



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE TECNOLOGIA
CURSO DE GRADUAÇÃO EM ENGENHARIA AMBIENTAL

TRABALHO DE CONCLUSÃO DE CURSO

PREVISÃO DE SECA NA BACIA HIDROGRÁFICA DO AÇUDE
EPITÁCIO PESSOA/PB COM BASE EM REDES NEURAIS
ARTIFICIAIS

JAMILLE FREIRE AMORIM

João Pessoa - PB
Junho de 2017

JAMILLE FREIRE AMORIM

**PREVISÃO DE SECA NA BACIA HIDROGRÁFICA DO AÇUDE
EPITÁCIO PESSOA/PB COM BASE EM REDES NEURAIAS
ARTIFICIAIS**

Trabalho de conclusão de curso
apresentado como pré-requisito para a
obtenção do título de Bacharel em
Engenharia Ambiental pela
Universidade Federal da Paraíba.

Área de concentração: Recursos
Hídricos

Orientador: Prof. Dr. Francisco Jácome
Sarmiento

JOÃO PESSOA

2017

A524p Amorim, Jamille Freire

Previsão de Seca na Bacia Hidrográfica do Açude Epitácio Pessoa/PB com Base em Redes Neurais Artificiais./ Jamille Freire Amorim. – João Pessoa, 2017.

69f. il.:

Orientador: Prof. Francisco Jácome Sarmento.

Monografia (Curso de Graduação em Engenharia Ambiental)
Campus I - UFPB / Universidade Federal da Paraíba.

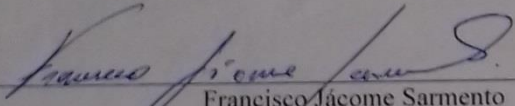
1. Índice de precipitação 2. Previsão de secas 3. Inteligência artificial I. Título.

FOLHA DE APROVAÇÃO

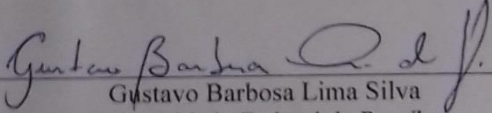
JAMILLE FREIRE AMORIM

**PREVISÃO DE SECA NA BACIA HIDROGRÁFICA DO AÇUDE EPITÁCIO
PESSOA/PB COM BASE EM REDES NEURAIS ARTIFICIAIS**

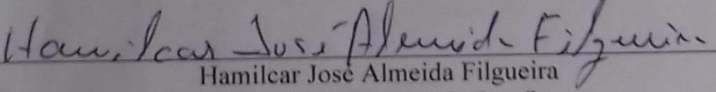
Trabalho de Conclusão de Curso apresentado em 30/05/2017 perante a seguinte Comissão Julgadora:


Francisco Jácome Sarmiento
Departamento de Engenharia Civil e Ambiental do Centro de
Tecnologia/UFPB

APROVADO


Gustavo Barbosa Lima Silva
Universidade Federal da Paraíba

APROVADO


Hamilcar José Almeida Filgueira
Universidade Federal da Paraíba

Aprovado

Prof. Adriano Rolim da Paz
Coordenador do Curso de Graduação em Engenharia Ambiental

AGRADECIMENTOS

Agradeço ao Universo, a Deus, ou seja lá como se queira chamar, por cada dia nesses anos de graduação. Foram anos de enorme evolução e aprendizado, seja através dos bons ou dos árduos, dolorosos e insanos momentos por mim vividos.

Agradeço aos meus amados pais, por todo o apoio e incentivo. Mesmo estando quilômetros longe de mim foram essenciais nesses anos.

Aos meus queridos irmãos Vinícius Amorim e Giselle Amorim pelo amor e carinho de sempre.

Ao Prof. Francisco Sarmiento por me orientar neste trabalho e por ser um exemplo de profissional brilhante, uma inspiração para todos aqueles que pretendem seguir os mesmos passos.

À Professora Carmem Gadelha, pelos anos em que estive ao meu lado me orientando nos projetos de pesquisa e me instruindo a se tornar melhor a cada dia.

Aos amigos que fiz no intercâmbio, que me mostraram o verdadeiro sentido de lealdade e companheirismo.

Aos amigos de curso Lairanne Costa e Cássio Lemos por estarem em todos os momentos ao meu lado.

À todos da Terra Consultoria.

RESUMO

As secas são eventos naturais associados principalmente a períodos de precipitação abaixo da média em uma dada região, que podem ocorrer em diversas partes do planeta e causar impactos na agricultura, na oferta de água, no turismo, no funcionamento e equilíbrio dos ecossistemas e consequentemente na provisão dos serviços ambientais. Entre o período de 1994 e 2013, esses eventos afetaram mais de um bilhão de pessoas no mundo. No Brasil as secas ocorrem com grande frequência no Nordeste brasileiro em virtude, dentre outros fatores, da elevada variabilidade espaço-temporal das precipitações. Tendo em vista a vulnerabilidade do semiárido brasileiro quanto ao fenômeno da seca, e a necessidade do conhecimento, monitoramento e previsão de tal fenômeno na bacia hidrográfica do Açude Epitácio Pessoa, localizada no município de Boqueirão - uma região semiárida do Estado da Paraíba, o presente trabalho foi proposto com o intuito de avaliar a aplicabilidade de redes neurais artificiais (RNAs) na identificação e previsão de secas utilizando o índice de precipitação padronizado (SPI). A princípio, os valores SPI 24, 12, 6, e 3 foram calculados para as séries históricas de três postos pluviométricos selecionados. Em seguida, foram criadas RNAs onde utilizaram-se diversas arquiteturas buscando a melhor descrição da relação entre as variáveis. O algoritmo *backpropagation* foi utilizado para ajustar os coeficientes envolvidos nas redes neurais como método alternativo ao cálculo do SPI – através da construção de RNAs sem defasagem entre os dados, e sua capacidade de previsão – através do uso de RNAs com defasagem dos dados. Os resultados obtidos mostraram uma boa capacidade de RNAs de 10 neurônios na camada intermediária para o cálculo do SPI em todos os postos pluviométricos estudados. No entanto, a construção das redes com defasagem de dados mostrou-se incapaz de prever secas.

