



APLICAÇÃO DAS REDES NEURAIS ARTIFICIAIS NO PREENCHIMENTO DE DADOS DIÁRIOS DE CHUVA NO ESTADO DE SÃO PAULO

Deborah Fernanda Santana Gimenez ^(a), Jonas Teixeira Nery ^(b)

^(a)Programa de Pós-graduação em Geografia da Universidade Estadual de Campinas - UNICAMP. (deborahsntna@gmail.com)

^(b)Livre Docente. Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP. (jonas@ourinhos.unesp.br).

Eixo: Climatologia em diferentes níveis escalares: Mudanças e Variabilidades

Resumo

A precipitação pluvial é uma grandeza que apresenta grande variabilidade espacial e temporal e para analisá-la de forma apurada é necessário dispor de séries de qualidade. No entanto, existem dificuldades em encontrar dados consistentes porque as falhas (humanas e de aparelhos) são muito comuns. Neste trabalho, foram utilizadas 171 séries temporais obtidas por meio do sistema *Hidroweb* abrangendo o período em comum de 37 anos (1976 a 2012). Os registros de cada posto pluviométrico foram preenchidos utilizando conjuntos de estações próximas com o uso da técnica das Redes Neurais Artificiais *Perceptron* Multicamadas através de sub-rotina no ambiente R. Os resultados apresentaram-se satisfatórios, 159 postos pluviométricos tiveram um bom preenchimento, 12 foram descartados por conterem uma grande quantidade de dados faltantes, o que impossibilitou a aprendizagem pelas Redes Neurais. Considerou-se que o preenchimento com esta técnica se mostrou uma boa alternativa para a obtenção de dados sólidos para análises da chuva.

Palavras chave: precipitação pluvial, redes neurais artificiais, dados diários, São Paulo

1. Introdução

A precipitação pluvial é uma grandeza que apresenta grande variabilidade espacial e temporal e para analisá-la de forma apurada é necessário dispor de séries de qualidade.

A medida da altura da precipitação pluvial é feita através dos pluviômetros instalados nas estações hidrométricas, onde os dados são armazenados, qualificados e retransmitidos à Agência Nacional de Águas (ANA), que depois os disponibiliza ao público geral por meio do sistema *Hidroweb*.

Os aparelhos devem ser instalados e mantidos obedecendo às orientações da ANA, para que os dados tenham qualidade. No caso dos pluviômetros, estes devem ser instalados em terreno plano, livre de obstáculos e de riscos de inundação, (ANA, 2011). “A superfície de captação do pluviômetro deve estar num plano horizontal, não deve apresentar deformações e estar a uma altura de 1,5 metros acima do solo.



Os obstáculos deverão estar a uma distância igual ou superior a duas vezes a altura do obstáculo com relação à superfície de captação dos pluviômetros.” (ANA, 2011, p. 12). Ver Figura 1.

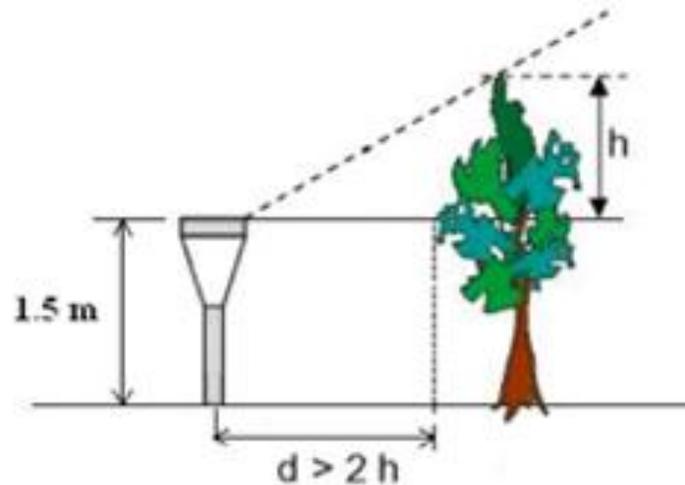


Figura 1 - Esquema de posicionamento do pluviômetro em relação a obstáculos.

