0,457	0,948
rna_ENC	12 h	-25,8	-12,7	1,7	18,5	32,6	19,7	0,080	0,930	0,461	0,955
rna_ENC	16 h	-27,8	-14,4	-1,6	12,9	25,4	20,3	0,074	0,834	0,104	0,913
rna_ENC	20 h	-27,9	-13,0	0,6	13,4	26,4	19,2	0,074	0,914	0,654	0,965
rna_ENC	24 h	-36,1	-15,8	-2,5	8,6	23,2	21,9	0,081	0,878	0,640	0,949

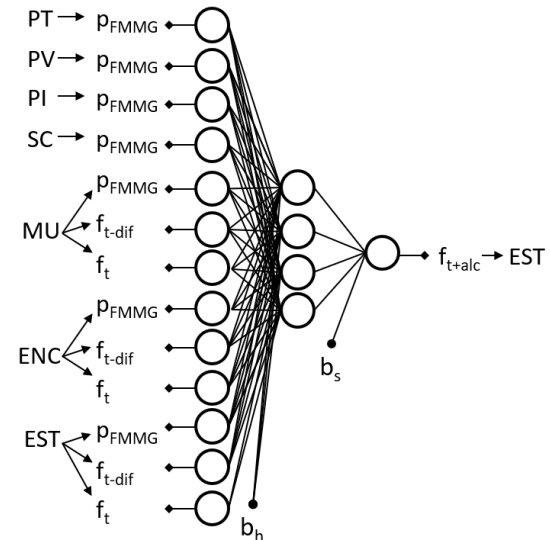
4.4.2. Estrela

As variáveis de entradas e suas transformações utilizadas para o modelo preliminar rna_EST (Estrela) são apresentadas na Tabela 10. O modelo é ilustrado na Figura 30.

Tabela 10. Tabela de apresentação, variáveis de entrada e saída para rna_EST (Estrela). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.

Variável	Tipo	Filtro	Instante
pPT	1	Entrada	FMMG
pPV	2	Entrada	FMMG
pPI	3	Entrada	FMMG
pSC	4	Entrada	FMMG
pMU	5	Entrada	FMMG
pEN	6	Entrada	FMMG
pES	7	Entrada	FMMG
fMU	8	Entrada	t - dif
fMU	9	Entrada	t
fENC	10	Entrada	t - dif
fENC	11	Entrada	t
fEST	12	Entrada	t - dif
fEST	13	Entrada	t
fEST	14	Saída	t + alc

Figura 30. Ilustração da rna_EST, referente à Tabela 10. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).



As métricas de avaliação para o CM rna_EST são apresentadas na Tabela 11.

Tabela 11. Métricas de avaliação da verificação para a rna_EST. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.

Modelo	Alc	E10	E25	E50	E75	E90	MEA	MPEA	NS	CEr Persist.	CEr Ext. Linear
rna_EST	4 h	-14,4	-7,6	-0,8	5,8	11,1	8,3	0,006	0,949	0,426	0,961
rna_EST	8 h	-22,4	-10,4	0,0	9,3	16,8	12,9	0,010	0,859	0,418	0,954
rna_EST	12 h	-27,7	-13,7	-1,2	8,3	16,8	14,9	0,011	0,814	0,466	0,956
rna_EST	16 h	-29,8	-15,0	-2,2	7,2	16,6	16,0	0,012	0,765	0,465	0,954
rna_EST	20 h	-36,2	-19,9	-8,8	2,0	10,8	17,4	0,014	0,792	0,837	0,961
rna_EST	24 h	-26,8	-14,9	-5,0	5,2	14,5	14,7	0,011	0,394	0,285	0,864

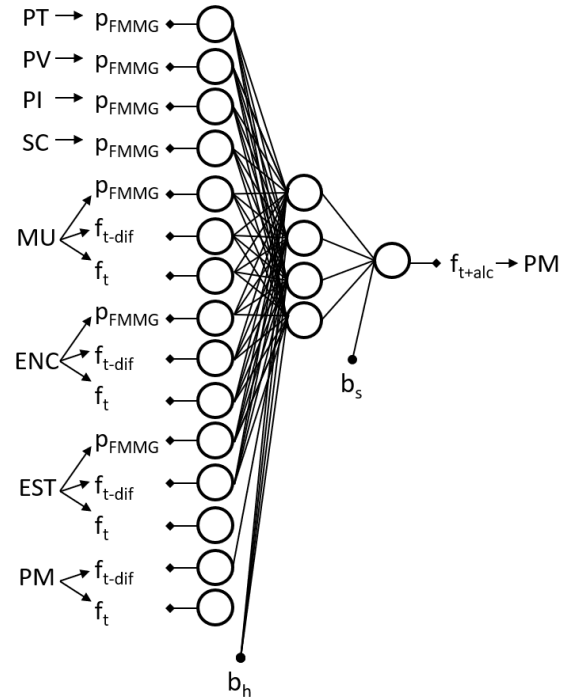
4.4.3. Porto Mariante

As variáveis de entradas e suas transformações utilizadas para o modelo preliminar rna_PM (Porto Mariante) são apresentadas na Tabela 12, O modelo é ilustrado na Figura 31.

Tabela 12. Tabela de apresentação, variáveis de entrada e saída para rna_PM (Porto Mariante). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, PM – Porto Mariante, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.

Variável	Tipo	Filtro	Instante
pPT	1	Entrada	FMMG
pPV	2	Entrada	FMMG
pPI	3	Entrada	FMMG
pSC	4	Entrada	FMMG
pMU	5	Entrada	FMMG
pEN	6	Entrada	FMMG
pES	7	Entrada	FMMG
fMU	8	Entrada	$t - dif$
fMU	9	Entrada	t
fENC	10	Entrada	$t - dif$
fENC	11	Entrada	t
fEST	12	Entrada	$t - dif$
fEST	13	Entrada	t
fPM	14	Entrada	$t - dif$
fPM	15	Entrada	t
fPM	16	Saída	$t + alc$

Figura 31. Ilustração da rna_PM , referente à Tabela 12. Cada FMMG depende de E e F , os quais variam com o alcance e a estação (ver 3.3.2).



As métricas de avaliação para o CM rna_PM são apresentadas na Tabela 13.

Tabela 13. Métricas de avaliação da verificação para a rna_PM . Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.

Modelo	Alc	E10	E25	E50	E75	E90	MEA	MPEA	NS	CEr Persist.	CEr Ext. Linear
rna_PM	4 h	-10,3	-4,9	1,2	8,3	16,0	8,7	0,036	0,984	0,669	0,956
rna_PM	8 h	-18,2	-9,4	-0,1	11,0	24,4	13,8	0,057	0,958	0,648	0,950
rna_PM	12 h	-26,1	-11,3	1,9	14,5	29,2	18,5	0,071	0,919	0,548	0,932
rna_PM	16 h	-31,4	-10,9	2,1	15,5	31,8	22,8	0,083	0,856	0,334	0,923
rna_PM	20 h	-41,1	-16,3	0,0	16,1	34,7	26,8	0,095	0,837	0,354	0,945
rna_PM	24 h	-54,1	-22,3	-3,1	17,1	36,4	30,6	0,110	0,829	0,485	0,950

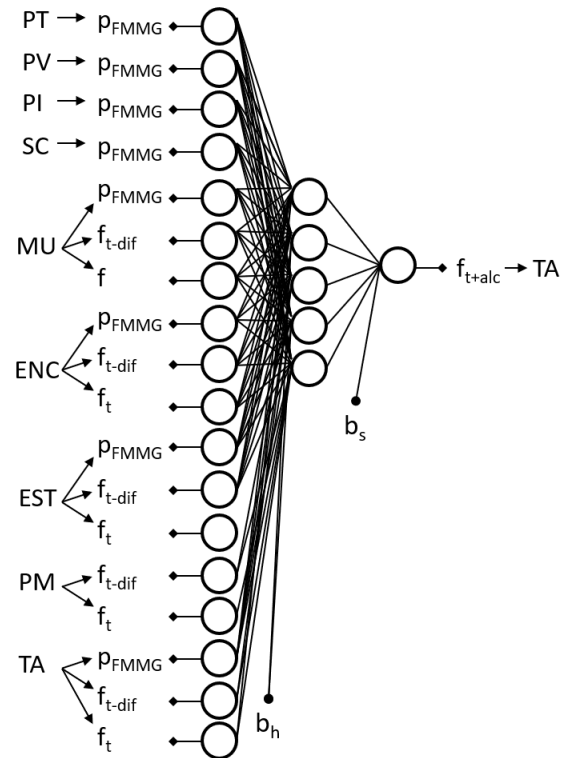
4.4.4. Taquari

As variáveis de entradas e suas transformações utilizadas para o modelo preliminar rna_TA (Taquari) são apresentadas na Tabela 14. O modelo é ilustrado na Figura 32.

Tabela 14. Tabela de apresentação, variáveis de entrada e saída para rna_TA (Taquari). PT – Passo Tainhas, PV – Vacaria, PI – Ibiraiaras, SC – Serafina Correa, MU – Muçum, ENC – Encantado, EST – Estrela, PM – Porto Mariante, TA – Taquari, f – nível, p – precipitação, dif – parâmetro do FDM, alc – alcance de previsão, FMMG – Filtro de Média Móvel Gama, t – Tempo atual.

Variável	Tipo	Filtro Aplicado	Instante Temporal
pPT	1	Entrada	FMMG
pPV	2	Entrada	FMMG
pPI	3	Entrada	FMMG
pSC	4	Entrada	FMMG
pMU	5	Entrada	FMMG
pEN	6	Entrada	FMMG
pES	7	Entrada	FMMG
pTA	8	Entrada	FMMG
fMU	9	Entrada	t - dif
fMU	10	Entrada	t
fEST	11	Entrada	t - dif
fEST	12	Entrada	t
fEST	13	Entrada	t - dif
fEST	14	Entrada	t
fPM	15	Entrada	t - dif
fPM	16	Entrada	t
fTA	17	Entrada	t - dif
fTA	18	Entrada	t
fTA	19	Saída	t + alc

Figura 32. Ilustração da rna_TA, referente à Tabela 14. Cada FMMG depende de E e F, os quais variam com o alcance e a estação (ver 3.3.2).



As métricas de avaliação para o CM rna_TA são apresentadas na Tabela 15.

Tabela 15. Métricas de avaliação da verificação para a rna_TA. Alc – alcance; E10, E25, E50, E75, E90 – quantis da distribuição do erro; MEA – média do erro absoluto; MPEA – média em percentil do erro absoluto; NS – coeficiente de desempenho de Nash-Sutcliffe; CEr Persist. – coeficiente de eficiência relativa de persistência; CEr Ext. Linear – coeficiente de eficiência relativa de extrapolação linear.

Modelo	Alc	E10	E25	E50	E75	E90	MEA	MPEA	NS	CEr Persist.	CEr Ext. Linear
rna_TA	4 h	-5,5	-2,7	0,5	4,0	9,3	4,5	0,017	0,978	0,053	0,912
rna_TA	8 h	-10,2	-4,0	1,7	7,7	15,2	7,8	0,028	0,938	0,051	0,887
rna_TA	12 h	-13,5	-5,9	2,3	10,0	20,0	10,8	0,039	0,878	-0,162	0,866
rna_TA	16 h	-14,0	-5,1	4,1	11,6	23,3	11,6	0,041	0,863	-0,012	0,889
rna_TA	20 h	-17,5	-6,9	1,9	10,5	24,3	12,8	0,044	0,806	-0,175	0,872
rna_TA	24 h	-20,4	-8,6	1,0	13,0	25,5	14,6	0,051	0,754	-0,174	0,866

4.4.5. Síntese dos modelos completos

Exceto para a localidade de Estrela, os modelos de previsão obtiveram NS maior ou igual a 0,83 para horizontes até 24 horas. As distribuições dos erros tenderam à simetria, com exceção de Estrela que, com uma distribuição com cauda à esquerda, exibiu tendência a subestimar os níveis. Os modelos de RNA superam os modelos hipotéticos de persistência e extrapolação linear na maioria dos casos. Em alcances maiores (acima de 12 horas), o modelo de persistência explicou melhor a variabilidade (maior NS) em Taquari, que é a maior bacia dessa pesquisa e é esperado um nível mais persistente. Para maiores alcances, a variação de nível pode depender de chuvas ainda não ocorridas no tempo atual (informação futura).

4.5. Modelos parciais de RNAs

Os modelos parciais (PM) foram treinados para lidar com eventuais indisponibilidades de dados — *i.e.* para realizar previsões em qualquer momento de falta de informação dos seis eventos identificados, com vistas a gerar imagens para análise por evento. Com os PMs aproveita-se para averiguar a importância de cada uma das variáveis dessas para o modelo completo (CM). Em cada localidade, um PM representa a situação em que não há informação de precipitação; outro PM, a falta de informação de níveis. Assim, pode-se investigar a contribuição desses dados para as previsões.

Na Tabela 16, resumem-se os tipos e números de modelos desenvolvidos. As referências de ordenamento das variáveis neste item correspondem aos ordenamentos apresentados nas Figura 29, 30, 31 e 32. As variáveis em cada PM estão representadas na seção “8. APÊNDICE”.

Tabela 16. Tabela-resumo quanto aos modelos desenvolvidos nessa pesquisa. CM é modelo completo. PM é modelo parcial. PM sem 1 variável é quando se retira uma variável, tal que PM5 é aquele modelo em que a 5ª variável foi retirada.

Modelo	CM	PM sem 1 variável	PM sem precipitação	PM sem níveis	Outros PMs	Total de Modelos
rna_ENC	1	PM1 até PM10	PM11	PM12	PM13 até PM30	31
rna_EST	1	PM1 até PM13	PM14	PM15	PM16 até PM49	50
rna_PM	1	PM1 até PM15	PM16	PM17	PM18 até PM45	46
rna_TA	1	PM1 até PM18	PM19	PM20	PM21 até PM59	60

Não se notou nenhuma grande diferença no desempenho dos PMs sem 1 variável durante o período de verificação (ano 2017). O PM que apresentou diferença mais pronunciada foi aquele sem a precipitação de Passo Vacarias para previsões em Encantado. Há uma pequena diferença na MEA a partir do alcance de 8 horas no NS a partir de 14 horas. O desempenho desse modelo parcial durante a verificação é apresentado na Figura 33. Por outro lado, o efeito da remoção por completo de precipitação (Figura 34) ou de níveis foi bem aparente a partir do alcance de 14 horas.

A baixa diferença no desempenho dos PMs sem 1 variável e os CMs não quer dizer necessariamente que os dados isoladamente não sejam de nenhuma maneira úteis para o CM. Ao analisar a totalidade de 2017, pode estar mascarando-se o desempenho durante eventos extremos. Por isso, no item 4.6, é apresentada uma análise por eventos.

O PM11 de Encantado (Figura 34) indica a importância dos dados de precipitação para previsões durante o ano de 2017, principalmente para os alcances maiores. A partir do alcance de 12 horas, a falta de informações sobre a precipitação começa a impactar na MEA, e a partir do alcance 16 o efeito se torna também visível no NS. O E10 e o E90 não aparentam grande mudança, o que leva a inferir que a distribuição de erros, na falta dos dados de precipitação, tendeu a achatar, propiciando previsões menos precisas (com erro mais uniforme).

Figura 33. Efeito da retirada dos dados de precipitação do posto Vacaria (pPV), representado pelo PM2 (modelo parcial 2) em relação ao modelo completo (CM) para Encantado nas métricas NS, MEA, E10 e E90.

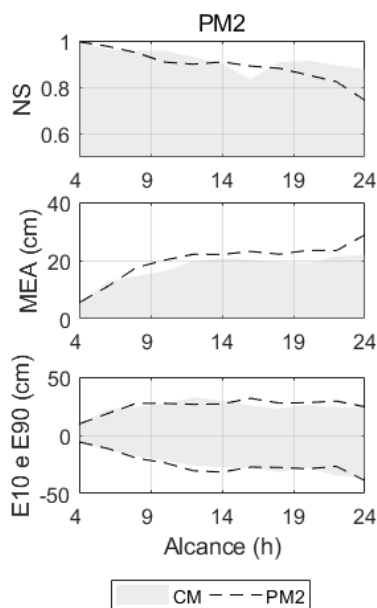
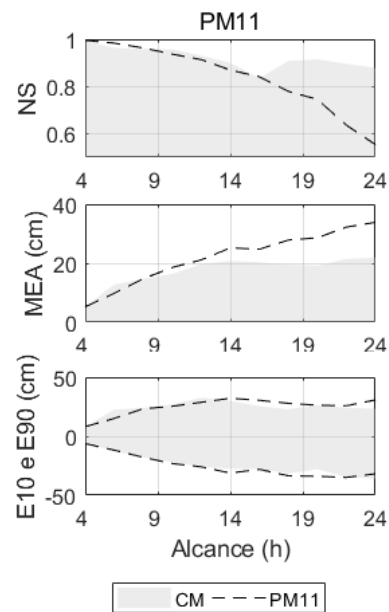


Figura 34. Efeito da retirada dos dados de precipitação, representado pelo PM11 (modelo parcial 11) em relação ao modelo completo (CM) desenvolvido para Encantado nas métricas NS, MEA, E10 e E90.