Palavras-chaves: índice de precipitação, previsão de secas, inteligência artificial.

ABSTRACT

Droughts are natural events associated with below-average precipitation periods in a given region, which can occur in various parts of the planet and cause impacts on agriculture, water supply, tourism, ecosystems and on the provision of environmental services. Between 1994 and 2013, these events affected more than one billion people in the world. In Brazil, droughts occur frequently in the Brazilian Northeast region, due, among other factors, to the high spatial and temporal variability of rainfall. Considering the vulnerability of the Brazilian semi-arid region to these event, this work aims to evaluate the applicability of artificial neural networks (ANNs) in the identification and prediction of droughts using the standardized precipitation index (SPI) in the Epitacio Pessoa watershed, located in the municipality of Boqueirão - a semi-arid region of the State of Paraíba – Brazil. At first, the SPI values 24, 12, 6, and 3 were calculated for the historical series of three selected pluviometric stations. After that, RNAs were created where several architectures were applied to evaluate those that better described the relationship between the variables. The backpropagation algorithm was used to adjust the coefficients involved in the ANNS as an alternative method to the calculation of the SPI - through the construction of ANNs without lag between the data, and its predictive capacity - through the use of RNAs with data lag. The results showed a good capacity of RNAs of 10 neurons in the intermediate layer for the calculation of SPI in all stations studied. However, the construction of networks with data lags proved unable to predict droughts.

Palavras-chaves: Precipitation Index, prediction of droughts, artificial intelligence.

LISTA DE FIGURAS

Figura 1 – Modelo de neurônio artificial.....	11
Figura 2 - Componentes de um neurônio artificial.....	12
Figura 3 - Rede <i>Feedforward</i> de duas camadas ocultas (<i>Multilayer Perceptron</i> - MLP)	13
Figura 4 - Fluxograma da metodologia aplicada no presente trabalho.....	15
Figura 5 – Mapa das bacias e sub-bacias hidrográficas do estado da Paraíba.....	16
Figura 6 – Localização geográfica dos postos pluviométricos estudados	18
Figura 7 - Função de ativação sigmóide	21
Figura 8 – Fluxo de processamento do algoritmo backpropagation.....	22
Figura 9- Valores SPI para o posto pluviométrico Barra de São Miguel, PB	28
Figura 10 - Valores SPI para o posto pluviométrico Passagem, PB.....	29
Figura 11 - Valores SPI para o posto pluviométrico Salgadinho, PB	30
Figura 12 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 24 sem defasagem temporal de dados	37
Figura 13 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 12 sem defasagem temporal de dados	38
Figura 14 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 – semestre 1 sem defasagem temporal de dados	38
Figura 15 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 – semestre 2 sem defasagem temporal de dados	39
Figura 16 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 24 com defasagem temporal de dados	41
Figura 17 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 12 com defasagem temporal de dados	41
Figura 18 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 com defasagem temporal de dados	42

LISTA DE TABELAS

Tabela 1 – Classificação da precipitação baseado no valor SPI.....	9
Tabela 2 - Postos pluviométricos selecionados	18
Tabela 3 - Resultados estatísticos das redes neurais artificiais (RNAs) sem defasagem temporal de dados para os Índices de Precipitação Padronizado (SPI).....	36
Tabela 4 - Resultados estatísticos SPI 24 sem defasagem temporal de dados	37
Tabela 5 - Resultados estatísticos SPI 12 sem defasagem temporal de dados	38
Tabela 6 - Resultados estatísticos SPI 6 – Semestre 1 sem defasagem temporal de dados.....	38
Tabela 7 - Resultados estatísticos SPI 6 – Semestre 2 sem defasagem temporal de dados.....	39
Tabela 8 - Resultados estatísticos das redes neurais artificiais (RNAs) com defasagem temporal de dados para os índices de precipitação padronizado (SPI).....	40
Tabela 9 - Resultados estatísticos SPI 24 com defasagem temporal de dados.....	41
Tabela 10 - Resultados estatísticos SPI 12 com defasagem temporal de dados.....	41
Tabela 11 - Resultados estatísticos SPI 6 com defasagem temporal de dados.....	42
Tabela A.1 - Pesos e vieses iniciais para RNAs 5	49
Tabela A.2 - Pesos e vieses iniciais para RNAs 10	49
Tabela A.3 - Pesos e vieses iniciais para RNAs 15	50
Tabela A.4 - Pesos e vieses iniciais para RNAs 20	50