Fonte: Plano de instalação da ANA (Agência Nacional de Águas)

Mesmo com a orientação, regulamentação e fiscalização, os equipamentos podem apresentar erros não esperados em suas medições, seja por falhas nos sensores, erros de transcrição ou até mesmo a indesejada falta de manutenção, o que pode prejudicar a qualidade dos dados. Desta forma, um dos problemas comuns quando se trabalha com a chuva é a falta de dados consistentes. Para modelar a variabilidade dessa grandeza de forma apropriada é necessário que os dados estejam adequados para que os resultados não se apresentem de forma incorreta, muito longe da realidade.

Para sanar o problema da falta de dados meteorológicos e hidrológicos, o método das Redes Neurais Artificiais (RNA's) tem sido amplamente empregado. Piazza et al. (2011) utiliza o descarte dos registros com dados faltantes. Entretanto, nos dias atuais, com o advento de computadores capazes de processar estatísticas refinadas, pode-se completar um grande número de lacunas em séries por meio de interpoladores. Convém destacar que o uso da técnica deve acompanhar a leitura atenta dos resultados obtidos, visto que as estimativas podem distar do real, pois existem a variabilidade natural da precipitação e a ocorrência de eventos anômalos.

Vários estudos já demonstraram técnicas de correção e preenchimento de falhas tanto diárias quanto mensais. Guijarro (2004) homogeneiza dados mensais de precipitação através do método do vizinho mais próximo e do inverso do quadrado da distância, utilizando uma série de referência. Nery e Carfan (2004)



utilizaram este mesmo método de homogeneização para consistir e preencher dados de precipitação da região sul do Brasil.

O uso das redes neurais é recente nestes tipos de estudo, mas vem sendo testada e segundo Teegavarapu e Chandramouli (2005), vem apresentando resultados satisfatórios em pesquisas hidrológicas que visam estimar dados faltantes de precipitação pluvial.

Quando o cérebro de um ser vivo nasce, ele tem a habilidade de construir suas próprias regras e comportamentos, e através da experiência adquirida ele se desenvolve ao longo do tempo, adapta-se e continua a se desenvolver. As Redes Neurais Artificiais, comumente chamadas de “redes neurais” são técnicas computacionais inspiradas na estrutura neural dos organismos vivos e que adquirem conhecimento através da experiência, da mesma forma que o cérebro (HAYKIN, 2008).

Ainda segundo Haykin (2008), o procedimento utilizado para realizar o processo de aprendizagem é chamado de algoritmo de aprendizagem, cuja função é modificar os pesos sinápticos da rede de forma ordenada para alcançar um objetivo de projeto desejado.

As redes neurais foram primeiramente introduzidas em 1943 por McCulloch e Pitts. Depois da introdução do conceito, Rosenblatt propôs em 1958 o perceptron (de camada única) como o primeiro modelo de aprendizagem supervisionada. A partir disso, diversos estudos foram aperfeiçoando a ideia de redes neurais até que se chegou ao MultilayerPerceptron (MLP), ou perceptronmulti-camadas, que é mais sofisticado, podendo ser utilizadas, como o próprio nome sugere, mais do que uma camada (HAYKIN, 2008).

Basicamente, as redes neurais perceptronmulti-camadas utilizam uma camada de entrada, camadas escondidas e a camada de saída. A Figura 2 refere-se ao esquema da estrutura de uma rede neural perceptronmulti-camada com duas camadas escondidas.