As entradas de precipitações não foram importantes para alcances pequenos (Figura 34) porque o efeito da persistência dos níveis, resultante do armazenamento, ainda se faz suficiente. Para alcances maiores, as entradas de precipitação fazem maior falta, podendo ser que precipitações futuras ao tempo atual fossem importantes. O efeito da remoção de dados de níveis foi tão grande que o NS ficou próximo a 0 e a MEA acima de 50 cm para todos alcances, indicando novamente a importância dos efeitos de persistência dos níveis.

As métricas de desempenho têm certa aleatoriedade. O NS para o alcance de 16 horas no CM é inesperadamente menor do que para os alcances próximos (14 e 18 horas). O NS do PM2 (Figura

33) parece ter aumentado em comparação ao CM para o alcance de 16 horas. O excesso de informação não deveria apresentar pior desempenho, quando o procedimento de treinamento fosse satisfatório. A pesquisa da importância de cada variável poderia ser mais clara se houvesse uma distribuição referente às repetições, apresentando o que geralmente acontece quando se retira cada entrada. Havendo informações quanto aos quantis dessas distribuições, poderia ser possível analisar estatisticamente a hipótese de qual variável retirada realmente faz falta.

Outras figuras apresentando o impacto da remoção de uma variável apenas para os modelos de Estrela, Porto Mariante e Taquari foram omitidas desse texto já que não apresentam potencial de retirar conclusões mais profundas dos modelos parciais (PM). Permite-se, porém, apresentar e discutir o efeito da remoção de dados de precipitação nessas localidades de Estrela (PM 14, Figura 35), Porto Mariante (PM16, Figura 36) e Taquari (PM19, Figura 37).

Figura 35. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Estrela nas métricas NS, MEA, E10 e E90.

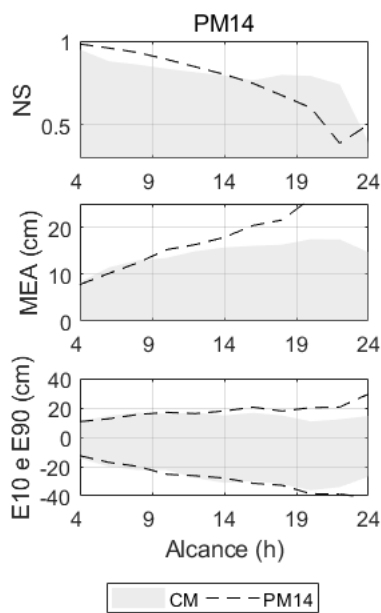


Figura 36. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Porto Mariante nas métricas NS, MEA, E10 e E90.

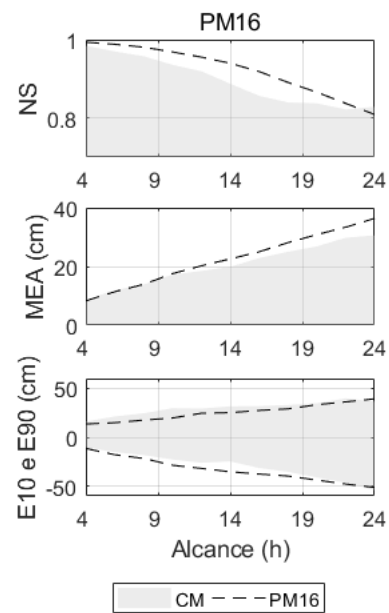
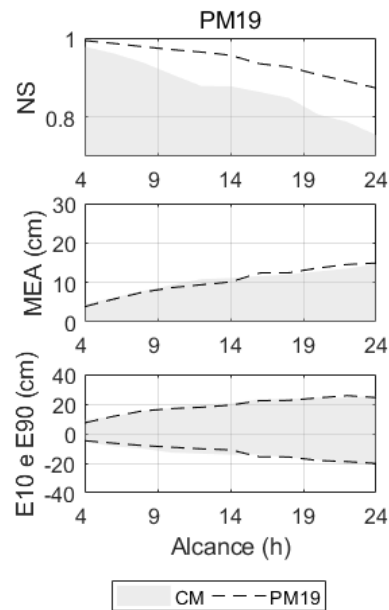


Figura 37. Efeito da retirada dos dados de precipitação do modelo desenvolvido para Taquari nas métricas NS, MEA, E10 e E90.



Para a localidade de Estrela, a falta de dados de precipitação fica mais pronunciada para um alcance maior de 10 horas pela métrica MEA e a partir de 16 horas pela métrica do NS. Em Porto Mariante e Taquari, a ausência de dados de precipitação aumentou a capacidade do modelo em descrever a variabilidade dos dados, visto que o NS aumentou.

Entre Estrela e Porto Mariante há a eclusa em Bom Retiro do Sul que trata de manter o nível do rio de tal maneira que garanta a navegação comercial até Estrela. Um dos impactos dessa eclusa deve ser a maior memória dos dados de nível em Porto Mariante e Taquari, já que o nível é regularizado. Por sua vez, pode ser a razão da métrica MEA não ter piorado tanto para essas localidades ao retirar a informação de precipitação, bem como a anomalia observada na métrica NS.

Os dados de nível são extremamente importantes para os modelos desenvolvidos nesta pesquisa. Os dados de precipitação são de maior importância para os modelos quando previsão em alcances maiores é pretendida, pois são nestes momentos que a memória do dado de cota já não é mais suficiente para modelar o futuro.

4.6. Análise por eventos com as Árvores de RNAs

Essa etapa consiste em uma análise mais prática dos modelos, emitindo previsões em tempo atual para seis eventos com picos pronunciados em todos os alcances estudados. Os dois primeiros eventos (E1 e E2) fizeram parte do conjunto de treinamento e validação dos modelos; já os eventos restantes (E3, E4, E5 e E6) não, eles estão contidos somente no período de 2017.

Os dois primeiros eventos indicam sobre um cenário irreal de importância teórica: qual o desempenho da rede em descrever em tempo atual um evento extremo que foi utilizado durante o treinamento? A rede deve estar apta a descrever esses eventos muito bem.

Os eventos restantes também foram selecionados para indicar o que aconteceria em um caso real em que se aplicasse os modelos. Qual o desempenho da rede em descrever em tempo atual um evento extremo que jamais foi visto durante o treinamento (em comparação com aqueles que foram vistos)?

Nesta pesquisa são apresentadas apenas três figuras por evento embora haja cerca de 500 figuras para os seis eventos de cada uma das quatro localidades (Encantado, Estrela, Porto Mariante e Taquari). As figuras podem ser visualizadas na íntegra no seguinte sítio: bit.ly/3bVCde5.

Durante os eventos 1 e 3 em Porto Mariante, houve falha momentânea no sistema de aquisição de dados. Durante os eventos 4 e 5, houve falha contínua no sistema de aquisição de dados. Assim, não há hidrogramas observados para algumas localidades nesses eventos.

Os eventos 1 (Figura 38-Figura 41) e 2 (Figura 42-Figura 45) puderam num geral ser bem estimados pelos modelos mesmo que a eventual falta de informações implicou no uso de modelos parciais (PM). Em Porto Mariante, durante o evento 1 (Figura 40, primeiro gráfico), as redes com as quais se realizam previsões a partir de 10 horas descrevem a ascensão do hidrograma de maneira caótica.

Figura 38. Previsão em tempo atual feita em ENC para o Evento 1 (participou do treinamento).

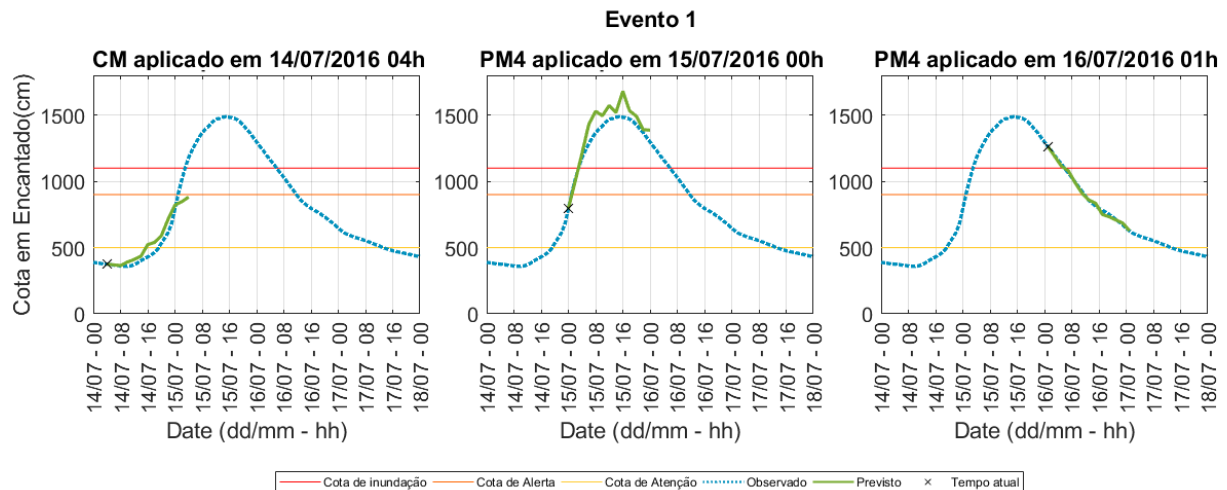


Figura 39. Previsão em tempo atual feita em EST para o Evento 1 (participou do treinamento).

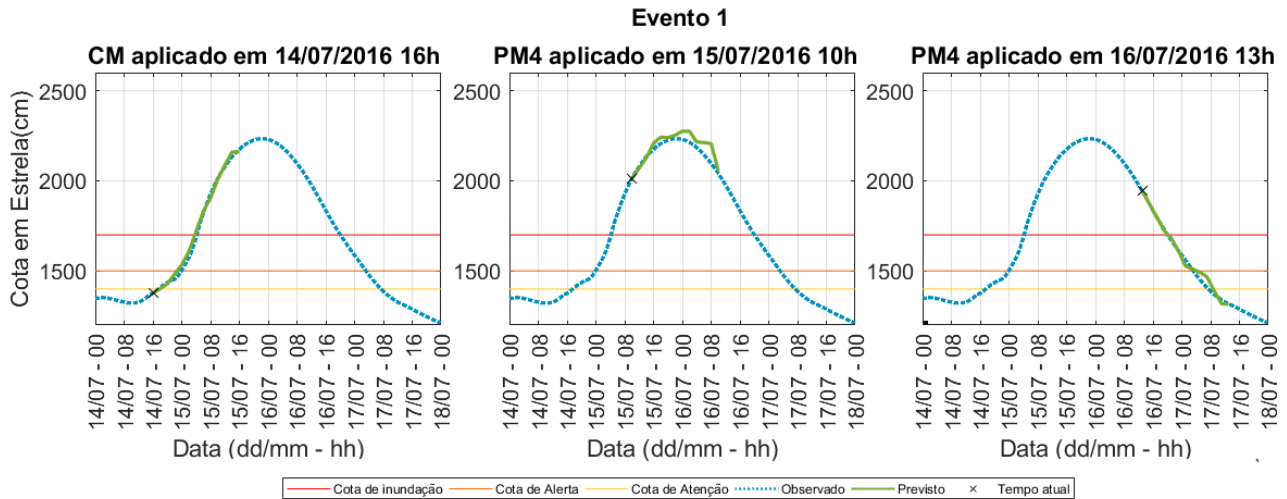


Figura 40. Previsão em tempo atual feita em PM para o Evento 1 (participou do treinamento).

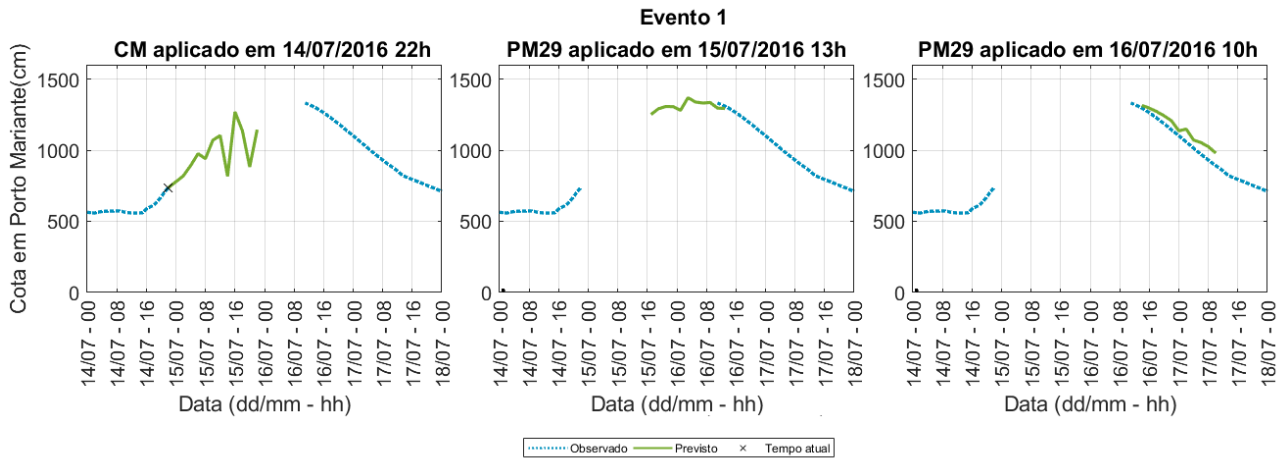
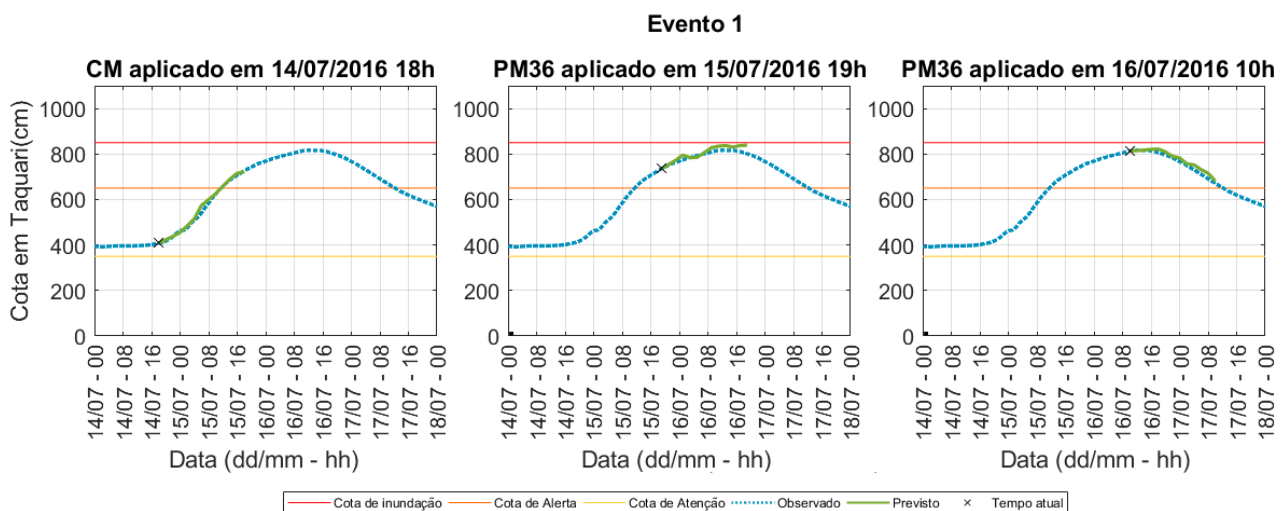


Figura 41. Previsão em tempo atual feita em TA para o Evento 1 (participou do treinamento).



Visualmente, os hidrogramas previstos e observados no evento 2 são semelhantes. Há raras discrepâncias mais pronunciadas no hidrograma previsto em Estrela para alcance maiores que 10

horas (Figura 43) quando o PM38 foi acionado. Poderia ser que essa discrepância estivesse relacionada com a falta de informação em Vacaria e Ibiraiaras (Apêndice A, Tabela 18). Porém, o mesmo evento em Taquari foi, visualmente, bem estimado utilizando o PM51 (Figura 45), que carece das mesmas variáveis.

A descrição dos eventos 1 e 2 não é exata por causa do critério da validação cruzada para impedir o sobreajustamento. Faz parte da metodologia que não haja uma descrição exata; o modelo deve estar apto à generalização.

Figura 42. Previsão em tempo atual feita em ENC para o Evento 2 (participou do treinamento).