LISTA DE ABREVIATURAS

ANA	Agência Nacional das Águas
ARIMA	<i>Regressive Integrated Moving Average</i>
CEC	<i>Comissão Europeia</i>
EEA	Agência Ambiental Europeia
ENOS	El Niño-Oscilação Sul
EQM	Erro Quadrático Médio
FF	<i>Feedforward</i>
FFBP	<i>feed forward backpropagation</i>
INMET	Instituto Nacional de Meteorologia
IPMA	Instituto Português do Mar e da Atmosfera
MLP	<i>Multilayer Perceptron</i>
NDVI	<i>Normalized Difference Indices Vegetation .</i>
NIDIS	Sistema Integrado de Informações sobre a seca
NOOA	Centro de Administração Nacional dos Oceanos e da Atmosfera dos Estados Unidos da América
RNA	Rede Neural Artificial
SPEI	<i>Standardized Precipitation-Evaporation Index</i>
SPI	Índice de Precipitação Padronizado
VCI	Índice de Condição de Vegetação
WMO	<i>World Meteorological Organization</i>

SUMÁRIO

1. INTRODUÇÃO	1
1.1 Contextualização do problema.....	1
1.2 Justificativa da pesquisa.....	3
1.3 Objetivos.....	5
1.3.1 Objetivo geral	5
1.3.2 Objetivos específicos	5
2. REVISÃO BIBLIOGRÁFICA.....	6
2.1 Secas	6
2.2 Índices de secas.....	7
2.2.1 Índice de precipitação padronizado (SPI).....	8
2.3 Previsão de secas	10
2.3.1 Redes neurais artificiais.....	11
2.3.2 Redes neurais do tipo <i>feedforward</i>	13
3. METODOLOGIA	14
3.1 Área de estudo	16
3.2 Dados	17
3.3 Cálculo do SPI	18
3.4 Concepção da rede neural	21
3.4.1 Treinamento da RNA.....	25
4. RESULTADOS E DISCUSSÃO	27
4.1 Valores SPI	27
4.2 Redes neurais	36
5. CONCLUSÃO E RECOMENDAÇÕES	43
REFERÊNCIAS BIBLIOGRÁFICAS	44
APÊNDICE A – Pesos e vieses iniciais	49
APÊNDICE B – Código de construção da rede neural artificial no software R	51

1. INTRODUÇÃO

1.1 Contextualização do problema

As secas são eventos naturais associados principalmente a períodos de precipitação abaixo da média em uma dada região, que podem ocorrer em diversas partes do planeta e causar impactos na agricultura, na oferta de água, no turismo, no funcionamento e equilíbrio dos ecossistemas e consequentemente na provisão dos serviços ambientais. Além disso, fatores humanos como a demanda por água e o manejo inadequado dos recursos hídricos podem intensificar os efeitos negativos de tal evento em uma dada região (MARATHE e DEMUTH, 2013).

Apesar de ser um evento temporário na variação das condições climáticas normais, os impactos adversos para a sociedade e o meio ambiente, de maneira geral, não são imediatamente observáveis, visto que as secas se comportam de forma gradual (ARCS, 2007). Como descrito por Batista (2012), isso se explica porque as secas iniciam-se normalmente de forma lenta, possuem longa duração e espalham-se por uma extensa área. Sendo assim, poucos fenômenos naturais são tão devastadores quanto as secas (CARRÃO, NAUMANN; BARBOSA, 2016).

Entre o período de 1994 e 2013, as secas afetaram mais de um bilhão de pessoas no mundo, dentre os quais a África foi o continente onde se registou o maior número durante esse período de tempo, cerca de 40% do total registrado no planeta (CRED, 2015).

Países desenvolvidos apesar de menos vulneráveis, devido à presença de recursos e políticas públicas eficazes, também sofrem com as secas. Nos Estados Unidos da América, estados como a Califórnia e Nevada são afetados frequentemente. Calcula-se que esses eventos causem uma perda média de US \$6 - 8 bilhões por ano. Estudos reportam ainda que entre os anos de 1980 e 2014 as secas resultaram em custos de mais de US \$ 200 bilhões para o país (NCDC, 2015).

Na Europa, a Agência Ambiental Europeia - EEA (2010), reportou que a média anual dos impactos econômicos desses eventos dobrou entre 1976 - 1990 e 1991 - 2006, subindo para €6,2 bilhões por ano no período mais recente. Além disso, entre 1976 e 2006, as áreas afetadas aumentaram em cerca de 20%. Estimativas da Comissão Europeia (CEC, 2007) indicam, ainda, que os danos das secas no continente nos últimos 30 anos foram de pelo menos €100 bilhões.

No Brasil, a região semiárida da região Nordeste é a que mais reporta ocorrências de seca no país, sofrendo recorrentemente com os efeitos desse evento. O semiárido nordestino, que abrange uma área de 969.589,4 km², compreende 1.133 municípios de nove estados do Brasil, Piauí, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia e Minas Gerais, onde vivem cerca de 22 milhões de pessoas que representam 11,8 % da população brasileira (IBGE, 2010).

Nessa região, o relato de períodos de secas remonta ao século XVI, sendo uma constante na literatura a abordagem sobre esse fenômeno e a convivência de grande parte do povo nordestino com o mesmo (PONTES; MACHADO, 2008).