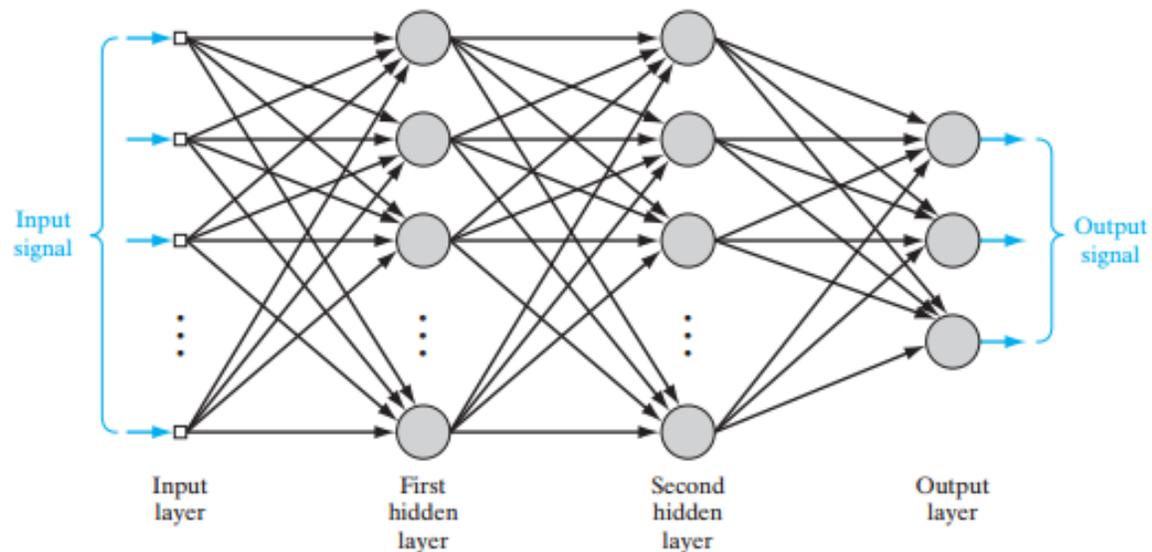


Figura 2 - Estrutura de uma rede neural artificial perceptron multicamada.
Fonte: HAYKIN (2008)

2. Materiais e métodos

Neste trabalho utilizou-se as RNA's como método para correção de dados diários faltantes. Os cálculos (entradas) foram realizados mediante sub-rotina criada para o software R, que é livre e multiplataforma (pode ser utilizado em diferentes sistemas operacionais). De acordo com Amaral e Cesário (2009), R é um conjunto integrado de facilidades de programas para manipulação de dados, cálculo e visualização gráfica, totalmente livre.

No R, existe uma grande diversidade de pacotes. Estes pacotes contêm um conjunto de funções que permitem ou facilitam a realização das análises. Neste trabalho, o pacote (sub-rotina) utilizado foi o *prefann* (*PrecipitationForecastingwith Artificial Neural Network andothers*), desenvolvido por Povia (2015), que tem como função o preenchimento de dados de precipitação faltantes por meio das Redes Neurais Artificiais perceptron multicamadas.

Para a obtenção dos dados, foi necessário realizar o download das séries de precipitação pluvial por meio do site da Agência Nacional de Águas (<http://hidroweb.ana.gov.br/>). Após o download de cerca de 5000 postos pluviométricos, abrangendo todo o estado de São Paulo e adjacências, foi feita a organização e tabulação dos dados para a obtenção de um período de estudo em comum superior a 30 anos. Desta forma, foi possível obter um total de 171 séries temporais bem distribuídas ao longo do estado de São Paulo (Figura 3) abrangendo o período entre 1976 a 2012 (37 anos).