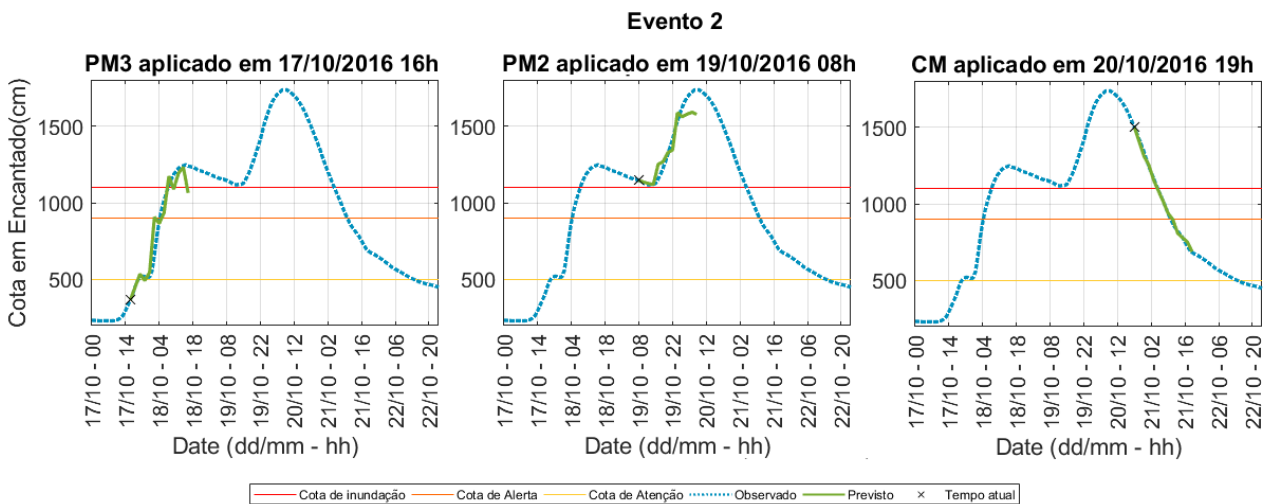


Figura 43. Previsão em tempo atual feita em EST para o Evento 2 (participou do treinamento).

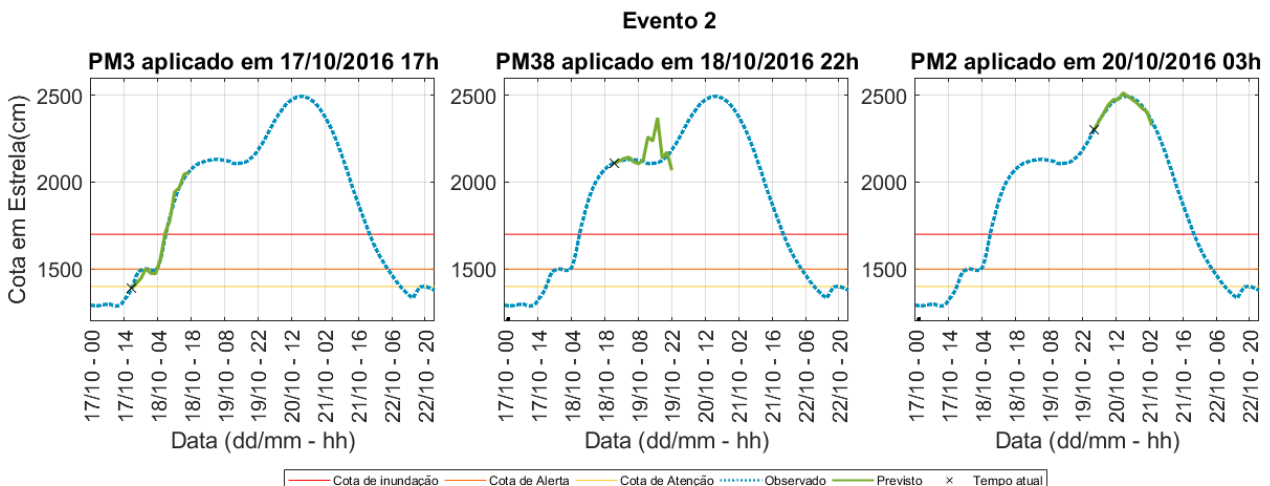


Figura 44. Previsão em tempo atual feita em PM para o Evento 2 (participou do treinamento).

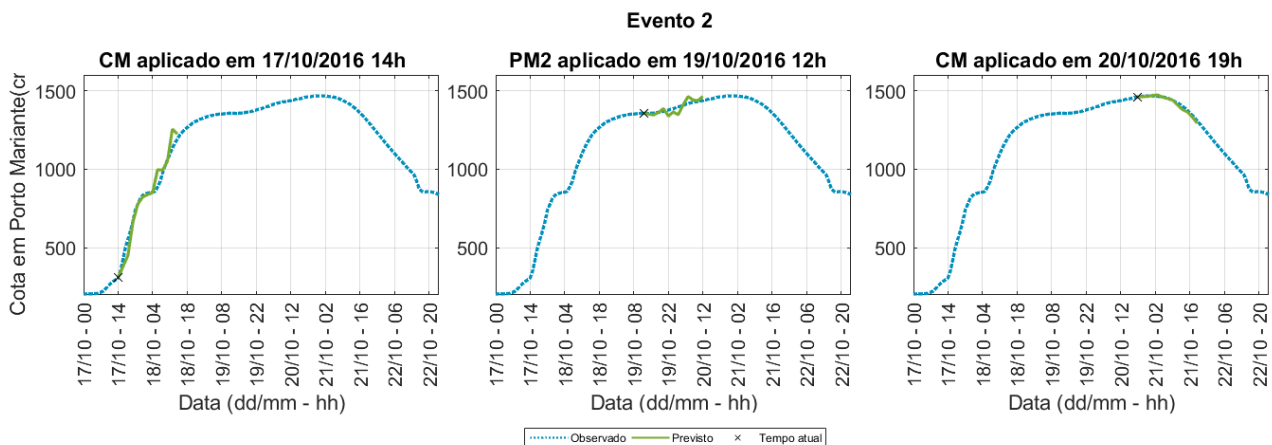
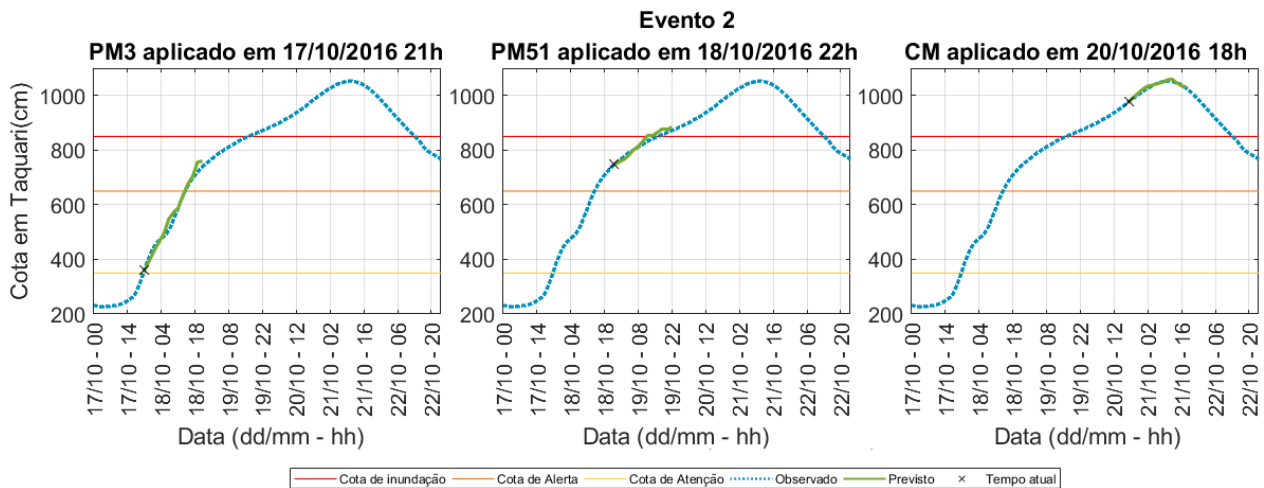


Figura 45. Previsão em tempo atual feita em TA para o Evento 2 (participou do treinamento).



Justo por isso é importante analisar o ocorrido nos eventos 3, 4, 5 e 6 para essas localidades. O pico do evento 3 em Encantado e Estrela foi consideravelmente superestimado em alguns momentos quando faltou a segunda variável dos modelos (precipitação em Vacaria, PV), especialmente quando faltava mais de 12 para o pico. Isso pode ser decorrência tanto da falta de uma boa descrição da precipitação na bacia nesse momento, ou devido a uma incerteza nos pesos sinápticos utilizados. Isso pode acontecer devido ao período de dados não ser muito extenso, pode ser que a maneira que a distribuição espacial de precipitação durante esse evento seja suficientemente diferente daqueles padrões de “estado inicial das variáveis hidrológicas VS variação na cota até o alcance previsto”, que estavam no conjunto de treinamento e validação.

Já para Porto Mariante e Taquari as cotas estimadas aproximam-se bastante da cota observada. Esse é o resultado mais encorajador para previsões nessas localidades.

Figura 46. Previsão em tempo atual feita em ENC para o Evento 3 (não presente do treinamento).

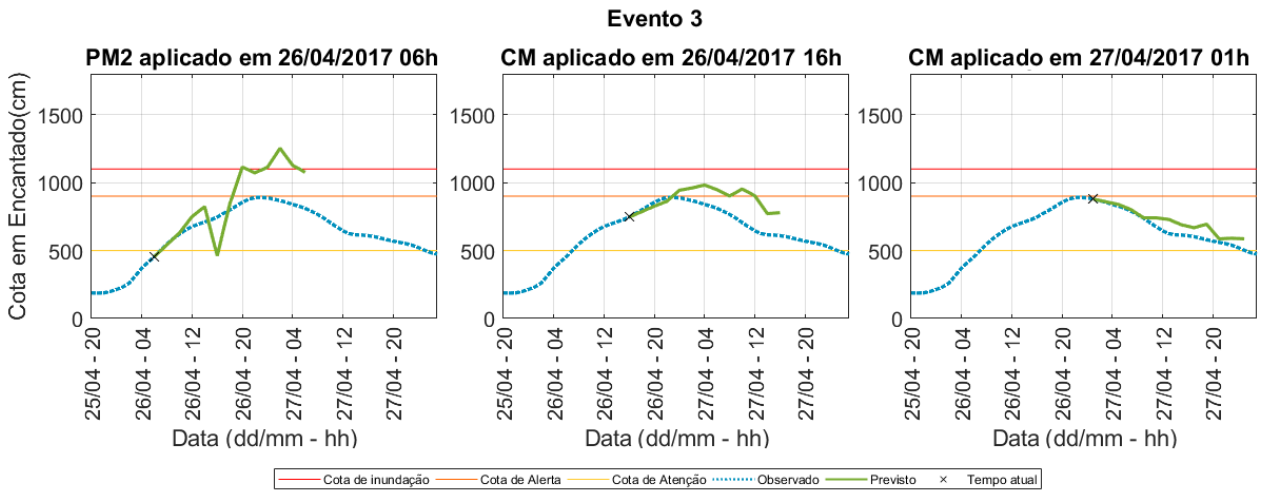


Figura 47. Previsão em tempo atual feita em EST para o Evento 3 (não presente do treinamento).

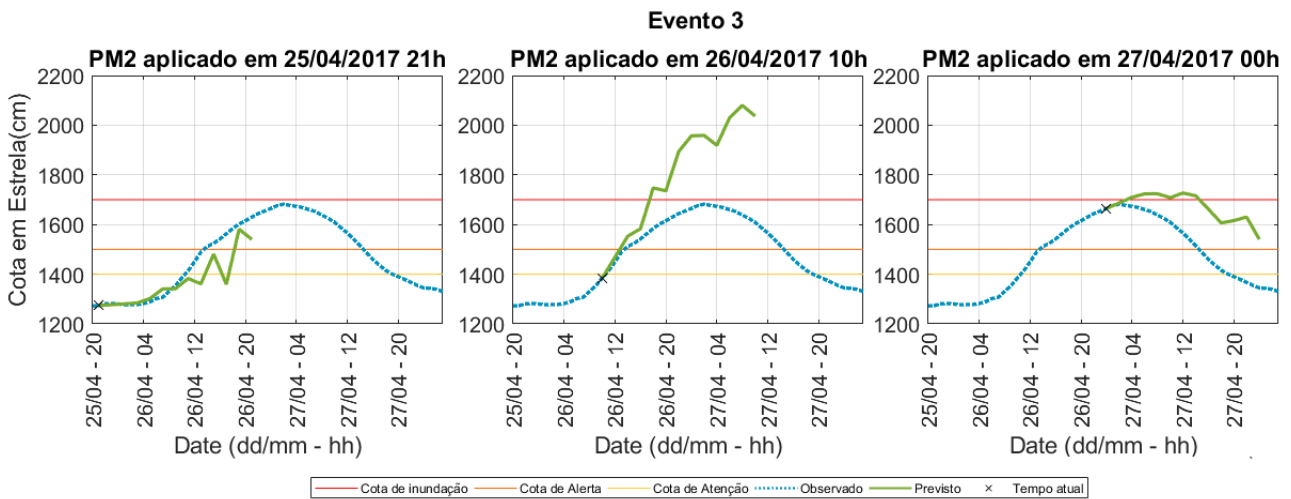


Figura 48. Previsão em tempo atual feita em PM para o Evento 3 (não presente do treinamento).

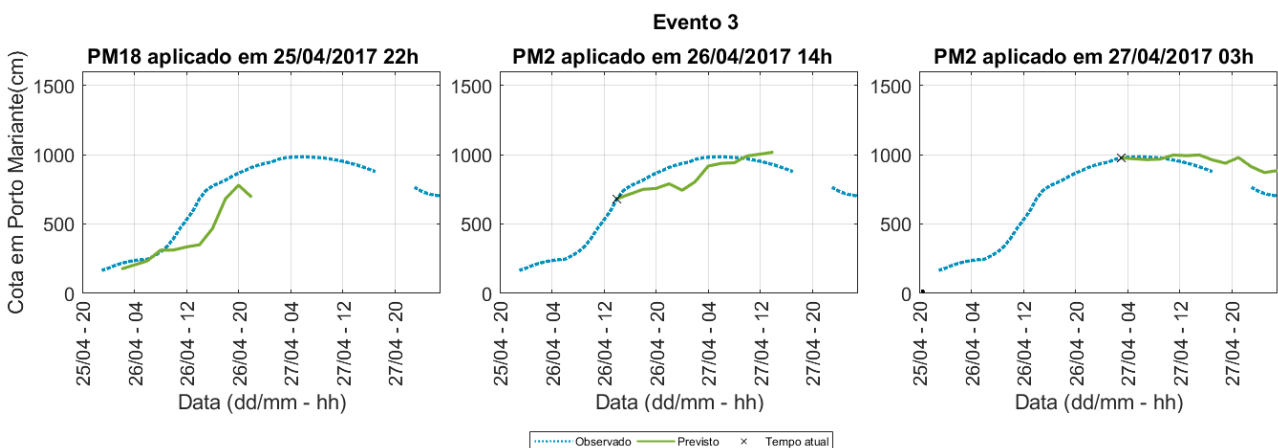
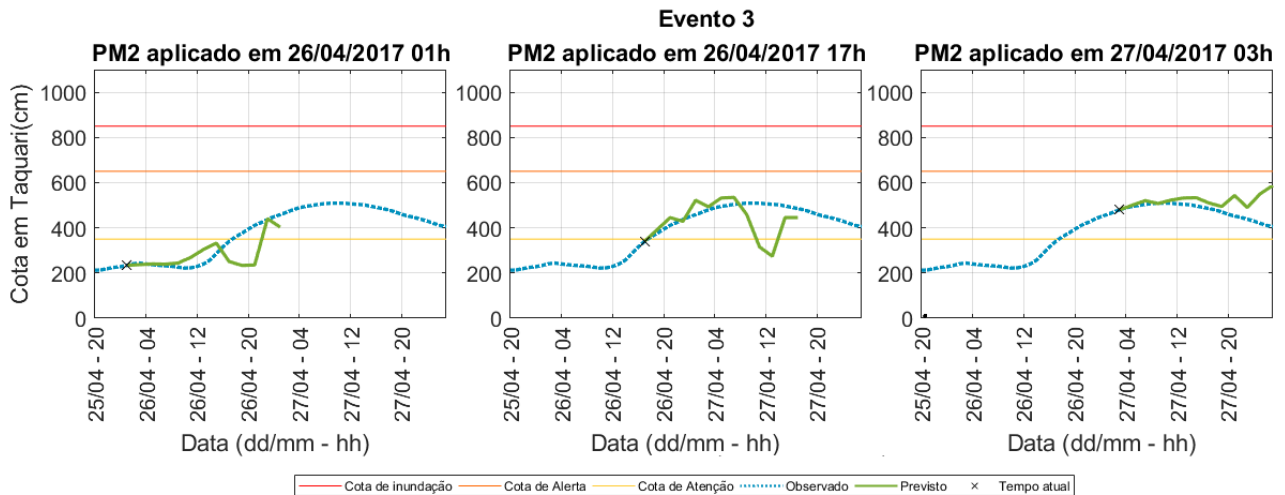


Figura 49. Previsão em tempo atual feita em TA para o Evento 3 (não presente do treinamento).



No evento 4, apesar da falta de dados sistemática (pPV, pMU, pENC, fPM), algumas previsões foram semelhantes aos observados (principalmente na recessão). Novamente, os hidrogramas observados são mais suaves que os previstos, que são caóticos.

Figura 50. Previsão em tempo atual feita em ENC para o Evento 4 (não presente do treinamento).