Nos últimos 500 anos, a região Nordeste do Brasil enfrentou vários períodos de secas. De acordo com Villa (2000) entre os séculos XIX e XX em torno de três milhões de pessoas morreram na região devido a esse fenômeno, relacionando que a quantidade de vítimas equivale a duas guerras do Vietnã (MARTINEZ, 2002). Villa (2000) ainda relata: “a seca de 1877-1879, foi uma das mais terríveis, teria dizimado cerca de 4% da população nordestina, erigindo o Nordeste, desde então, em “região-problema”. (VILLA, 2000, p.253).

Comparado a outras regiões semiáridas do mundo, o semiárido nordestino possui um volume anual de precipitação pluviométrica relativamente elevado, variando entre 400 mm a 800 mm. No entanto, as temperaturas médias anuais elevadas e a alta variabilidade espaço temporal da precipitação, combinado ao solo cristalino, em sua maior parte de difícil infiltração, com alto teor de cloreto de sódio, aumenta a vulnerabilidade da região quando da ocorrência de secas (NÓBREGA; SANTIAGO; SOARES, 2016). Além disso, o índice de evapotranspiração (3 mil mm/ano) é maior que as chuvas anuais, o que dificulta o armazenamento da água e torna a região ainda mais vulnerável à esses eventos extremos (ASA, 2014)

Além das vulnerabilidades ambientais do semiárido, Silva et al. (2013) evidencia ainda as interações do desastres das secas nas consequências negativas sobre o processo de desenvolvimento social da região. Os efeitos das secas sobre a economia regional e os grandes prejuízos promovem, por exemplo, o fenômeno das migrações, orientadas, ao longo dos anos, para quase todo o Brasil.

Dessa forma, considerando os efeitos econômicos, ambientais e sociais das secas, é de fundamental importância que haja o monitoramento e previsão de tais eventos. O monitoramento permite que observar como a seca está progredindo e a sua previsão desempenha um papel importante na mitigação dos seus impactos, principalmente, sobre os sistemas de recursos hídricos. Sob essa perspectiva, e considerando a vulnerabilidade do

semiárido brasileiro quanto ao fenômeno da seca, esta pesquisa questiona de que maneira o monitoramento e predição do fenômeno da seca na bacia hidrográfica do açude Epitácio Pessoa, estado da Paraíba, auxilia na tomada de decisão e contribui para a mitigação desse evento na região.

1.2 Justificativa da pesquisa

A fim de reduzir as ameaças globais das secas, vários países têm investido em pesquisas capazes de detectar, quantificar e prever tais eventos. Centros de estudo e monitoramento de secas existentes no mundo desenvolvem pesquisas, elaboram planos e promovem a colaboração entre pesquisadores e o público.

Um número crescente de iniciativas internacionais vem tentando influenciar governos a desenvolver políticas e fornecer recursos para que evolua uma sociedade resiliente às secas (CARRÃO; NAUMANN; BARBOSA, 2016). Projetos e relatórios como “*High-level Meeting on National Drought Policy*” (WMO, 2013) e “*The Human Cost of Natural Disasters – 2015*” (CRED, 2015), diagnosticam e apontam possíveis soluções para esse problema.

Os Estados Unidos da América, por exemplo, lançaram em 2006 o Sistema Integrado de Informações sobre a seca (NIDIS), a fim de reunir informações sobre o fenômeno no país. O programa promove a informação dinâmica e de fácil acesso para a nação, utilizando infraestrutura e informação disponível entre parceiros federais, estaduais e locais. Dessa forma, o público tem acesso às pesquisas e informações desenvolvidas por uma série de colaboradores como o Departamento de Agricultura dos Estados Unidos, o Centro de Mitigação Nacional das Secas, grupos de pesquisas de universidades, entre outros.

Na Europa, um relatório da União Europeia revelou que em 2009 mais de 100.000 estações de monitoramento estavam em operação (MARATHE; DEMUTH, 2013). Além disso, uma plataforma online (Observatório de Secas da Europa – EDO), fruto de pesquisas conjuntas de universidade e centros de estudos de países da União Europeia, fornece ao público informações relevantes de monitoramento sobre as secas. Mapas e outras ferramentas possibilitam que o público possa entender a dinâmica do fenômeno na região e que tomadores de decisão possam se planejar.

No Brasil, a Agência Nacional das Águas (ANA) lançou em 2014 o Monitor de Secas, ferramenta de acompanhamento (monitoramento) regular e periódico da situação da seca no Nordeste. O Monitor de Secas acompanha, mês a mês, como a seca está progredindo em todos os estados do Nordeste, e produz um “mapa da seca”. Originado da associação da ANA com

outros órgãos federais e estaduais, o Monitor de Secas objetiva integrar o conhecimento técnico e científico já existente em diferentes instituições estaduais e federais para alcançar um entendimento comum sobre as condições do fenômeno, como: sua severidade, a evolução espacial e temporal, e seus impactos sobre os diferentes setores envolvidos (ANA, 2016).

Além do trabalho de organizações e órgãos governamentais, pesquisadores em todo o mundo que objetivam avaliar os aspectos ambientais, econômicos e sociais relacionados às secas discutem e desenvolvem técnicas para auxiliar na gestão desse fenômeno. Estudiosos investigam as secas com o intuito de desenvolver e aprimorar métodos de mapeamento, observação, monitoramento, análise e previsão.