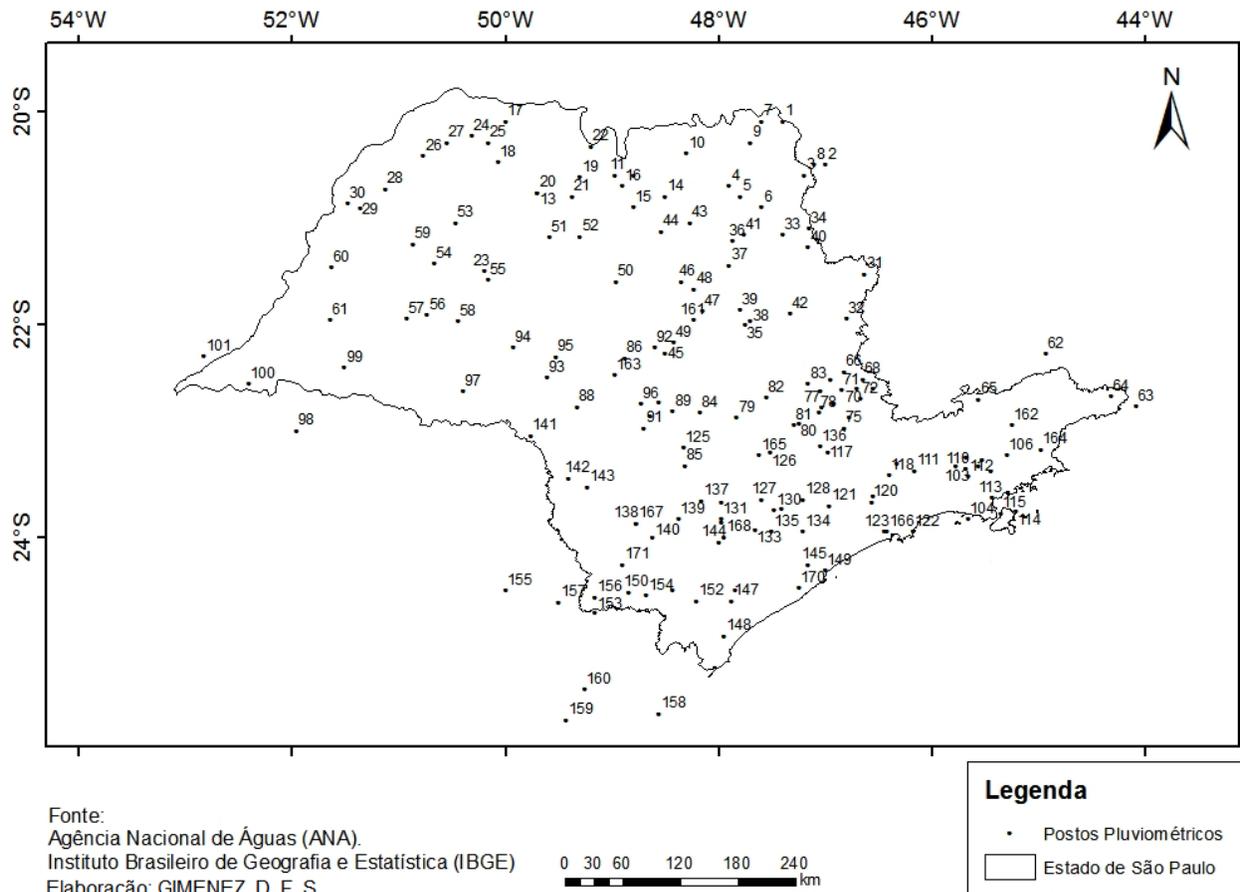


Figura 3 - Mapa de localização dos postos pluviométricos do estado de São Paulo.

Os arquivos (planilha ou documento do bloco de notas) obtidos foram organizados e salvos em formato “.csv”. Para o preenchimento, foi necessário iniciar o R e carregar o pacote `prefann` inserindo o comando `require(prefann)`, em seguida, os comandos para efetuar o preenchimento. Os dados obtidos através do preenchimento foram salvos em planilha eletrônica novamente, no mesmo formato da entrada dos dados.

Para o preenchimento de falhas foram utilizadas, no mínimo, três séries históricas de postos vizinhos à estação em questão. Neste estudo estabeleceu-se que seriam utilizadas quantas séries fossem necessárias para o preenchimento, conforme a disponibilidade, proximidade, a altitude e observando sempre o critério de haver pelo menos uma série histórica sem dados faltantes (ou com falhas ínfimas).

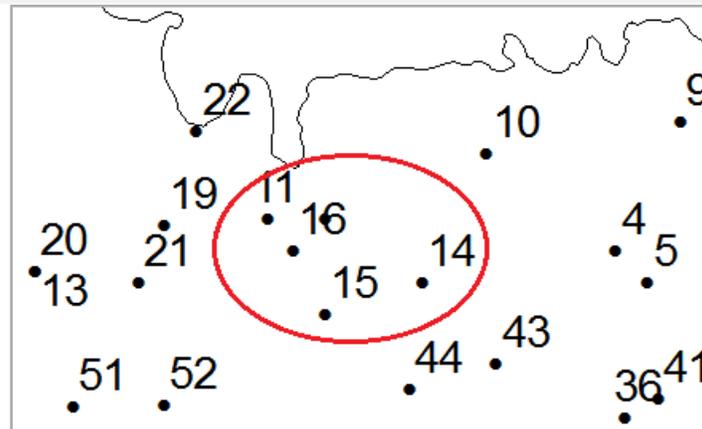


Figura 4 - Estações pluviométricas utilizadas no preenchimento de uma série de dados.

As Redes Neurais MultilayerPerceptron (perceptron multicamadas) trabalham com uma organização em camadas, como o nome sugere. Na camada de entrada, no caso do preenchimento de dados de precipitação, são apresentadas a série a ser corrigida e séries completas (ou com pouquíssimos dados faltantes) próximas a ela. A rede basicamente recebe como entrada o dia, mês, ano e a precipitação correspondente e nas camadas “escondidas” é onde a aprendizagem sobre os padrões de precipitação é realizada. Nestas camadas, também chamadas de intermediárias, a sub-rotina realiza cálculos, estipula pesos e a partir do aprendizado estima o valor faltante. A camada de saída consiste no resultado final do processamento, nela é apresentada a série com os dados faltantes preenchidos. O esquema abaixo (Figura 5) refere-se a esse processo.