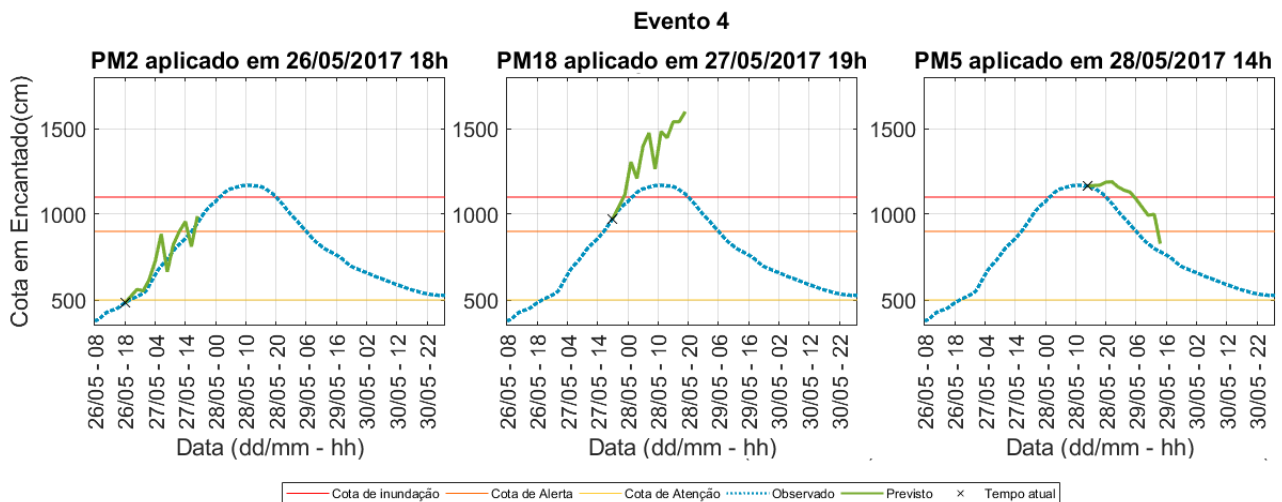


Figura 51. Previsão em tempo atual feita em EST para o Evento 4 (não presente do treinamento).

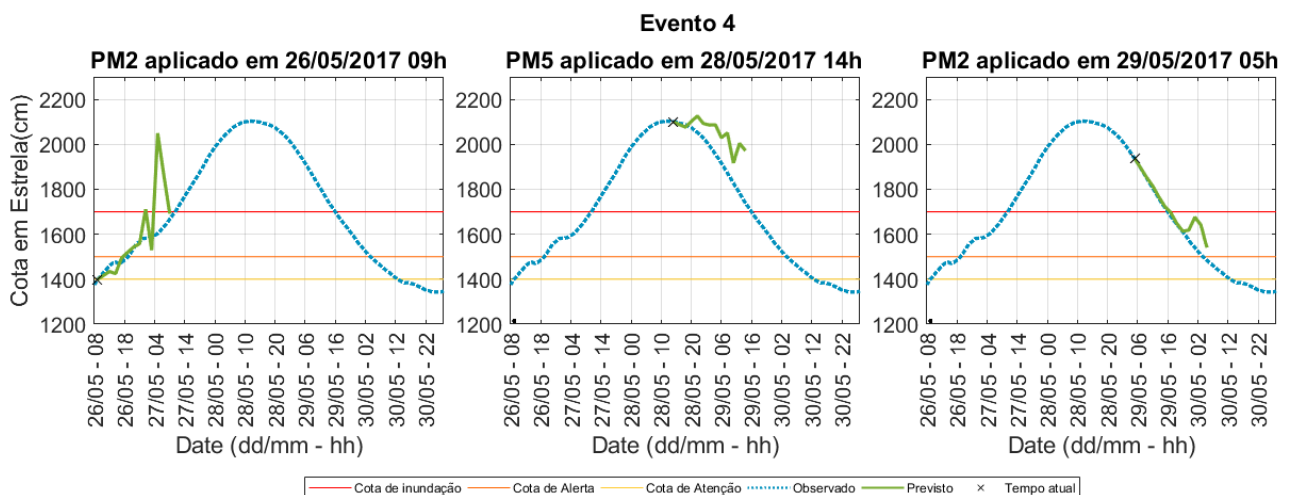


Figura 52. Previsão em tempo atual feita em PM para o Evento 4 (não presente do treinamento).

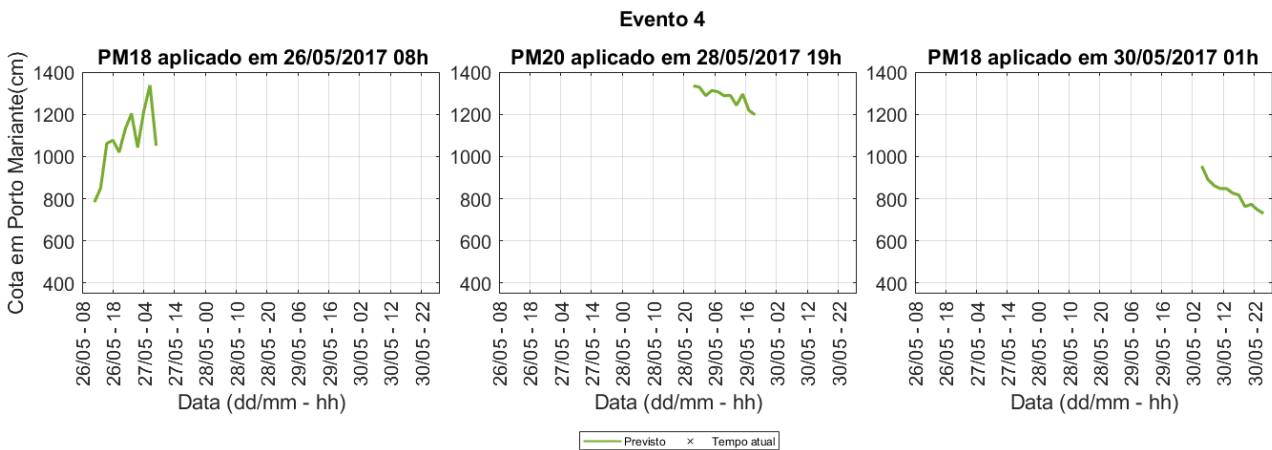
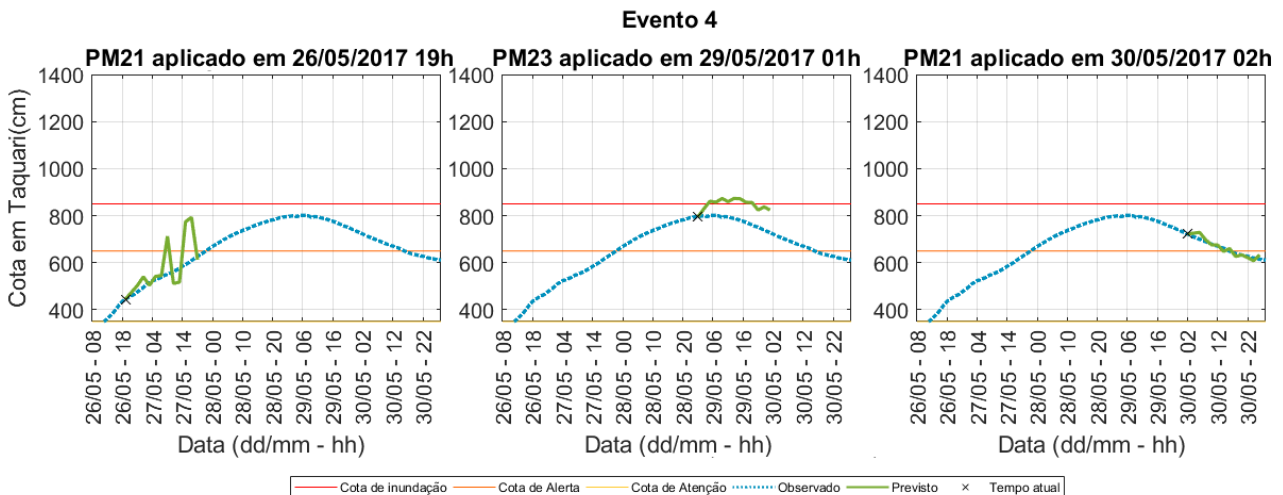


Figura 53. Previsão em tempo atual feita em TA para o Evento 4 (não presente do treinamento).



Durante o evento 5, houve falha sistemática (pPT, pSC, pMU, pENC, pTA e fPM). Faltou informação quanto à precipitação de cabeceira em Passo Tainhas, mais a montante, e Serafina Correa durante todo o evento, o que impossibilitou o cálculo do FMMG para essas entradas. Além disso, os dados de precipitação em Muçum, Encantado e Taquari faltaram por algum momento também. Não havia informações em tempo atual quanto à cota em Porto Mariante, tampouco. Apesar desse cenário de indisponibilidade de dados, mesmo assim algumas previsões puderam ser realizadas, o que justifica o tempo investido em criar modelos parciais.

Figura 54. Previsão em tempo atual feita em ENC para o Evento 5 (não presente do treinamento).

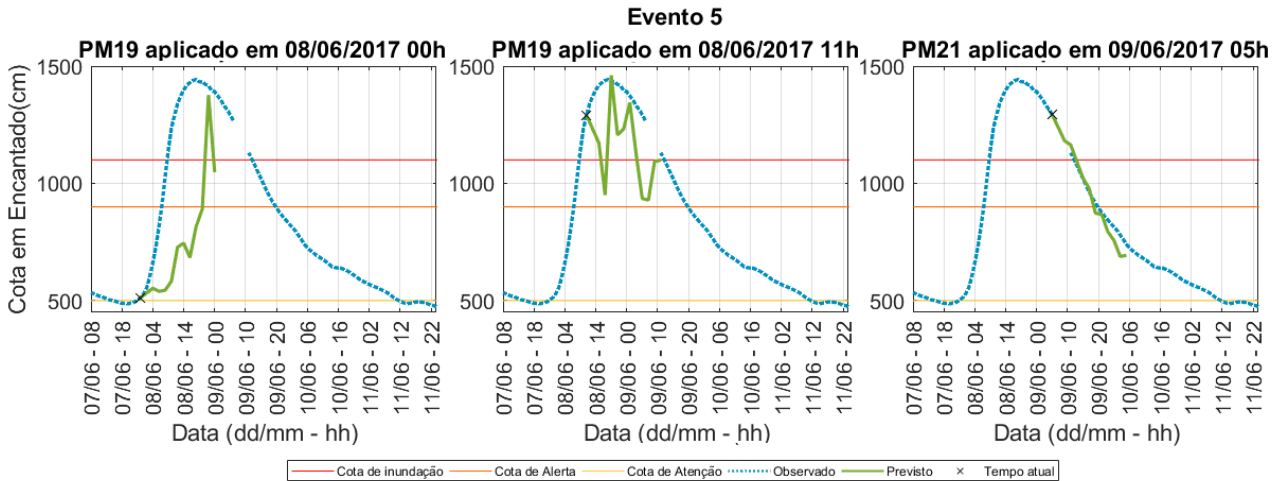


Figura 55. Previsão em tempo atual feita em EST para o Evento 5 (não presente do treinamento).

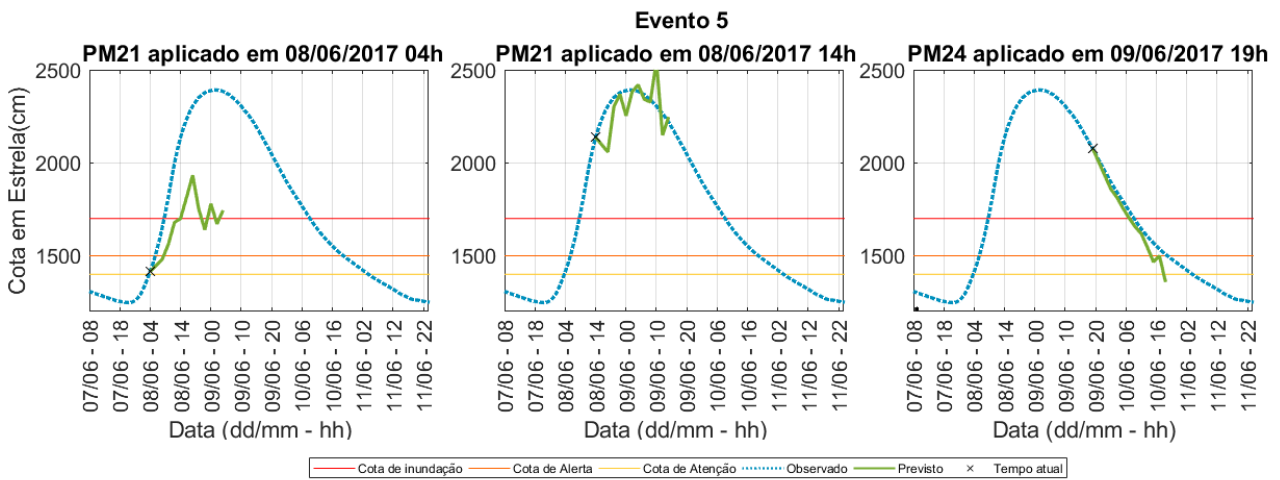


Figura 56. Previsão em tempo atual feita em PM para o Evento 5 (não presente do treinamento).

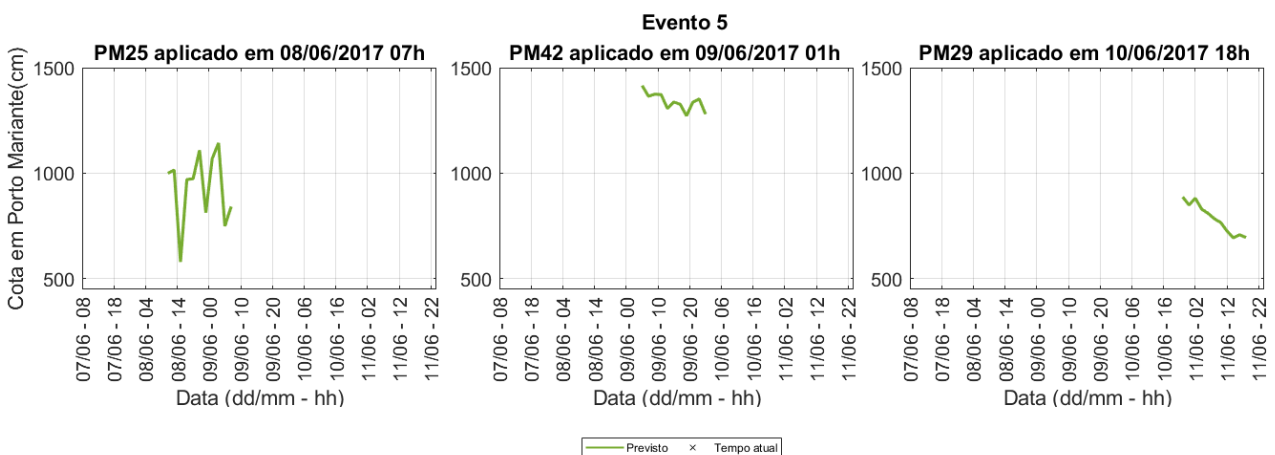
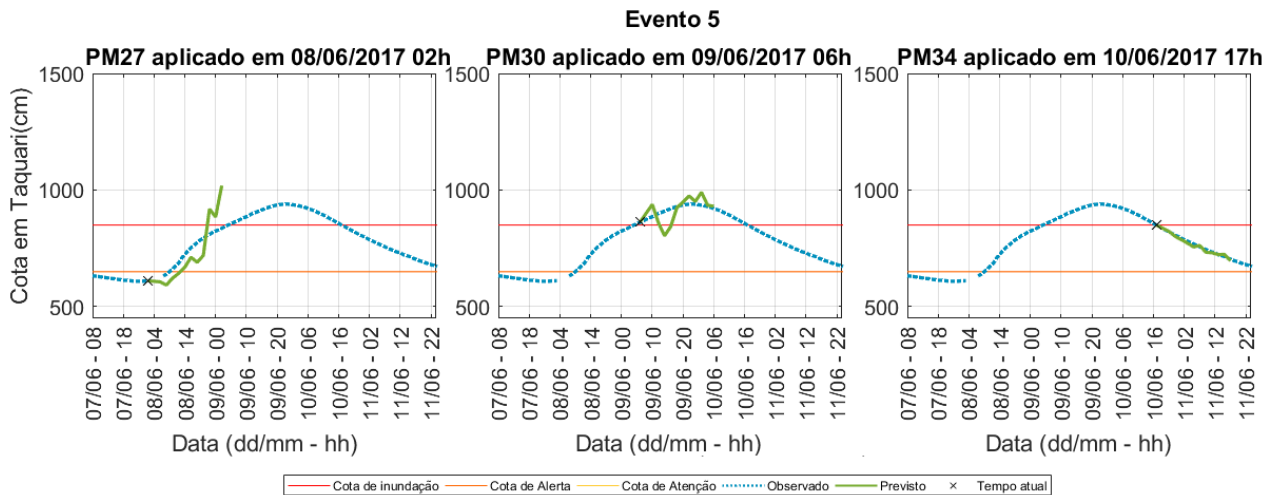


Figura 57. Previsão em tempo atual feita em TA para o Evento 5 (não presente do treinamento).