No Brasil, vários trabalhos, especialmente na região semiárida nordestina, já foram publicados sobre o tema. Moraes et al. (2012), por exemplo, estudaram os eventos de seca na região Nordeste e avaliou um método computacional de previsibilidade. Silva et al (2013), discutiu sobre as questões contemporâneas do desastre no Nordeste Brasileiro. Pontes (2014) realizou um estudo de caso comparativo onde analisou a segurança hídrica de famílias rurais no município de Afogados da Ingazeira, em Pernambuco (Brasil) – visando avaliar como o acesso irrestrito e descentralizado à água nas comunidades rurais pode impulsionar o desenvolvimento local, com base na convivência com o semiárido. Ferreira et al. (2015) calcularam e analisaram o Índice de Condição de Vegetação (VCI) para o interior do estado de Alagoas, tendo em vista uma análise da seca ocorrida em 2012 em comparação com os últimos 11 anos, utilizando imagens orbitais da Normalized Difference Indices Vegetation (NDVI). Silva e Pacheco (2016) investigaram o processo de desertificação como consequência da seca no município de Irauçuba/CE, cidade localizada no semiárido nordestino.

Dessa forma, levando em consideração o exposto e a necessidade do conhecimento, monitoramento e previsão do fenômeno da seca e considerando a necessidade de aprofundamento nas informações de tal evento na bacia hidrográfica do açude Epitácio Pessoa, é que essa pesquisa se justifica.

1.3 Objetivos

1.3.1 Objetivo geral

Avaliar a aplicabilidade de redes neurais artificiais (RNAs) na identificação e previsão de secas utilizando o índice de precipitação padronizado (SPI), em diferentes escalas temporais, no âmbito da bacia hidrográfica do açude Epitácio Pessoa.

1.3.2 Objetivos específicos

1. Estudar os eventos de seca por meio do Índice de Precipitação Padronizado (SPI);
2. Investigar a eficiência do cálculo do SPI com o uso de inteligência artificial;
3. Avaliar a predição do evento da seca por meio da aplicação de inteligência artificial.

2. REVISÃO BIBLIOGRÁFICA

2.1 Secas

A seca é um evento que pode ser observado em diversas partes do mundo, sendo definida na literatura como um longo período de escassez de eventos de precipitação capaz de desencadear inúmeros problemas para os diversos segmentos da sociedade, no qual seus impactos resultam da relação entre o evento natural e a demanda humana por água.

A fim de entender as causas das incidências de tal evento, pesquisadores vêm investigando a relação de outros eventos naturais com as ocorrências de seca, tais como a variabilidade natural dos oceanos e as mudanças atmosféricas. Pesquisas em todo o mundo vêm sendo desenvolvidas, por exemplo, com o intuito de entender os eventos de secas no Oriente Médio e na África Oriental, Hoell et al. (2013), com o intuito de entender os eventos de secas no Oriente Médio e na África Oriental, mostraram em seu estudo que as variações de umidade e advecção de ar seco, devido às circulações anômalas em eventos do La Niña, intensificaram os eventos de seca nas regiões estudadas. Em outro estudo, Funk et. al (2014) mostraram que a variabilidade da precipitação e consequentes secas na África Oriental estão ligadas às flutuações de temperatura superficial do Pacífico Ocidental.

No Nordeste brasileiro, estudos mostraram a importância do Atlântico Sul na variabilidade de precipitação da região estudada (GIANNINI et al., 2004; ANDREOLI et al., 2005). Consistente com tais autores, Nóbrega e Santiago (2014) identificaram relações entre as secas no estado de Pernambuco e a ação conjunta dos eventos do El Niño-Oscilação Sul (ENOS).

É importante destacar que o entendimento dos eventos de seca é um processo complexo, pois envolve diversos fatores sejam eles climáticos globais, como exemplificados acima, sejam eles hidrológicos, geológicos e meteorológicos (ROSA, 2011). Além disso, fatores sociais e ações humanas também apresentam um papel fundamental na intensidade e nos impactos negativos quando da ocorrência de tais eventos.

A complexidade das secas verifica-se, portanto, não somente em relação aos fatores acima citados, mas também ao nível de definição. Palmer (1965) define seca como o intervalo de tempo, geralmente da ordem de meses ou até mesmo de anos, durante o qual a precipitação diminui consideravelmente em relação ao valor climatologicamente esperado ou apropriado. Para Wilhite e Glantz (1985), as secas estão relacionadas às condições abaixo da relação de balanço ‘média’ entre precipitação e evapotranspiração em uma área em particular.

Vale salientar, no entanto, que a definição de secas pode variar de acordo com as formas que pode ser sentida pelos diferentes setores da sociedade. Assim, um fator que influencia nessa grande heterogeneidade de formas de incidência é o fato de que a mesma quantidade de precipitação pode ter consequências diferentes dependendo das características de cada região, como os elementos meteorológicos, o tipo de ecossistema e das atividades econômicas desenvolvidas (CANAMARY, 2015). Todavia, de maneira geral, a maior parte dos trabalhos científicos reconhecem quatro tipos de seca: meteorológica, hidrológica, agrícola e socioeconômica (BLAIN, 2005).