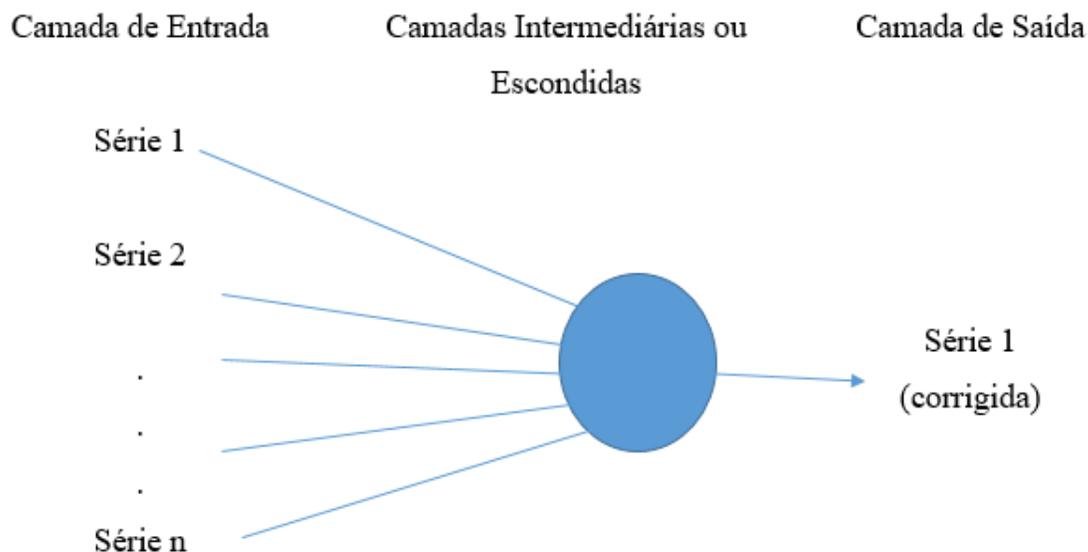


Figura 5 - Esquema do funcionamento de uma Redes Neural Artificial MultilayerPerceptron (perceptron multicamadas)



3. Resultados e discussões

Ao final da coleta, organização e triagem dos dados, 171 postos pluviométricos foram obtidos. Os dados foram preenchidos com relativa precisão. Dos 171, foram descartados 12 conjuntos de dados por conterem muitas falhas, que impossibilitavam a aprendizagem pelas Redes Neurais Artificiais, ou que mesmo com a utilização de procedimentos de preenchimento e consistência, prejudicariam os resultados finais, por não se apresentarem confiáveis. Na Figura 6, visualiza-se as séries de dados descartadas.

Tabela I – Postos pluviométricos descartados

Código	Nome da estação	Município	Latitude	Longitude	Altitude
2148168	Guarapiranga	Ribeirão Bonito	-21,96	-48,23	-
2245053	Bonfim	Aparecida	-22,95	-45,25	690
2248019	Agudos	Agudos	-22,48	-48,98	620
2344012	Bairro Paraibuna	Cunha	-23,18	-44,98	1120
2345001	Taubaté	Taubaté	-23	-45,56	610
2347015	Usina Santa Rosa (IAA)	Boituva	-23,23	47,62	520
2348076	Fazenda Boa Esperança	Capão Bonito	-23,95	-46,42	720
2348088	Engenheiro Bacelar	Itapeva	-23,87	-48,77	680
2447001	Usina Turvinho	São Miguel Arcanjo	-24	-47,95	660
2447008	Pedra do Largo	Miracatu	-25,5	-47,28	15
2447024	Cachoeira do Guilherme	Iguape	-24,48	-47,25	4
2448009	Pinara	Ribeirão Branco	24,26	-48,9	900

Identificou-se uma pequena modificação na base de dados utilizada neste trabalho. Pode-se perceber que as correções foram realizadas de forma esperada, visto que os dados obtidos foram comparados às estações de referência e os ajustes foram aceitáveis.