Para o evento 6 também aconteceu de modelos parciais de índice mais elevado terem sido necessários (os índices mais elevados são referentes a 2 ou mais dados de entradas estarem indisponíveis). Esse evento possibilitou um outro aprendizado quanto ao sistema de aquisição: no segundo gráfico da

Figura 58, referente a nove horas antes do pico do evento 6 em Encantado, previsões são emitidas utilizando o CM (modelo completo). Mesmo com todas as informações de entrada disponíveis, não foi possível descrever a ascensão do hidrograma muito bem. Esse se encontra descrito em um ritmo agitado de subidas e descidas. Isso pode ser um indício de que faltam dados na bacia para aprender as relações entre as variáveis hidrológicas. Outra hipótese é que daquelas 10 repetições realizadas durante o treinamento, não se escolheu aquela que melhor generalizasse o que foi aprendido. Dentre as 10 foi sempre selecionada aquela repetição que possuía o erro quadrático mínimo no conjunto de validação, o que, a princípio, garante, antes de tudo, que se escolha a repetição que melhor generaliza o aprendido para o conjunto de validação, mas não necessariamente a que melhor generalizará para os dados jamais vistos. Não é plausível montar outro conjunto de validação porque esse problema persistiria (estaríamos selecionando agora uma repetição para dois conjuntos de validação).

Acredita-se haver uma variabilidade natural, uma incerteza no método de escolher a repetição. Se a maioria das repetições são elegíveis, no sentido de que passaram por um treinamento com o algoritmo de ajuste dos pesos, por que não utilizar todas elas? Assim haveria uma descrição da incerteza nos pesos sinápticos escolhidos. Qual seria a tendência central das previsões geradas se fossem utilizadas todas as repetições, todos os conjuntos de pesos?

Figura 58. Previsão em tempo atual feita em ENC para o Evento 6 (não presente do treinamento).

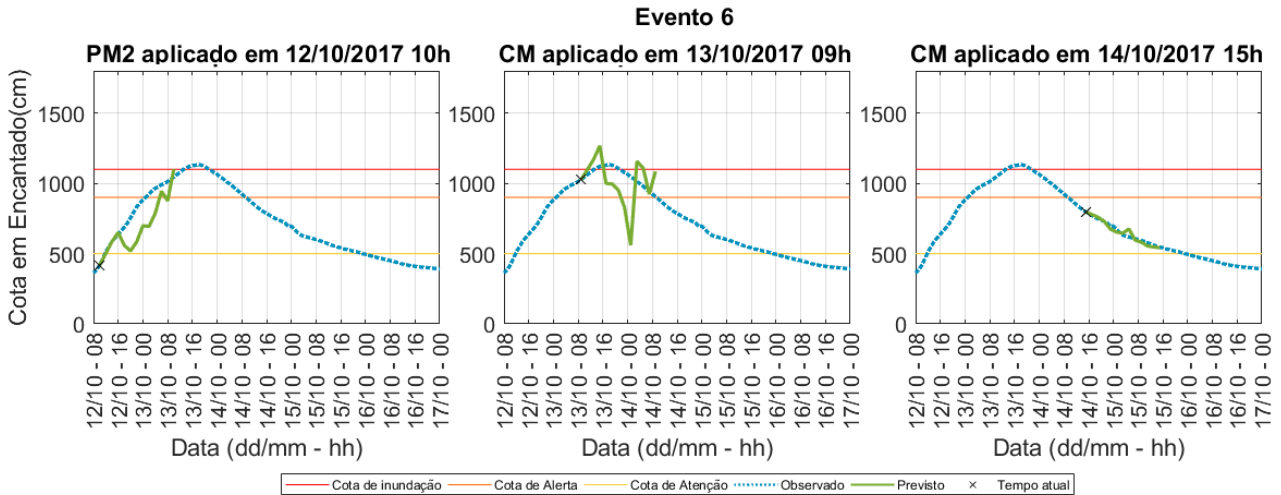


Figura 59. Previsão em tempo atual feita em EST para o Evento 6 (não presente do treinamento).

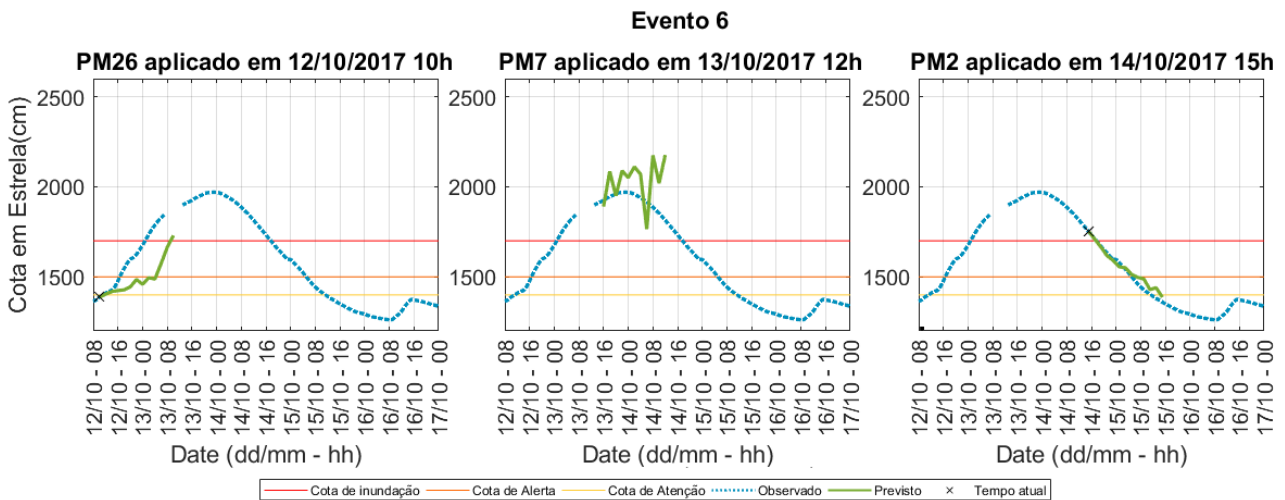


Figura 60. Previsão em tempo atual feita em PM para o Evento 6 (não presente do treinamento).

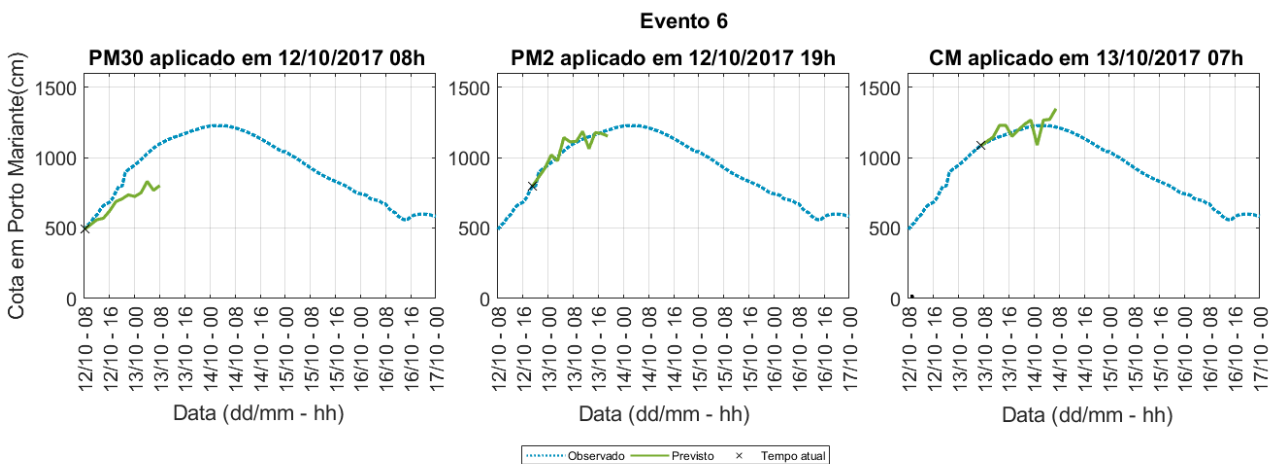
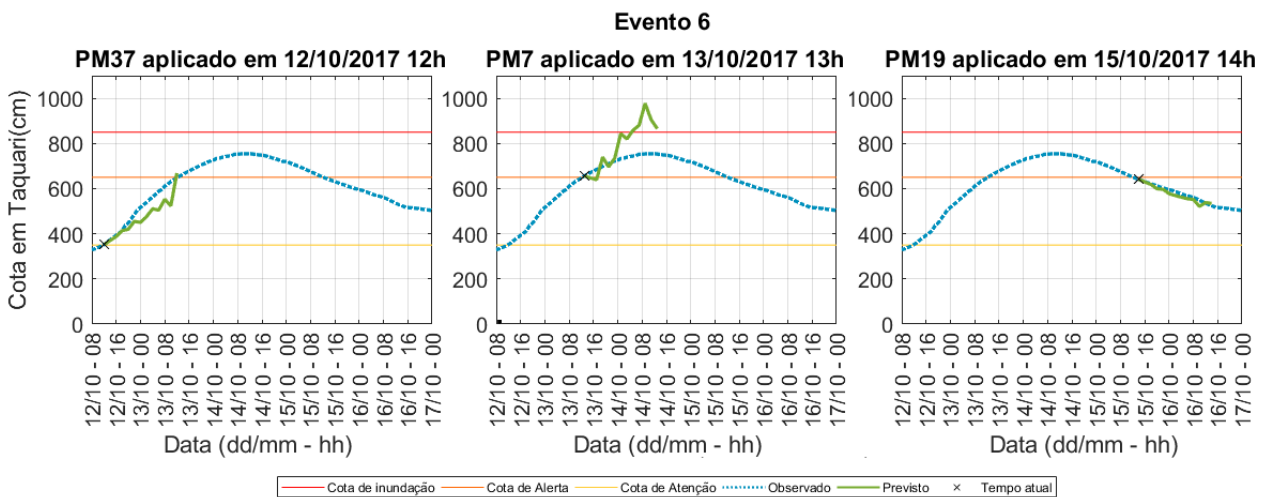


Figura 61. Previsão em tempo atual feita em TA para o Evento 6 (não presente do treinamento).



Quanto às diferenças entre os eventos do período de treinamento e validação, e do período de verificação, é notável pior desempenho nos eventos do período de verificação com relação à previsibilidade. Pode ser que os cenários da verificação não tenham sido bem representados nas séries utilizadas para treinamento e validação. Um ponto crítico da previsão, notado nas figuras apresentadas, é a ascensão do hidrograma. Pode ser que faltem exemplos de eventos semelhantes na série disponível para o treinamento; a distribuição espacial e temporal do evento chuvoso no período de validação pode ser sido consideravelmente diferente. Por outro lado, a recessão do hidrograma costuma ser melhor prevista, o que pode ser útil para estimar a quantidade de recursos necessário para o tempo que famílias precisarão ser abrigadas.

Quanto aos alcances inspecionados nesse trabalho, para horizontes menores de até 10 horas, dados de cotas foram suficientes para previsão. Para horizontes maiores uma boa descrição da precipitação recente na bacia é importante. Aquelas previsões com horizontes que já se aproximam de 20 horas talvez necessitem de informação de chuvas futuras, advindas de modelos meteorológicos. Por fim, os hidrogramas previstos são bem menos suaves que os observados, isso é reflexo de que cada previsão é feita independentemente por uma RNA com pesos sinápticos diferentes.

4.7. Discussão dos resultados com referência à literatura

Os resultados são importantes por trazer contribuições científicas e fortalecer certas abordagens utilizadas na metodologia. As pesquisas recentes com RNAs contam com artifícios para agregar informações temporais e espaciais, reduzindo o número de entradas necessárias para a RNA. Uma das técnicas bastante utilizada é a das ondaletas (SHOAIB *et al.*, 2017; NOURANI, 2014). Assim como em outras pesquisas (BRAVO *et al.*, 2008; OLIVEIRA *et al.*, 2015c; GERHARD, 2017), a alternativa dos filtros FMMU e FMME foram investigados. O novo FMMG, também investigado,

foi escolhido para os dados de precipitação por necessitar de menos registros passados para seu cálculo em alcances pequenos, o que é importante para previsões em tempo atual que dependem de uma rede de monitoramento eventualmente sujeita falha de sinal durante eventos extremos. A abordagem dos filtros permite identificar os dados agregados e a ponderação realizada, que, no caso do FMMG, indica também a precipitação mais importante para explicar a variação de nível.

A análise do número de registros de treinamento indicou que mais registros não implica sempre em mais informação, podendo dificultar o procedimento de treinamento, inclusive. Diferente da análise feita em Nourani (2017) a validação foi mantida fixa, garantindo uma comparação mais justa das estratégias de divisão de registros entre os conjuntos.

Adicionar neurônios internos não aumenta a capacidade de representar a não-linearidade dos dados após um certo limiar. A análise da complexidade interna foi uma maneira de identificar o número mínimo de neurônios internos suficientes para descrever a não-linearidade dos dados de maneira semelhante a uma rede com número de neurônios internos superdimensionado (vinte). A vantagem dessa técnica está em seu fundamento conceitual lógico, sendo uma alternativa a simples adoção da arquitetura expressa no teorema de Hecht-Nielsen (1987) de duas vezes o número de neurônios de entrada mais um ($2n+1$), ou ao teste exaustivo de números de neurônios internos. Foi possível identificar indícios do platô de desempenho atingido com o número mínimo de neurônios internos, embora houvesse uma banda aleatória.

Tratando-se da literatura científica na mesma bacia, há somente um estudo com RNAs. Pedrollo (2017) apresentou dois modelos de previsão baseados em RNAs para Estrela. Um deles, com alcance de 5 horas, tem seguinte desempenho: NS de 0,97, EA95 de 32 cm e MEA de 9 cm; outro, com alcance de 11 horas: NS de 0,93, EA95 de 53 cm e MEA de 14 cm. Os NSs encontrados em Pedrollo (2017) são um pouco maiores. As MEAs são próximas às obtidas na presente pesquisa para a localidade de Estrela com os alcances de 4 e 12 horas (Tabela 11).

A diferença entre os NS obtidos em Pedrollo (2017) e nesta presente pesquisa advém dos dados (diferentes resoluções temporais, períodos e estações) e da metodologia aplicada. Em Pedrollo (2017), por exemplo, foram utilizados dez neurônios internos para Estrela, em contraste aos quatro dessa pesquisa de mestrado. Portanto, comparações devem ser realizadas com cautela, considerando o período de verificação, que, infelizmente, não está explícito em Pedrollo (2017). De qualquer maneira, a abordagem utilizada por Pedrollo (2017), chamada na presente pesquisa de mestrado de *Árvore de RNAs*, foi útil como alternativa a gerar previsões em tempo atual multialcance para cenários de falta de sinal, acionando os respectivos PMs. Porém, o desempenho dos PMs é geralmente pior que do CM (quando a rede de monitoramento está funcionando em sua totalidade).

O principal limitador dos modelos desenvolvidos está relacionado à série de dados utilizada. Essa é de 26/11/2015 às 19 h até 24/04/2019 às 6 h. Como apresentado na Figura 11, enquanto a

maior inundação registrada em Lajeado foi de cerca de 29,92 metros, a maior cota de inundação atingida durante o período utilizado no desenvolvimento dos modelos foi de 24,93 metros. Tratando-se de modelos empíricos, não é de se encorajar que esses sejam empregues para além do domínio; extrapolações devem ser evitadas. Essa uma limitação só poderá ser sanada quando num futuro ocorrerem inundações maiores que tenham sido devidamente monitoradas pela rede. Os modelos regressão de cotas-máximas de inundação presentes em Gerhard (2017) podem ser úteis para eventos mais extremos do que aqueles presentes na série de dados telemétricos horários do SACE-Taquari.

Quantos parâmetros há nos modelos? Quais parâmetros são esses? As arquiteturas dos CM foram 10-4-1 (Encantado), 13-4-1 (Estrela), 15-4-1 (Porto Mariante) e 18-5-1 (Taquari). Considerando o termo independente da camada interna e aquele da camada de saída, os modelos de previsão tiveram 49, 61, 69 e 101 pesos sinápticos, respectivamente. Seis séries pluviométricas são filtradas com FMMG — que tem dois parâmetros, E e F — no modelo de Encantado, sete em Estrela e Porto Mariante, e oito em Taquari, totalizando, junto com os parâmetros das RNAs, para cada localidade: 61 parâmetros (Encantado), 75 (Estrela), 83 (Porto Mariante) e 117 (Taquari). Esses parâmetros tem valores específicos, na metodologia utilizada, para cada um dos onze alcances. Os modelos parciais, por serem formas reduzidas, possuem menor quantidade de parâmetros.

5. CONCLUSÕES

A previsão hidrológica pode ser útil para evitar perdas econômicas e sociais. Esta pesquisa objetivou determinar a aplicabilidade de RNAs para previsão em tempo atual de níveis fluviais com vistas ao alerta de cheias para os postos do sistema de alerta da bacia do rio Taquari-Antas/RS (SACE-Taquari), utilizando exclusivamente os dados horários da rede telemétrica do SACE-Taquari do período de 26/11/2015 a 24/04/2019. Foram montados modelos de previsão com alcance de 4 até 24 horas para as localidades de Encantado, Estrela, Porto Mariante e Taquari.