A Organização Meteorológica Mundial (*World Meteorological Organization* – WMO, 2006) agrupa as definições ou tipos de secas nas quatro categorias citadas. A seca meteorológica é definida como um período onde ocorre uma deficiência de precipitação em relação aos valores normais de uma região. Tal tipo de seca ocorre com maior frequência que as demais, em virtude do tempo necessário para que os impactos gerados pelo déficit de chuva sejam percebidos no meio (WMO, 2006).

Ainda de acordo com a WMO (2006), a seca agrícola ocorre caso essa carência se mantenha por um longo período, sendo verificada uma redução da umidade do solo até um ponto que comece a prejudicar o desenvolvimento dos cultivos, ocasionando o estresse hídrico. Já a seca hidrológica é desencadeada por uma diminuição da infiltração e do escoamento e, conseqüentemente, da recarga do lençol freático e das reservas hídricas superficiais.

A seca intitulada socioeconômica é evidenciada quando a deficiência no suprimento de água começa a afetar a produção de bens ofertados para a população. Ela é, portanto, associada à questão monetária e dependente dos tipos de uso da água, da densidade de usuários e da sua distribuição (WMO, 2006; HEIM JUNIOR, 2002).

Assim, baseado na premissa de que todos os tipos de seca são originados por um déficit de precipitação que resulta em uma baixa disponibilidade hídrica para a atividade que a requer, McKee et al. (1993) afirmam ser o tempo requerido, de ocorrência de tal deficiência, que separa os diferentes conceitos desse fenômeno climático.

2.2 Índices de secas

A elevada complexidade de avaliação da seca, por ser um fenômeno capaz de se manifestar de diferentes formas e gerar uma grande gama de impactos, estimulou a realização de pesquisas, visando o desenvolvimento de instrumentos que procuram caracterizar esse

evento por meio de um único valor. A partir de tal premissa, pesquisadores desenvolveram os índices de seca, dos quais é possível que se investigue e compreenda de forma mais eficaz e prática a seca e seus impactos, contribuindo para sua monitoração, e até certo grau, predição (ROSA, 2011).

Em suma, um índice é uma variável que avalia o efeito e define diversos parâmetros de seca, que incluem duração, intensidade e gravidade. Assim, um índice deve ser capaz de quantificar a seca por diferentes escalas de tempo, para o qual uma longa série de monitoramento é essencial (DANTAS, 2015). Dessa forma, tal ferramenta é capaz de extrair informações de eventos de seca como a severidade, a extensão geográfica e a duração, auxiliando no processo de tomada de decisão (CANAMARY, 2015).

Existe uma grande variabilidade de índices de seca disponíveis na literatura. Assim a escolha da melhor ferramenta para cada tipo de estudo deve considerar fatores como as características hidroclimáticas da região analisada, o tipo de seca, a proposta do trabalho, os dados disponíveis e a vulnerabilidade da sociedade. A análise dos índices calculados a partir dos dados históricos permite a identificação dos locais mais afetados pelos déficits de água (CANAMARY, 2015).

2.2.1 Índice de precipitação padronizado (SPI)

O Índice de precipitação padronizado - *Standardized Precipitation Index* (SPI) é um índice que foi desenvolvido por McKee et al. (1993), com o objetivo de identificar os períodos de seca e avaliar a sua severidade, em várias escalas temporais.

O índice se baseia na distribuição de probabilidades da precipitação e traduz o desvio das condições ‘normais’, ou seja, suas anomalias. Assim, o SPI expressa o afastamento entre a precipitação observada e a média, em unidades de desvio padrão, no mês considerado, ou seja, é um índice normalizado, espacial e temporal (PORTELA et al., 2014). Esta característica permite efetuar, portanto, comparações entre valores de SPI obtidos em diferentes locais ou em períodos diferentes, mesmo que tais regiões apresentem climas bastante distintos (SANTOS; PORTELA, 2010).

As escalas temporais mais usualmente analisadas são as de 1 mês, 3 meses, 6 meses, 12 meses e 24 meses. A escolha da escala temporal usada no cálculo do SPI relaciona-se com a sequência temporal das repercussões da seca sobre os recursos hídricos (PAULO, 2004). Dessa forma, como descrito por Santos e Portela (2010), escalas menores de até 6 meses, permitem avaliar a intensidade desse evento com o intuito de analisar seus efeitos sobre a seca

meteorológica e agrícola. Já as escalas de tempo mais longas (12 ou 24 meses) são mais adequadas para avaliar as anomalias nas reservas hídricas superficiais e as consequentes perturbações na sua oferta.

De acordo com a metodologia do SPI, o evento seca ocorre no período em que o valor do SPI é continuamente negativo. Quando o SPI se torna positivo tal evento cessa. Assim, a análise de uma sequência histórica de valores SPI permite avaliar a frequência de eventos anômalos de secas e chuvas extremas em uma região (Tabela 1).

Tabela 1 – Classificação da precipitação baseado no valor SPI

Valores de SPI	Classe
2,00 ou mais	Extremamente Úmido
1,50 a 1,99	Severamente Úmido
1,00 a 1,49	Moderadamente Úmido
0,10 a 0,99	Umidamente Incipiente
0 a -0,99	Seca Incipiente
-1,00 a -1,49	Moderadamente Seco
-1,50 a -1,99	Severamente Seco
-2,00 ou menos	Extremamente Seco

Fonte: McKee (1993).