Uma das limitações do uso das redes neurais artificiais na correção dos dados de chuva é a falta de justificativa para os resultados apresentados além disso, a precisão deste modelo não é tão alta, pois os dados são muito variáveis em séries diárias de precipitação. A variabilidade dos dados diários de chuva dificulta a atribuição dos pesos, visto que estes são ajustados de acordo com os padrões apresentados. Se os padrões são muito variáveis, os pesos também serão, logo os resultados podem ficar mais imprecisos.

Wanderley et al. (2014) utilizaram as RNA's para interpolar dados de precipitação do estado de Alagoas (entre 1965 e 1980) e os dados estimados pelas redes neurais se mostraram muito parecidos com os das estações base, no entanto a partir da aplicação do teste t de Student, os autores obtiveram que os resultados variaram em função do procedimento utilizado e o mês de novembro foi o que apresentou melhores correções, visto que este mês tem menor variabilidade da precipitação no estado do Alagoas.



A aprendizagem, no presente estudo, foi realizada de modo aceitável, visto que dados de três a cinco séries foram utilizadas no treinamento e resultaram em apenas uma série. Entretanto, é imprescindível testar e analisar as correções dos dados para a melhor compreensão do funcionamento do método.

4. Conclusão

No estudo da precipitação pluvial é necessário dispor de séries longas e consistentes. Para tanto, é fundamental o tratamento dos dados disponíveis. Formas manuais de preenchimento costumam demandar um longo tempo para tamanha quantidade de estações pluviométricas existentes no estado de São Paulo e o descarte de muitas estações pode, até mesmo, inutilizar os resultados de cálculos que são feitos utilizando dados de chuva. No caso de interpolação desta variável, por exemplo, poucos postos pluviométricos podem prejudicar de forma significativa os cálculos geoestatísticos.

No presente trabalho, o preenchimento dos dados de chuva faltantes permitiu a obtenção de 159 séries temporais do estado de São Paulo, com 37 anos de dados (1976 a 2012) sem falhas.

Neste sentido, salienta-se que mesmo com a precisão do preenchimento das falhas não definida por completo, o método se faz importante para que as análises da precipitação pluvial possam ser realizadas cada vez mais de forma mais precisa, com bases de dados mais consistentes, possibilitando os mais diversos cálculos e estudos.

5. Referências

AGÊNCIA NACIONAL DE ÁGUAS – ANA. Portal Do Sistema Nacional De Informações Sobre Recursos Hídricos – SNIRH. Disponível em: portalsnirh.ana.gov.br. Acesso em: 10 de Mar. 2016.

AMARAL, M. R. S. do; CESARIO, C. V. Apostila do Minicurso: Software R. IX Semana de Estatística da UERJ. Solução Estatística Júnior, Rio de Janeiro, 2009. Disponível em: <<http://www.ime.uerj.br/~mrubens/slae/minicursosoftwareR.pdf>>. Acesso em: 15 de agosto de 2016.

GUIJARRO, J. A. CLIMATOL: Software libre para ladepuración y homogeneización de datos climatológicos. In: CONCHA D. L.; CODRÓN, J. C. G.; ALVAREZ, D. F. R.; HERNÁEZ, P. F. A.; PEDRAJA, C. G. (coord.) El clima, entre el mar y lamontaña. Santander: Universidad de Cantabria. p. 493–502, 2004.

HAYKIN, S. O. Neural networks and learning machines (3ª edição). Upper Saddle River, Nova Jersey: Prentice Hall, 2008.

NERY, J. T.; CARFAN, A. C. Re-analysis of pluvial precipitation in southern Brazil. *Atmósfera*, México, v. 27, n. 2, p. 103-115, 2014.

PIAZZA, A. et al. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geo-information*, v. 13, p. 396-408, 2011.

POVOA, L. V. prefann: Precipitation forecasting with Artificial Neural Network and others. Pacote do R. Versão 0.0.1. Disponível em: <<https://github.com/lucasvenez/pfann>>. Acesso em agosto de 2016.



TEEGAVARAPU, R. S. V.; CHANDRAMOULI, V. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, v. 312, p. 191–206, 2005.

WANDERLEY, H. S. et al. Interpolação espacial de dados médios mensais pluviométricos com redes neurais artificiais. *Revista Brasileira de Meteorologia*, 29(3), p. 389-396, 2014.