A verificação dos modelos completos de RNAs no período de 2017 teve resultados encorajadores, mesmo utilizando exclusivamente os dados do SACE-Taquari. Para um horizonte de 12 horas em Encantado, Estrela, Porto Mariante e Taquari, as métricas MEA e NS foram, respectivamente, (19,7 cm; 14,9 cm; 18,5 cm; e 10,8 cm) e (0,930; 0,813; 0,919; e 0,878). Para horizontes menores de 4, 6 e 8 horas, os desempenhos foram ainda melhores.

Notou-se algumas discrepâncias entre os níveis fluviais observados e previstos na análise por eventos, a qual foi necessária ver a previsão com vistas ao de alerta de inundação (4.6). Alcances acima de 14 horas serviram para indicar a tendência de longo prazo, e os alcances abaixo de 12 horas para realizar previsões pontuais, por exemplo.

A aplicabilidade de modelos de previsão em tempo atual baseados em Árvores de RNAs com vistas ao sistema de alerta foi evidenciada pelos resultados de desempenho no período de verificação (4.4) e na análise por eventos entre hidrogramas observados e previstos (4.6). Pôde-se prever continuamente com as Árvores de RNAs, mesmo com uma rede de monitoramento mínima sujeita a eventuais cenários de indisponibilidade. Acessando os dados pelo portal do SACE-Taquari, os operadores da rede podem prever em tempo atual em alcances de 4 até 24 horas à frente, dado que os cenários de indisponibilidade de dados tenham sido contemplados durante o desenvolvimento dos PMs das Árvores de RNAs.

É necessário o uso de dados horários para prever eventos de inundação no vale do Taquari porque a ascensão da cota de inundação até o pico do evento de alguns eventos foi de cerca de 12 horas. Isso indica que dados horários são importantes para descrição dos fenômenos.

Os filtros usados para agregar informações, contribuindo para modelos mais parcimoniosos. Para níveis, o FDM foi o que apresentou maiores correlações com as diferenças entre o tempo atual e os alcances testados. Para precipitação, o FMMG foi selecionado por necessitar menos dados antecedentes, não utilizar o parâmetro de defasagem, e ainda assim apresentar desempenho semelhante ao FMMU, segundo melhor filtro. Com os filtros pôde-se identificar o exato número de registros utilizados e, no caso do FMMG, o dado mais relevante à agregação.

Por serem empíricos, uma limitação dos modelos é que o domínio de treinamento indica os limites para sua aplicação (Tabela 4). Já que os dados são do período de 11/2015 a 04/2019, há certos eventos extremos de maior magnitude, como o de 1941, para os quais a RNA não foi treinada.

A razão clássica proposta em Hecht-Nielsen (1990) de 2:1:1 (conjuntos de treinamento, validação e verificação de uma RNA) não foi utilizada nessa pesquisa. Utilizou-se cerca de 1.200 registros de treinamento, 7.200 de validação e 3.600 de verificação. Houve indícios de que quanto mais informação (na forma de registros que representem maior diversidade de cenários) no treinamento de uma RNA, melhor é seu desempenho, como visto nas métricas para o conjunto de validação partindo de 200 registros de treinamento. Registros em excesso, porém, dificultam o treinamento — às vezes nem 300 repetições foram suficientes para o treinamento com 2000 registros de treinamento (4.2). Fixar o conjunto de validação é importante para prevenir análises injustas na pesquisa pelo número de registros de treinamento.

O salto de cinco para dez neurônios internos ao pesquisar a complexidade interna (4.3) dificultou identificar se o tal platô para Taquari. De qualquer maneira, a pesquisa serviu para demonstrar a aplicabilidade do método visual para complexidade interna.

Essa pesquisa de mestrado não esgotou a necessidade de outras pesquisas na previsão hidrológica realizada com Redes Neurais Artificiais com vistas ao SACE-Taquari. Principais pontos que deveriam ser considerados, futuramente, tanto em pesquisas (por potencial de inovação) como na operação do SACE-Taquari são apresentados como recomendações.

6. RECOMENDAÇÕES

Sugerem-se modelos complementares para auxiliar no processo de decisão, como os modelos conceituais de Siqueira *et al.* (2016) para alcances maiores que 24 horas e, em caso de eventos extremos superiores àqueles da presente pesquisa, as regressões de cotas máximas de Gerhard (2017).

A análise adicional de eventos dessa pesquisa pode ser refeita quantitativamente, com métricas comparativas. As previsões multialcance podem apresentar comportamento frenético quando cada horizonte é previsto por uma RNA, e talvez possam suavizar em rede com saídas multialcance. Pode ser difícil, porém, ajustar filtros para as entradas, cujos parâmetros dependeram dos alcances nessa presente pesquisa. Sugere-se identificar tipos de registros extremos, visionando uma amostragem tipológica para montar os conjuntos de treinamento, validação e verificação (*e.g.* eventos com precipitação de verão/inverno vinda do Norte/Sul). Amostragem por curvas de permanência ou por conjuntos difusos, considerando a imprecisão na hora dessa seleção, podem ser caminhos.

Um critério de treinamento específico para eventos extremos talvez possa dirigir a melhores desempenhos, mas o impasse seria obter suficientes registros de “eventos extremos”. Futuros modelos de previsão com RNAs para sistemas de alerta poderiam utilizar um critério de seleção mais rigoroso para repetição no treinamento, ou usar todas as repetições satisfatoriamente treinadas para previsões, analisando simultaneamente a distribuição de previsões. As incertezas do procedimento de treinamento são propagadas ao selecionar parâmetros estruturais e de treinamento. Pesquisar essas pode reafirmar certas hipóteses adotadas nessa metodologia. A complexidade interna, por exemplo, talvez possa ser analisada mais rigorosamente com algum critério estatístico, substituindo a abordagem visual. Uma maneira para lidar com a aleatoriedade das métricas entre as repetições (satisfatoriamente treinadas) auxiliaria no processo de decisão.

Sugere-se testar entradas de chuva interpolada na bacia. Agregar dados de sensoriamento remoto para zonas não monitoradas e de outros atores (*e.g.* INMET, CEMADEN) à rede de monitoramento pode ser positivo, apesar do tempo de treinamento possivelmente aumentar bastante. Precisa-se desenvolver maneiras, ainda assim, de inserir e sincronizar os dados nas RNAs (frequência, período de dados com informação de eventos variados, em oposição ao excesso de registros). Deve-se atentar que diferentes fontes podem ter diferente qualidade (*e.g.* as estações da CPRM são submetidas aos mesmos padrões de calibragem e instalação dos mesmos técnicos e engenheiros).

A manutenção e o desenvolvimento do sistema é um serviço constante, o que pode ser um impasse para a modelagem com RNAs. Quando um posto é descontinuado num local e deslocado para quilômetros a montante ou a jusante, está sendo renunciada uma série de dados que auxiliaria na descrição do estado hidrológico da bacia durante eventos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ASCE - TASK COMMITTEE ON APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY. Artificial neural networks in hydrology. I: Preliminary concepts. **Journal of Hydrologic Engineering**, United States, v. 5, n. 2, p. 115-123, abr. 2000a. DOI: 10.1061/(ASCE)1084-0699(2000)5:2(115).
- ASCE - TASK COMMITTEE ON APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN HYDROLOGY. Artificial neural networks in hydrology. II: Hydrologic applications. **Journal of Hydrologic Engineering**, United States, v. 5, n. 2, p. 124-137, abr. 2000b. DOI: 10.1061/(ASCE)1084-0699(2000)5:2(124).
- AWCHI, T. A. River discharges forecasting in northern Iraq using different ANN techniques. **Water Resources Management**, Netherlands, v. 28, n. 3, p. 801-814, 2014. DOI: 10.1007/s11269-014-0516-3.
- BATAINEH, S.; AL-ANBUKY, A.; AL-AQTASH, S. An expert system for unit commitment and power demand prediction using fuzzy logic and neural networks. **Expert Systems: The journal of Knowledge Engineering**, United Kingdom, v. 13, n. 1, p. 29-40, feb. 1996. DOI: 10.1111/j.1468-0394.1996.tb00281.x.
- BEVEN, K. Towards integrated environmental models of everywhere: Uncertainty, data and modelling as a learning process. **Hydrology and Earth System Sciences**, Germany, v. 11, n. 1, p. 460-467, 2007. DOI: 10.5194/hess-11-460-2007.
- BOTH, G. C.; HAETINGER, C.; FERREIRA, E. R.; DIEDRICH, V. L.; AZAMBUJA, J. L. F. de; Uso de modelagem matemática para a previsão de enchentes no Vale do Taquari – RS. In: VI Simpósio de Engenharia Ambiental, 6., 2008, Serra Negra/SP. **Anais VI SBEA**. 2008. Serra Negra, SP. 7 f.
- BOX, G. E. P. Science and statistics. **Journal of the American Statistics Association**, United States, v. 71, n. 356, p. 791-799, 1976. DOI: 10.1080/01621459.1976.10480949.
- BRASIL. Lei nº 6.766, de 19 de dezembro de 1979. Dispõe sobre o Parcelamento do Solo Urbano e dá outras Providências. **Diário Oficial [da] República Federativa do Brasil**, Brasília-DF, 20 dez. 1979. Disponível em: <planalto.gov.br/ccivil_03/leis/L6766.htm>. Acesso em: 18 mar. 2017.
- BRASIL. Estimativas da população para estados e municípios com data de referência em 1º de julho de 2018 feitas pelo IBGE. **Diário Oficial da União**, Brasília-DF, n. 167, p. 55-78, 29 ago. 2018. IBGE - Instituto Brasileiro de Geografia e Estatística. Disponível em: <pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=29/08/2018jornal=515pagina=55totalArquivos=134>. Acesso em: 5 abr. 2020.
- BRAGA, A. de P.; CARVALHO, A. C. P. de L.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2ª Edição. São Paulo: LTC, 2007. 248 p. ISBN 13: 978-8521615644.
- BRAVO, J. M.; PEDROLLO, O. C.; COLLISCHONN, W.; PAZ, A. R.; UVO, C. B. Previsões de curto prazo de vazão afluente ao reservatório de Furnas utilizando redes neurais artificiais. **RBRH**, Porto Alegre, v. 13, n. 2, p. 77-88, abr./jun. 2008. DOI: 10.21168/rbrh.v13n2.p77-88.
- BRUN, G. W.; TUCCI, C. E. M. Previsão em tempo real do volume afluente ao reservatório de Ernestina. **RBRH**, Porto Alegre, v. 6, n. 2, p. 73-79, abr./jun. 2001. DOI: 10.21168/rbrh.v6n2.p73-79.
- CAVALCANTI, I. F. A.; FERREIRA, N. J.; SILVA, M. G. A. J. da; DIAS, M. A. F. da S. **Tempo e Clima no Brasil**. 1ª Edição. São Paulo: Oficina de Textos, 2009. 464 p. ISBN 13: 978-85-86238-92-5.

- CHAGAS, A.; CASTILHO, A.; PEDROLLO, M.; GERMANO, A.; SOTERIO, P. The Warning Flood System in Taquari River Basin. In: 6th International Conference on Flood Management, 6., sept. 2014, São Paulo. **Proceedings of the ICFM 6**. Porto Alegre: ABRH, Sept. 2014. 12 f.
- CHIEW, F. H.; ZHOU, S. L.; MCMAHON, T. A. Use of seasonal streamflow forecasts in water resources management. **Journal of Hydrology (Amsterdam)**, Netherlands, v. 270, n. 1-2, jan. 2003. DOI: 10.1016/S0022-1694(02)00292-5.
- CHOW, V. T.; MAIDMENT, D. R.; MAYS, L. W. **Applied Hydrology**. 1^a Ed. New York: McGraw Hill, 1988. 572 p. ISBN 13: 9780070108103. Disponível em: <ponce.sdsu.edu/Applied_Hydrology_Chow_1988.pdf>. Acesso em: 17 dez. 2017.
- COLLISCHONN, W. **Simulação hidrológica de grandes bacias**. 2001. 270 f. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Saneamento Ambiental, Porto Alegre, 2001. Disponível em: <lume.ufrgs.br/handle/10183/2500>. Acesso em: 5 abr. 2020.
- COLLISCHONN, W.; TUCCI, C. E. M. Simulação hidrológica de grandes bacias. **RBRH**, v. 6, n. 1, p. 95-118, abr./jun. 2001. DOI: 10.21168/rbrh.v6n1.p95-118.
- CORPO DE ENGENHEIROS DO EXÉRCITO DOS ESTADOS UNIDOS, 2019. **HEC-HMS**. Disponível em: <hec.usace.army.mil/software/hec-hms/>. Acesso em: 4 maio 2019.
- CPRM/SGB – Serviço Geológico Brasileiro. **Bacias monitoradas**: Bacia do Caí, SACE Monitoramento. Disponível em: <cprm.gov.br/sace/index_bacias_monitoradas.php>. Acesso em: 05 abr. 2020.
- CPRM/SGB – Serviço Geológico Brasileiro. **Mapa geológico regional das unidades geológico-ambientais**. GeoSGB: dados, informações e produtos do Serviço Geológico do Brasil. Disponível em: <bit.ly/2Rg9ZTc>. Acesso em: 05 abr. 2020.
- DARCY, H. **Les fontaines publiques de la ville de Dijon**: exposition et application. 1^a Ed. Paris: Victor Dalmont, 1856. 647 p. Disponível em: <bit.ly/3bV3SvK>.
- DORNELLES, F. **Previsão contínua de níveis fluviais com redes neurais utilizando previsão de chuvas**. 2007. 99 p. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Porto Alegre, 2007. Disponível em: <lume.ufrgs.br/handle/10183/12886>. Acesso em: 5 abr. 2020.
- DRH/SEMA. **Plano de Bacia do Taquari-Antas**. Secretaria Estadual do Meio Ambiente. Estado do Rio Grande do Sul. 2011. Disponível em: <bit.ly/39LwdDg>. Acesso em: 6 abr. 2020.
- ECKHARDT, R. R. **Geração de modelo cartográfico aplicado ao mapeamento das áreas sujeitas às inundações urbanas na cidade de Lajeado/RS**. 2008. 116 p. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul, Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, Programa de Pós-Graduação em Sensoriamento Remoto, Porto Alegre, 2008. Disponível em: <lume.ufrgs.br/handle/10183/13755>. Acesso em: 5 abr. 2020.
- EMBRAPA. **Mapa pedológico do Brasil**. Mapa de Solos do Brasil na escala 1:5.000.000. Esta base de dados representa a distribuição geográfica dos solos de acordo com Sistema Brasileiro de Classificação de Solos (SiBCS, 2006). Vector data. GeoINFO, Infraestrutura de dados espaciais da EMBRAPA. Disponível em: <bit.ly/2Rg9ZTc>. Acesso em: 5 abr. 2020.

- FAN, F. M. **Previsão por conjunto de vazões afluentes a reservatórios em grandes bacias hidrográficas brasileiras**. 2015. 424 f. Tese (Doutorado) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Porto Alegre, 2015. Disponível em: <lume.ufrgs.br/handle/10183/127309>. Acesso em: 16 jun. 2019.
- FAN, F. M.; PAIVA, R. C. D.; COLLISCHONN, W. Hydrological Forecasting Practices in Brazil. In: ADAMS, T. E.; PAGANO, T. C. **Flood Forecasting: A Global Perspective**. 1ª Ed. [s. l.]: Elsevier, 2016, p. 41-66. ISBN 13: 9780128018842.
- FAN, F. M.; FLEISCHMANN, A. S.; SIQUEIRA, V. A.; NECTOUX, M.; GIACOMELLI, L.; COLLISCHONN, W. Development of a Flood Forecasting System for the State of Rio Grande do Sul, Brazil. In: EGU General Assembly 2019, 7-12 Apr. 2019, Vienna, Austria. **Geophysical Research Abstracts**. Vienna: EGU, 2019. v. 21. p. 10547-10547.
- FEPAM - Fundação estadual de proteção ambiental Henrique Luiz Roessler. **Qualidade ambiental - região hidrográfica do Guaíba**: qualidade das águas da bacia hidrográfica do rio das Antas e rio Taquari. Porto Alegre, [2012]. Disponível em: <fepam.rs.gov.br/qualidade/qualidade_taquari_antas/taquariantas.asp>. Acesso em: 24 jun. 2019.
- FERNANDO, D. A. K.; JAYAWARDENA, A. W. Runoff forecasting using RBF networks with OLS algorithm. **Journal of hydrologic engineering**, United States, v. 3, n. 3, p. 203-209, Jul. 1998. DOI: 10.1061/(ASCE)1084-0699(1998)3:3(203).
- FINCK, J. S. **Estudo preliminar de sistema de alerta para a parte brasileira da bacia do rio Uruguai utilizando modelo de previsão cota-cota**. 2018. 104 p. Trabalho de conclusão de curso (Graduação) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas e Escola de Engenharia, Porto Alegre, 2018. Disponível em: <lume.ufrgs.br/handle/10183/179893>. Acesso em: 6 abr. 2020.
- FURTADO, C. R. Intervenção do Estado e (re) estruturação urbana: um estudo sobre gentrificação. **Cad. Metropolitano**. v. 16, n. 32, p. 341-363, nov. 2014. São Paulo. DOI: 10.1590/2236-9996.2014-3203.
- GERHARD, F. **Previsão de níveis fluviais de inundação para a cidade de Lajeado, Rio Grande do Sul, a partir de diferentes abordagens matemáticas**. 2017. 136 p. Trabalho de Conclusão de Curso II (Graduação) – Universidade do Vale do Taquari, Centro de Ciências Exatas e Tecnológicas, Lajeado, 2017. Disponível em: <univates.br/bdu/handle/10737/1912>. Acesso em: 5 abr. 2020.
- GERMANO, A. de O.; CASTILHO, A.; CANDIDO, M.; REIS, F. Cheia de 2014 na bacia do rio Madeira. In: XXI Simpósio Brasileiro de Recursos Hídricos, 21., 22-27 nov. 2015. Brasília – DF. **Anais do XXI Simpósio Brasileiro de Recursos Hídricos**. Porto Alegre: ABRH. Disponível em: <rigeo.cprm.gov.br/jspui/handle/doc/15049>. Acesso em: 19 jun. 2019.
- GERMANO, A. de O.; MATOS, A. J. S.; Metodologia para definição de cotas de referência em sistemas de alerta e previsão hidrológica. In: XXII Simpósio Brasileiro de Recursos Hídricos, 22., 2017, Florianópolis, SC. **Anais do XXII Simpósio Brasileiro de Recursos Hídricos**. Porto Alegre: ABRH, 2017. Disponível em: <evoluedoc.com.br/xxiisbrh/detalhes-766_metodologia-para-definicao-de-cotas-de-referencia-em-sistemas-de-alerta-e-previsao-hidrológica>. Acesso em: 30 dez. 2017.
- GRIMM, A. M.; FERRAZ, S. E. T.; GOMES, J. Precipitation anomalies in southern south america associated with El Niño and La niña events. **Journal of Climate**, United States, v. 11, p. 2863-2880, nov. 1998. DOI: 10.1175/1520-0442(1998)011<2863:PAISBA>2.0.CO;2.