Vale salientar que o SPI é baseado apenas na precipitação. Assim, a avaliação de seca não exige o uso de outros dados hidrometeorológicos e não é afetado negativamente pela topografia (HAYES et al. 1999). Devido às vantagens apresentadas, tal índice tem sido amplamente usado no estudo da variabilidade temporal e espacial das secas, e é recomendado pela Organização Mundial de Meteorologia desde 2009, como o índice padrão para a monitoração desses eventos (WMO, 2012).

A utilização desse índice se encontra em expansão por todo o mundo. Centros de pesquisa governamentais de meteorologia fornecem dados atualizados de monitoração das secas por meio do índice SPI, é o caso do Instituto Português do Mar e da Atmosfera (IPMA) e do Centro de Administração Nacional dos Oceanos e da Atmosfera dos Estados Unidos (NOOA). Além disso, pesquisadores em todo o mundo utilizam esse índice para a investigação de eventos extremos. Dogan (2014), por exemplo, estudou os eventos extremos de seca em Sidney e Melbourne na Austrália desde 1850, por meio do SPI. Dash et al. (2012) estudou as características das secas meteorológicas em Bangladesh utilizando o SPI.

No Brasil, diversos autores vêm utilizando o índice SPI para analisar os eventos de seca no país. Além disso, no Instituto Nacional de Meteorologia (INMET) o SPI é um dos métodos preconizados para caracterizar anomalias de precipitação. Desde janeiro de 2002 o

SPI vem sendo calculado para todo país e, recentemente, esses resultados têm sido disponibilizados para o público em forma de mapas (MORAIS et al., 2012).

2.3 Previsão de secas

A possibilidade de previsão de eventos de seca permite que as autoridades competentes possam colocar em prática planos de ação, antes que o problema esteja instalado, reduzindo possíveis impactos.

Tradicionalmente, modelos estocásticos foram usados para prever secas, sendo esses baseados em métodos que lançam mão de séries temporais disponíveis. No entanto, esses modelos são basicamente lineares e tem uma capacidade limitada de capturar a não-linearidade inerente a dados hidrológicos (MORAIS; SILVA; SANTOS, 2012).

Em décadas recentes, as Redes Neurais Artificiais (RNAs) têm sido usadas e demonstrado uma grande capacidade na modelagem de séries temporais hidrológicas, despontando como uma ferramenta útil na previsão de fenômenos naturais.

Diversos estudos vêm sendo realizados nesse âmbito. Um dos aspectos que tem motivado o uso frequente de RNAs em diversos campos da ciência, particularmente em estudos de previsão de séries temporais, é a sua capacidade comprovada de representar adequadamente relações fortemente não-lineares entre variáveis (FREITAS, 2000).

Valipour, et al. (2013) previu a vazão de entrada da barragem localizada na bacia de Dez utilizando modelos de redes neurais: *Auto Regressive Moving Average* e *Auto Regressive Integrated Moving Average* (ARIMA), onde concluiu-se que redes neurais artificiais autoregressivas estáticas e dinâmicas com funções de ativação sigmoide são capazes de prever a entrada afluente de barragens a partir de dados dos últimos 60 meses.

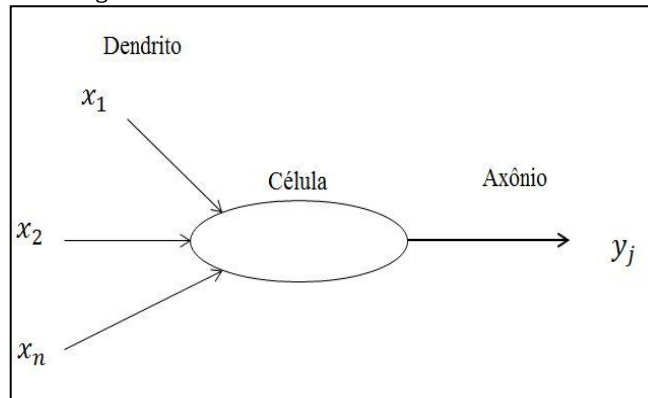
Partal et al. (2014) para prever a ocorrência de precipitação diária, usaram três algoritmos de redes neurais diferentes: *feed forward backpropagation* (FFBP); *radial basis function*; *generalized regression neural network* e *wavelet transformation*. Os resultados indicaram que o uso de *wavelet transformation* com redes neurais podem prover significativas vantagens para o processo de estimação de precipitações.

Dayal; Deo e Apan (2017) investigaram o uso de algoritmos de RNA para a predição do índice de seca: *Standardized Precipitation-Evaporation Index* (SPEI), em duas regiões de Queensland, Austrália. Através da avaliação do estudo e de análises estatísticas concluiu-se que modelos de redes neurais podem ser aplicados como ferramentas para a predição de eventos de secas.

2.3.1 Redes neurais artificiais

As Redes Neurais Artificiais funcionam baseadas na capacidade biológica de informação e processamento do cérebro humano, a partir das conexões entre neurônios. Logo, o modelo dos seus componentes apresenta uma analogia direta aos componentes de um neurônio biológico (Figura 1).

Figura 1 – Modelo de neurônio artificial



Fonte: adaptado de Tayfur (2012).