- GUASELLI, L. A.; OLIVEIRA, G. G. de; ALVES, R. de C. M. **Eventos Extremos no Rio Grande do Sul: inundações e movimentos de massa**. 1ª Ed. Porto Alegre: Evangraf, 2013. 205 p. FINEP – Financiadora de Estudos e Projetos, UFRGS. ISBN 13: 9788577276196.
- GUHA-SAPIR, D.; BELOW, R.; HOYOIS, P. **EM-DAT: International Disaster Database**, Université Catholique de Louvain, Bruxelas, Bélgica. Disponível em: <emdat.be/database>. Acesso em: 22 fev. 2019.
- GUIMARÃES, G. M. **Análise de previsibilidade de cheias na bacia do rio Uruguai através do modelo MGB-IPH**. 2018. 173 p. Trabalho de conclusão de curso (Graduação) – Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas e Escola de Engenharia, Porto Alegre, 2018. Disponível em: <lume.ufrgs.br/handle/10183/179900>. Acesso em: 6 abr. 2020.
- HAGHVERDI, A.; CORNELIS, W. M.; GHAHRAMAN, B. A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. **Journal of Hydrology (Amsterdam)**, Netherlands, v. 442, p. 46-54, jun. 2012. DOI: 10.1016/j.jhydrol.2012.03.036.
- HAYKIN, S. **Neural networks: a comprehensive foundation**. 3ª Ed. New York: Prentice Hall, 2008. 937 p. ISBN 13: 978-0-13-147139-9.
- HE, Z.; WEN, X.; LIU, H.; DU, J. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. **Journal of Hydrology (Amsterdam)**, Netherlands, v. 509, p. 379-386, Feb. 2014. DOI: 10.1016/j.jhydrol.2013.11.054.
- HEBB, D. O. **The organization of behaviour: a neuropsychological theory**. 1ª Ed. New York: John Wiley Sons, 1949. 335 p. ISBN 13: 978-0805843002.
- HECHT-NIELSEN, R. Kolmogorov's mapping neural network existence theorem. In: IEEE International Conference on Neural Networks, 1. 1987, San Diego. **Proceedings of the international conference on Neural Networks**, v. 3, p. 11-14, 1987. IEEE Press New York.
- HECHT-NIELSEN, R. **Neurocomputing**. 1ª Ed. Boston: Addison-Wesley, 1990. 433 p. ISBN 13: 978-0201093551
- HJELMFELT, A. T.; WANG, M. Predicting runoff using artificial neural networks. In: International Conference on Hydrology and Water Resources, 1996 **Proceedings of the International Conference on Hydrology and Water Resources**, New Delhi, India, December 1993. Springer, Dordrecht, 1996. p. 233-244.
- HORNIK, K. Approximation capabilities of multilayer feedforward networks. **Neural networks**, United States, v. 4, n. 2, p. 251-257, 1991. DOI: 10.1016/0893-6080(91)90009-T.
- HORTON, R. E. The rôle of infiltration in the hydrologic cycle. **Eos, Transactions American Geophysical Union**, United States, v. 14, n. 1, p. 446-460, Jun. 1933. DOI: 10.1029/TR014i001p00446.
- HSU, K.; GUPTA, H. V.; SOROOSHIAN, S. Artificial neural network modelling of the rainfall-runoff process. **Water Resources Research**, United States, v. 31, n. 10, p. 2517-2530, Oct. 1995. DOI: 10.1029/95WR01955.
- IBGE – Instituto Brasileiro de Geografia e Estatística. **Mapa físico da região sul: 1:1.500.000**. 2013. Disponível em: <170.84.40.11/cartas_e_mapas/mapas_regionais/fisico/sul_fisico1500_k_2013.pdf>. Acesso em: 6 abr. 2020.

- JAFARZADEGAN, K.; ABED-ELMDOUST, A.; KERACHIAN, R. A stochastic model for optimal operation of inter-basin water allocation systems: a case study. **Stochastic environmental research and risk assessment**, Germany, v. 28, n. 6, p. 1343-1358, Jun. 2014. DOI: 10.1007/s00477-013-0841-8.
- JAIN, A.; KUMAR, S. Dissection of trained neural network hydrologic models for knowledge extraction. **Water Resources Research**, United States, v. 45, n. 7, Jul. 2009. DOI: 10.1029/2008WR007194.
- JAYAWARDENA, A. W.; FERNANDO, D. A. K.; ZHOU, M. C. Comparison of multilayer perceptron and radial basis function networks as tools for flood forecasting. In: LEAVESLEY, G. H.; LINS, H. F.; NOBILIS, F.; PARKER, R. S.; SCHNEIDER, V. R. **Destructive water: water-caused natural disasters - their abatement and control**. 1ª Ed. Wallingford: IAHS Press, 1997. p. 173-182. ISBN 13: 978-1901502008.
- KARAMOUZ, M.; ARAGHINEJAD, S. Drought mitigation through long-term operation of reservoirs: Case study. **Journal of Irrigation and Drainage Engineering**, v. 134, n. 4, p. 471-478, Aug. 2008. DOI: 10.1061/(ASCE)0733-9437(2008)134:4(471).
- KARUNANITHI, N.; GRENNEY, W. J.; WHITLEY, D.; BOVEE, K. Neural networks for river flow prediction. **Journal of Computing in Civil Engineering**, v. 8, n. 2, p. 201-220, Apr. 1994. DOI: 10.1061/(ASCE)0887-3801(1994)8:2(201).
- KICH, E. de M.; MELATI, M. D.; MARCUZZO, F. F. N. Estudo do regime hídrico pluvial e fluvial na sub-bacia 86 visando a determinação do seu ano hidrológico. In: XXI Simpósio Brasileiro de Recursos Hídricos, 21., 22-27 nov. 2015. Brasília – DF. **Anais do XXI Simpósio Brasileiro de Recursos Hídricos**. Porto Alegre: ABRH.
- KIRCHNER, J. W. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. **Water Resources Journal**, v. 42, n. 3, 5 f., Mar. 2006. DOI: 10.1029/2005WR004362.
- KITANIDIS, P. K.; BRAS, R. L. **Real time forecasting of river flows**. Cambridge: MIT, 1978.
- KOBIYAMA, M.; MENDONÇA, M.; MORENO, D. A.; MARCELINO, I. P. V. de O.; MARCELINO, E. V.; GONÇALVES, E. F.; BRAZETTI, L. L. P.; GOERL, R. F.; MOLLERI, G. S. F.; RUDORF, F. de M. **Prevenção de desastres naturais: conceitos básicos**. 1ª Ed. Florianópolis – SC: Organic Trading, 2006. 109 p. ISBN 10: 858775503X.
- KOLMOGOROV, A. N. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In: **Doklady Akademii Nauk**. Russian Academy of Sciences, v. 114, n. 5, 1957. p. 953-956. (Russo) Disponível em: <bit.ly/2JKxXSp>. Acesso em: 6 abr. 2020.
- KÖPPEN, W.; GEIGER, R. Das geographische System der Klimate. In: KÖPPEN, W. (Org.); GEIGER, R. (Org.); ALT, E.; BERG, L.; BIRKELAND, B. J.; BORCHARDT, W.; BRAAK, C.; BROOKS, F.; CASTENS, G.; CONNOR, A.; CONRAD, V.; COX, G. W.; FÖYN, N. J.; FUJIWHARA, S.; KAMINSKI, A.; KIDSON, E.; KNOCH, K.; KUHLBRODT, E.; LOEWE, F.; MEINARDUS, W.; MILANKOVITCH, M.; NORMAND, C. W. B.; OKADA, T.; PETERSEN, H.; ROBERTSON, C. L.; RUBISTEIN, E.; SAPPER, K.; SCHOTT, G.; SELICK, N. P.; SEMMELHACK, W.; SVERDRUP, H. U.; TAYLOR, T. Gr.; WAGNER, A.; WALKER, Sir Gilbert T.; WARD, R. DeC.; WEGENER, K.; WEICKMANN, L. **Handbuch der Klimatologie: in fünf Bände**. Berlin: Gebrüder Bornträger, 1936, Band 1, Teil C, p. 44.

- KUREK, R. K. M. **Análise das inundações no vale do Taquari/RS como subsídio à elaboração de um modelo de previsão**. 2016. 178 p. Dissertação (Mestrado) – Universidade Federal de Santa Maria, Programa de Pós-Graduação em Engenharia Civil, Área de concentração Recursos Hídricos e Saneamento Ambiental, Santa Maria, 2016. Disponível em: <repositorio.ufsm.br/handle/1/7908>. Acesso em: 5 abr. 2020.
- LANNA, A. E.; INDURSKY, A. B. Predição de vazões médias mensais com modelos empíricos. **Revista Brasileira de Engenharia**, Caderno de Recursos Hídricos, ABRH, Brasil, v. 1, p. 59-91, 1983.
- LEK, S.; DELACOSTE, M.; BARAN, P.; DIMOPOULOS, I.; LAUGA, J.; AULAGNIER, S. Application of neural networks to modelling nonlinear relationships in ecology. **Ecological Modelling**, Netherlands, v. 90, n. 1, p. 39-52, Sept. 1996. DOI: 10.1016/0304-3800(95)00142-5.
- LEGATES, D. R.; DAVIS, R. E. The continuing search for an anthropogenic climate change signal: limitations of correlation-based approaches. **Geophysical Research Letters**, United States, v. 24, n. 18, p. 2319-2322, Sept. 1997. DOI: 10.1029/97GL02207.
- LIU, Z.; GUO, S.; ZHANG, H.; LIU, D.; YANG, G. Comparative study of three updating procedures for real-time flood forecasting. **Water Resources Management**, Netherlands, v. 30, n. 7, p. 2111-2126, Mar. 2016. DOI: 10.1007/s11269-016-1275-0.
- LOPES, J. E.; BRAGA, B. F. F.; CONEJO, J. L. “SMAP - A Simplified Hydrologic Model”, Applied Modeling in Catchment Hydrology. **Water Resources Publication**, 1982. 1 ed. V. P. Singh, Colorado.
- LORENZ, E. N. The predictability of a flow which contains many scales of motion. **Tellus**, United Kingdom, v. 21, n. 3, p. 289-307, Jun. 1969. DOI: 10.1111/j.2153-3490.1969.tb00444.x.
- MapBiomass Project. **Collection v. 4.0 of the Annual Land Use Land Cover Maps of Brazil**, accessed in 2019 through the link: mapbiomas.org/. MapBiomass Project - is a multi-institutional initiative to generate annual land use land cover maps based on automatic classification processes applied to satellite images. The complete project description can be found at mapbiomas.org.
- MATHWORKS. **Matlab** for Windows, 2012b. Matick, Massachusetts.
- MATOS, A. B. de; PEDROLLO, O. C.; CASTRO, N. M. dos R. Efeito do controle de montante de sub-bacias embutidas na previsão hidrológica de curto prazo com redes neurais: aplicação à bacia de ponte mística. **RBRH**, v. 19, n. 1, p. 87-99, jan./mar. 2014. DOI: 10.21168/rbrh.v19n1.p87-99.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **Bulletin of mathematical biophysics**, United States, v. 5, p. 115-133, Dec. 1943. DOI: 10.1007/BF02478259.
- MELO, T. M. de; PEDROLLO, O. C. Artificial neural networks for estimating soil water retention curve using fitted and measured data. **Applied and Environmental Soil Science**, United Kingdom, v. 2015, 16 f., Mar. 2015. DOI: 10.1155/2015/535216.
- MINSKY, M.; PAPERT, S. **Perceptrons: An introduction to computational geometry**. 1ª Ed. Cambridge (EUA): Cambridge Press, 1969. ISBN 13: 9780262630221.
- MOORE, R. J. **Transfer functions, noise predictors and the forecasting of flood events in real-time**. In: Singh, V.P. (ed.), *Statistical Analysis of Rainfall and Runoff*, p. 229-250, Water Resources Publications, Littleton, Colorado, USA. 1982.

- MORIASI, D. N.; ARNOLD, J. G.; VAN LIEW, M. W.; BINGNER, R. L.; HARMEL, R.D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. **American Society of Agricultural and Biological Engineers**, v. 50, n. 3, p. 885-900, May 2007. DOI: 10.13031/2013.23153.
- NASA/NGA – National Aeronautics and Space Administration/National Geospatial-Intelligence Agency. **Shuttle Radar Topography Mission 1 Arc-Second Global**: Feb. 11-22 2000. Imagens obtiveis pelo portal Earth Explorer da USGS-EUA desde 2014. DOI: 10.5066/F7PR7TFT. Disponível em: <earthexplorer.usgs.gov/>. Acesso em: 20 jun. 2017.
- NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part I: a discussion of principles. **Journal of Hydrology**. v.10, n.3, p. 282-290, 1970. DOI: 10.1016/0022-1694(70)90255-6.
- NIH/UNIVATES – **Núcleo de Informações Hidrometeorológicas da Universidade do Vale do Taquari**. 2019. Informações disponibilizadas através do e-mail:<nih@univates.br> no dia 25/06/2019.
- NOURANI, V.; BAGHANAM, A. H.; ADAMOWSKI, J.; KISI, O. Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. **Journal of Hydrology (Amsterdam)**, Netherlands v. 514, p. 358-377, Sept. 2014. DOI: 10.1016/j.jhydrol.2014.03.057.
- NOURANI, V. An Emotional ANN (EANN) approach to modeling rainfall-runoff process. **Journal of Hydrology (Amsterdam)**, Netherlands, v. 544, p. 267-277, Jan. 2017. DOI: 10.1016/j.jhydrol.2016.11.033.
- OGURA, A.; MACEDO, E. S. Procesos y riesgos geologicos. In: **II Curso internacional de aspectos geologicos de protección ambiental**: notas de clases. Montevideo: UNESCO, 2002. p. 114-137.
- OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R.; BRAVO, J. M. Simulações hidrológicas com diferentes proporções de área controlada na bacia hidrográfica. **RBRH**, Brasil, v. 18, n. 3, p. 193-204, jul./set. 2013. DOI: 10.21168/rbrh.v18n3.p193-204.
- OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. O desempenho das Redes Neurais Artificiais (RNAs) para simulação hidrológica mensal. **RBRH**, Brasil, v. 19, n. 2, p. 251-265, abr./jun. 2014. DOI: 10.21168/rbrh.v19n2.p251-265.
- OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. As incertezas associadas às condições climáticas obtidas pelo modelo ETA CPTEC/HADCM3: avaliação comparativa entre os dados simulados e observados de precipitação, evapotranspiração e vazão na bacia hidrográfica do rio Ijuí, Brasil. **Revista Brasileira de Meteorologia**, Brasil, v. 30, n. 1, p. 101-121, mar. 2015a. DOI: 10.1590/0102-778620140041.
- OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. Stochastic approach to analyzing the uncertainties and possible changes in the availability of water in the future based on scenarios of climate change. **Hydrology and Earth System Sciences**, Germany, v. 19, n. 8, p. 3585-3604, Aug. 2015b. DOI: 10.5194/hess-19-3585-2015.
- OLIVEIRA, G. G.; PEDROLLO, O. C.; CASTRO, N. M. R. Simplifying artificial neural network models of river basin behaviour by an automated procedure for input variable selection. **Engineering Applications of Artificial Intelligence**, United Kingdom, v.40, p. 47-61. Apr. 2015c. DOI: 10.1016/j.engappai.2015.01.001.
- OLIVEIRA, M. H. C. de; SARI, V.; CASTRO, N. M. dos R.; PEDROLLO, O. C. Estimation of soil water content in watershed using artificial neural networks. **Hydrological Sciences Journal**, United Kingdom, v. 62, n. 13, p. 2120-2138, Aug. 2017. DOI: 10.1080/02626667.2017.1364844.

- PAGANO, T.; WOOD, A.; RAMOS, M.; CLOKE, H.; PAPPENBERGER, F.; CLARK, M.; CRANSTON, M.; KAVETSKI, D.; MATHEVET, T.; SOROOSHIAN, S.; VERKADE, J. Challenges of Operational River Forecasting. **Journal Hydrometeorology**, United States, v. 15, p. 1692-1707, Jul. 2014. DOI: 10.1175/JHMD-13-0188.1.
- PEDROLLO, M.; GERMANO, A.; SOTÉRIO, P.; RODRIGUES, E.; MADUELL, J. C. Alerta Hidrológico da Bacia do Rio Caí: concepção e implantação do sistema. In: XIX Simpósio Brasileiro de Recursos Hídricos, 19., 27 nov./1º dez. 2011, Maceió - AL. **Anais do XIX SBRH**. Porto Alegre: ABRH, 2011. Disponível em: <bit.ly/39Merj9>. Acesso em: 6 abr. 2020. 14 f.
- PEDROLLO, O.; PEDROLLO, M. Sistema de alerta de cheias da bacia do rio Caí: previsão de níveis com redes neurais artificiais. In: GUASELLI, L. A.; OLIVEIRA, G. G. de; ALVES, R. de C. M. **Eventos Extremos no Rio Grande do Sul: inundações e movimentos de massa**. 1ª Ed. Porto Alegre: Evangraf, 2013. cap. 5, p. 101-120. ISBN 13: 9788577276196.
- PEDROLLO, M.; SOTÉRIO, P.; PEDROLLO, O. Operation of the hydrological warning system in the Caí river basin preliminary studies. In: 6th International Conference on Flood Management, 6., sept. 2014, São Paulo. **Proceedings of the ICFM 6**. Porto Alegre: ABRH, Sept. 2014. 12 f.
- PEDROLLO, O. C. Previsão de curto prazo de níveis com redes neurais artificiais para a cidade de Estrela (RS): resultados preliminares. In: XXII Simpósio Brasileiro de Recursos Hídricos, 22., 26 nov./1 dez. 2017, Florianópolis – SC. **Anais do XXII SBRH**. Porto Alegre: ABRH, 2017. 9 f. Disponível em: <lume.ufrgs.br/handle/10183/173852>. Acesso em: 6 abr. 2020.
- QGIS, Development Team. QGIS: sistema de informação geográfica para área de trabalho de computador. Versão 3.10.4. 2020. Licença de software livre GNU GPL. Disponível em: <qgis.org/en/site/forusers/download.html>. Acesso em: 6 abr. 2020.
- RIO GRANDE DO SUL. **Plano estratégico de desenvolvimento do Vale do Taquari**. 2017. Disponível em: <bit.ly/3e4MBCf>. Acesso em: 6 abr. 2020.
- RITTER, A.; MUÑOZ-CARPENA, R. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. **Journal of Hydrology (Amsterdam)**, Netherlands, v 480, p. 33-45, Feb. 2013. DOI: 10.1016/j.jhydrol.2012.12.004.
- ROBAINA, L. E. de S. Áreas de risco por processos de dinâmica geomorfológica: introdução ao tema. In: GUASELLI, L. A.; OLIVEIRA, G. G. de; ALVES, R. de C. M. **Eventos Extremos no Rio Grande do Sul: inundações e movimentos de massa**. 1ª Ed. Porto Alegre: Evangraf, 2013. cap. 1, p. 11-24. ISBN 13: 9788577276196.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, United States, v. 65, n. 6, p. 386, 1958. DOI: 10.1037/h0042519.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, United Kingdom, v. 323, n. 6088, p. 533, 1986. DOI: 10.1038/323533a0.
- SARTORI, A.; LOMBARDI NETO, F.; GENOVEZ, A. M. Classificação hidrológica de solos brasileiros para a estimativa da chuva excedente com o método do Serviço de Conservação do Solo dos Estados Unidos Parte 1: Classificação. **RBRH**, Brasil, v. 10, n. 4, p. 05-18, out./dez. 2005. DOI: 10.21168/rbrh.v10n4.p5-18.
- SEMA, 2019. **Sistema Estadual de Gestão Integrada de Risco de Desastres**. Disponível em: <sema.rs.gov.br/sistema-estadual-de-gestao-integrada-de-risco-de-desastres>. Acesso em: 16 jun. 2019.

- SILVA, J. X. da; R. T. ZAIDAM (org). **Geoprocessamento Análise Ambiental: aplicações**. 7ª Ed. Rio de Janeiro: Bertrand Brasil, 2004. 366 p. ISBN 13: 978-8528610765.
- SINGH, K. P.; MALIK, A.; MOHAN, D.; SINHA, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India): a case study. **Water Research (Oxford)**, United Kingdom, v. 38, n. 18, p. 3980-3992, Nov. 2004. DOI: 10.1016/j.watres.2004.06.011.
- SIQUEIRA, V. A.; FLEISCHMANN, A. S.; COLLISCHONN, W.; FAN, F. M. Desenvolvimento de um sistema operacional de previsão de cheias na bacia do Taquari-Antas/RS. In: XXI Simpósio Brasileiro de Recursos Hídricos, 21., 22-27 nov. 2015. Brasília – DF. **Anais do XXI Simpósio Brasileiro de Recursos Hídricos**. Porto Alegre: ABRH. 2015a.
- SIQUEIRA, V. A. **Previsão de cheias por conjunto em curto a médio prazo: bacia do Taquari-Antas/RS**. 2015. 168 p. Dissertação de mestrado (Mestre em Engenharia) – Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015b. Disponível em: <lume.ufrgs.br/handle/10183/147137>. Acesso em: 6 abr. 2020.
- SIQUEIRA, V. A.; COLLISCHONN, W.; FAN, F. M.; CHOU, S. C. Ensemble flood forecasting based on operational forecasts of the regional Eta EPS in the Taquari-Antas basin. **RBRH**, v. 21, n. 3, p. 587-602, Jul./Sept. 2016. DOI: 10.1590/2318-0331.011616004.
- SHOAIB, M.; SHAMSELDIN, A. Y.; KHAN, S.; KHAN, M. M.; KHAN, Z. M.; SULTAN, T.; MELVILLE, B. W. A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. **Water Resources Management**, Netherlands, v. 32, n. 1, p. 83-103, Aug. 2017. DOI: 10.1007/s11269-018-2151-x.
- TAPIA, R. P.; PINEDA, R. V.; OLIVARES, C.; GONZÁLEZ, P. A. Development of upstream data-input models to estimate downstream peak flow in two mediterranean river basins of Chile. **Open Journal of Modern Hydrology (Print)**, United States, Oct. 2014. v. 4, n. 4 p. 132-143. DOI: 10.4236/ojmh.2014.44013.
- TAWFIK, M.; IBRAHIM, A.; FAHMY, H. Hysteresis sensitive neural network for modelling rating curves. **Journal of Computing in Civil Engineering**, United States, v. 11, n. 3, p. 206-211, Jul. 1997. DOI: 10.1061/(ASCE)0887-3801(1997)11:3(206).
- TOKAR, A. S.; JOHNSON, P. A. Rainfall-runoff modelling using artificial neural networks. **Journal of Hydrologic Engineering**, United States, v. 4, n. 3, p. 232-239, Jul. 1999. DOI: 10.1061/(ASCE)1084-0699(1999)4:3(232).
- TUCCI, C. E. M. **Modelos hidrológicos: Revista e Ampliada**. 2ª Ed. Porto Alegre: UFRGS. 2005. 678 p. Colaboração da ABRH/ABRH. ISBN 13: 9788570258236.
- TUCCI, C. E. M.; PORTO, R. L. L.; BARROS, M. T. de. **Drenagem Urbana**. Porto Alegre: ABRH, 1995. 428 p. Coleção ABRH de Recursos Hídricos. ISBN 13: 9788570253644.
- UNDRR - United Nations International Strategy for Disaster Reduction/United Nations Office for Disasters Risk Reduction. **Prevention Web**. Terminology. (*anteriormente sob a sigla UNIS DR*) Disponível em: <preventionweb.net/terminology>. Acesso em: 25 Mar. 2020.
- VALIPOUR, M.; BANIHABIB, M. E.; BEHBAHANI, S. M. R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. **Journal of hydrology (Amsterdam)**, Netherlands, v. 476, p. 433-441, Jan. 2013. DOI: 10.1016/j.jhydrol.2012.11.017.
- VOGL, T. P.; MANGIS, J. K.; RIGLER, A. K.; ZINK, W. T.; ALKON, D. L. Accelerating the convergence of the back-propagation method. **Biological cybernetics**, Germany, v. 59, n. 4-5, p. 257-263, Sept. 1988. DOI: 10.1007/BF00332914.

- WANG, L.; LI, X.; MA, C.; BAI, Y. Improving the prediction accuracy of monthly streamflow using a data-driven model based on a double-processing strategy. **Journal of Hydrology (Amsterdam)**, Netherlands, v. 573, p. 733-745, Jun. 2019. DOI: 10.1016/j.jhydrol.2019.03.101.
- WIDROW, B.; HOFF, M. E. **Adaptive switching circuits**. In: IRE WESCON Convention, [] Aug. 23-26 1960, Los Angeles. **IRE WESCON Convention Record**, Part 4. New York: Institute of Radio Engineers, 1960. p. 96-104.
- ZADEH, L. A. Fuzzy sets. **Information and Control** continued as **Information and Computation (Printed)**, United States, v. 8, p. 338-353, Jun. 1965. DOI: 10.1016/S0019-9958(65)90241-X.
- ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: The state of the art. **International journal of forecasting**, Netherlands, v. 14, n. 1, p. 35-62, Mar. 1998. DOI: 10.1016/S0169-2070(97)00044-7.

8. APÊNDICE

Tabela 17. Modelos que compõe a Árvore de RNAs “rna_ENC”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.

Modelo	pPT t	pPV t	pPI t	pSC t	pMU t	pEN t	fMU t-dif	fMU t	fEN t-dif	fEN t
CM	1	1	1	1	1	1	1	1	1	1
PM1	0	1	1	1	1	1	1	1	1	1
PM2	1	0	1	1	1	1	1	1	1	1
PM3	1	1	0	1	1	1	1	1	1	1
PM4	1	1	1	0	1	1	1	1	1	1
PM5	1	1	1	1	0	1	1	1	1	1
PM6	1	1	1	1	1	0	1	1	1	1
PM7	1	1	1	1	1	1	0	1	1	1
PM8	1	1	1	1	1	1	1	0	1	1
PM9	1	1	1	1	1	1	1	1	0	1
PM10	1	1	1	1	1	1	1	1	1	0
PM11	0	0	0	0	0	0	1	1	1	1
PM12	1	1	1	1	1	1	0	0	0	0
PM13	1	0	1	1	0	1	0	0	1	1
PM14	1	0	1	1	0	1	1	1	1	1
PM15	1	0	1	1	0	0	0	0	1	1
PM16	1	1	1	1	0	0	0	0	1	1
PM17	1	0	1	1	0	0	1	1	1	1
PM18	1	1	1	1	0	0	1	1	1	1
PM19	0	1	1	0	1	1	1	1	1	1
PM20	0	1	1	0	0	1	0	0	1	1
PM21	0	1	1	0	0	1	1	1	1	1
PM22	0	1	1	0	0	0	1	1	0	0
PM23	0	1	1	0	0	0	1	1	1	1
PM24	0	1	1	0	1	0	1	1	1	1
PM25	1	1	1	0	1	0	1	1	1	1
PM26	0	1	0	0	0	0	1	1	1	1
PM27	0	1	1	1	0	0	1	1	1	1
PM28	0	1	0	1	0	0	1	1	1	1
PM29	0	1	1	1	0	1	1	1	1	1

Tabela 17 (conclusão). Modelos que compõe a Árvore de RNAs “rna_ENC”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.

Modelo	pPT t	pPV t	pPI t	pSC t	pMU t	pEN t	fMU t-dif	fMU t	fEN t-dif	fEN t
PM30	1	0	0	1	1	1	1	1	1	1

Tabela 18. Modelos que compõe a Árvore de RNAs “rna_EST”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.

Modelo	pPT t	pPV t	pPI t	pSC t	pMU t	pEN t	pES t	fMU t-dif	fMU t	fEN t-dif	fEN t	fES t-dif	fES t
CM	1	1	1	1	1	1	1	1	1	1	1	1	1
PM1	0	1	1	1	1	1	1	1	1	1	1	1	1
PM2	1	0	1	1	1	1	1	1	1	1	1	1	1
PM3	1	1	0	1	1	1	1	1	1	1	1	1	1
PM4	1	1	1	0	1	1	1	1	1	1	1	1	1
PM5	1	1	1	1	0	1	1	1	1	1	1	1	1
PM6	1	1	1	1	1	0	1	1	1	1	1	1	1
PM7	1	1	1	1	1	1	0	1	1	1	1	1	1
PM8	1	1	1	1	1	1	1	0	1	1	1	1	1
PM9	1	1	1	1	1	1	1	1	0	1	1	1	1
PM10	1	1	1	1	1	1	1	1	1	0	1	1	1
PM11	1	1	1	1	1	1	1	1	1	1	0	1	1
PM12	1	1	1	1	1	1	1	1	1	1	1	0	1
PM13	1	1	1	1	1	1	1	1	1	1	1	1	0
PM14	0	0	0	0	0	0	0	1	1	1	1	1	1
PM15	1	1	1	1	1	1	1	0	0	0	0	0	0
PM16	1	0	1	1	0	1	1	0	0	1	1	1	1
PM17	1	0	1	1	0	1	1	1	1	1	1	1	1
PM18	1	0	1	1	0	0	1	0	0	1	1	1	1
PM19	1	1	1	1	0	0	1	1	1	1	1	1	1
PM20	1	1	1	1	0	0	1	0	0	1	1	1	1
PM21	0	1	1	0	1	1	1	1	1	1	1	1	1
PM22	0	1	1	0	0	1	1	0	0	1	1	1	1
PM23	0	1	1	0	0	0	1	1	1	0	0	1	1
PM24	0	1	1	0	0	0	1	1	1	1	1	1	1
PM25	1	1	1	0	1	0	1	1	1	1	1	1	1
PM26	1	0	1	1	1	1	0	1	1	1	1	1	1

Tabela 20. Modelos que compõe a Árvore de RNAs “rna_TA”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.

Nome	pPT t	pPV t	pPI t	pSC t	pMU t	pEN t	pES t	pTA t	fMU t-dif	fMU t	fEN t-dif	fEN t	fES t-dif	fES t	fPM t-dif	fPM t	fTA t-dif	fTA t
CM	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM3	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM4	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM5	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
PM6	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
PM7	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
PM8	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
PM9	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
PM10	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1
PM11	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
PM12	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
PM13	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
PM14	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
PM15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
PM16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
PM17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
PM18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
PM19	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
PM20	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
PM21	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
PM22	1	0	1	1	0	1	1	1	0	0	1	1	1	1	0	0	1	1
PM23	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1
PM24	1	0	1	1	0	0	1	1	0	0	1	1	1	1	0	0	1	1
PM25	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1
PM26	1	1	1	1	0	0	1	1	1	1	1	1	1	1	0	0	1	1
PM27	0	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1
PM28	0	1	1	0	1	1	1	0	1	1	1	1	1	1	0	0	0	0
PM29	0	1	1	0	1	1	1	0	1	1	1	1	1	1	0	0	1	1
PM30	0	1	1	0	0	1	1	0	1	1	1	1	1	1	0	0	1	1
PM31	0	1	1	0	0	0	1	0	1	1	0	0	1	1	0	0	1	1
PM32	0	1	1	0	0	0	1	0	1	1	1	1	1	1	0	0	1	1
PM33	0	1	1	0	0	0	1	1	1	1	1	1	1	1	0	0	1	1
PM34	0	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1
PM35	1	1	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	1
PM36	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	1
PM37	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
PM38	1	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1
PM39	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1

Tabela 21 (conclusão). Modelos que compõe a Árvore de RNAs “rna_TA”. Se 1, a variável de entrada é utilizada no modelo; Se 0, a variável de entrada não é utilizada no modelo.

Nome	pPT t	pPV t	pPI t	pSC t	pMU t	pEN t	pES t	pTA t	fMU t-dif	fMU t	fEN t-dif	fEN t	fES t-dif	fES t	fPM t-dif	fPM t	fTA t-dif	fTA t
PM40	1	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1
PM41	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM42	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1
PM43	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
PM44	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1
PM45	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1
PM46	1	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
PM47	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1
PM48	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
PM49	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1
PM50	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
PM51	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM52	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
PM53	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1
PM54	0	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
PM55	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
PM56	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
PM57	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1
PM58	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PM59	0	1	1	0	0	1	1	0	0	0	1	1	1	1	0	0	1	1