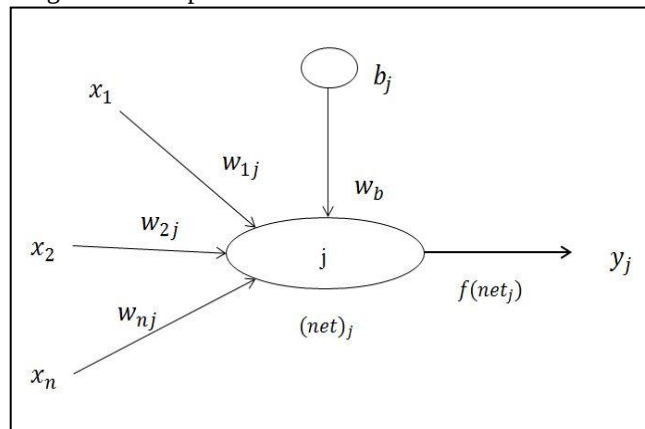
As RNAs são, portanto, conjuntos bem estruturados de unidades de processamento dispostas em camadas interconectadas entre si, que tentam reconhecer regularidades e padrões de entrada para aprender com a experiência e realizar generalizações. Podem, portanto, representar qualquer relação funcional complexa, por meio de funções não lineares entre conjunto de dados (FERNANDES et al. 1995).

Existem quatro componentes e duas operações básicas quando criado um modelo funcional de RNA:

- Vetor de entrada (x_i)
- Pesos (w_{ij})
- Viés (bias b_j)
- Saída (y_j)
- Informação de rede ($\Sigma, \Pi, \min, \max$)
- Função de ativação (f)

Basicamente uma RNA recebe uma entrada, processa-a, e produz uma resposta (Figura 2).

Figura 2 - Componentes de um neurônio artificial



Fonte: adaptado de Tayfur (2012).

Para o funcionamento de uma RNA, os dados de entrada fornecidos para a rede devem:

- (1) ser multiplicados pelos pesos das conexões entre o neurônio e o seu adjacente, simulando sua entrada em uma célula neural;
- (2) A seguir é feita a computação das entradas no neurônio adjacente. Vale salientar que nessa etapa o método mais utilizado em aplicações de engenharia de recursos hídricos é a soma ponderada (TAYFUR, 2012);
- (3) O valor computado é então processado por uma função de ativação, que determina a resposta de um neurônio ao sinal de entrada total que recebe;
- (4) Essa função emite uma resposta ao estímulo recebido para os demais neurônios adjacentes.

Além dos processos descritos, existe também um elemento externo b_j , chamado de viés ou “bias” que é empregado na rede com o objetivo de evitar ruídos aleatórios que possa vir a afetar o sistema.

É importante destacar que há inúmeros modelos de redes neurais artificiais que se diferenciam, principalmente, pela sua arquitetura e neurodinâmica. Assim, a aplicação de RNAs passa basicamente por três fases: definição da arquitetura da rede, aprendizagem, e verificação da rede (BAFFA et al., 2008).

A arquitetura é definida basicamente pela determinação do número de camadas, número de neurônios por camada, tipo de conexão entre as camadas (funções de ativação) e topologia da rede. Já o processo de aprendizagem consiste em fornecer a rede um conjunto de exemplos e alterar os seus pesos de modo a tornar a rede capaz de representar a relação entre os dados de entrada e de saída.

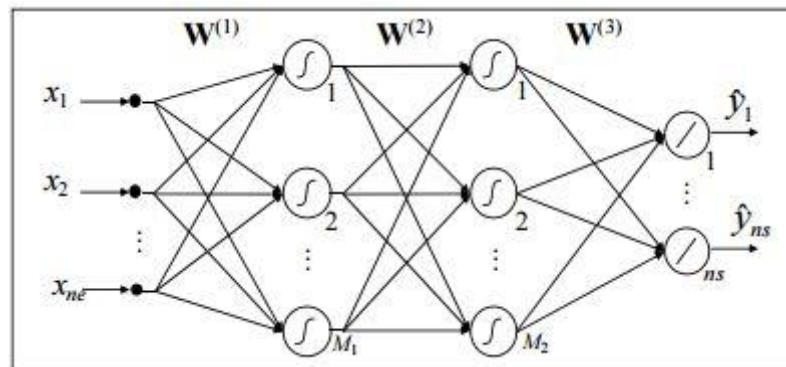
2.3.2 Redes neurais do tipo *feedforward*

O tipo de rede mais popular para a previsão de séries temporais consiste na rede progressiva *Feedforward* (FF) com três camadas (MORAIS et al., 2012). Uma das razões para isso é a sua capacidade de aproximação universal e sua flexibilidade para solucionar uma grande classe de problemas, incluindo reconhecimento de padrões, processamento de sinais, controle e otimização, problemas de classificação e previsão de séries temporais (TAYFUR, 2012).

Em uma rede neural do tipo FF, os neurônios são arranjados em camadas, no qual neurônios de uma camada são conectados aos da próxima. A informação flui na direção adiante - da camada de entrada, para a camada intermediária, até a camada de saída.

Dentre os principais tipos de redes FF, destaca-se o *Multilayer Perceptron* (MLP) ilustrado na Figura 3.

Figura 3 - Rede *Feedforward* de duas camadas ocultas (*Multilayer Perceptron* - MLP)



Fonte: Villanueva (2011).

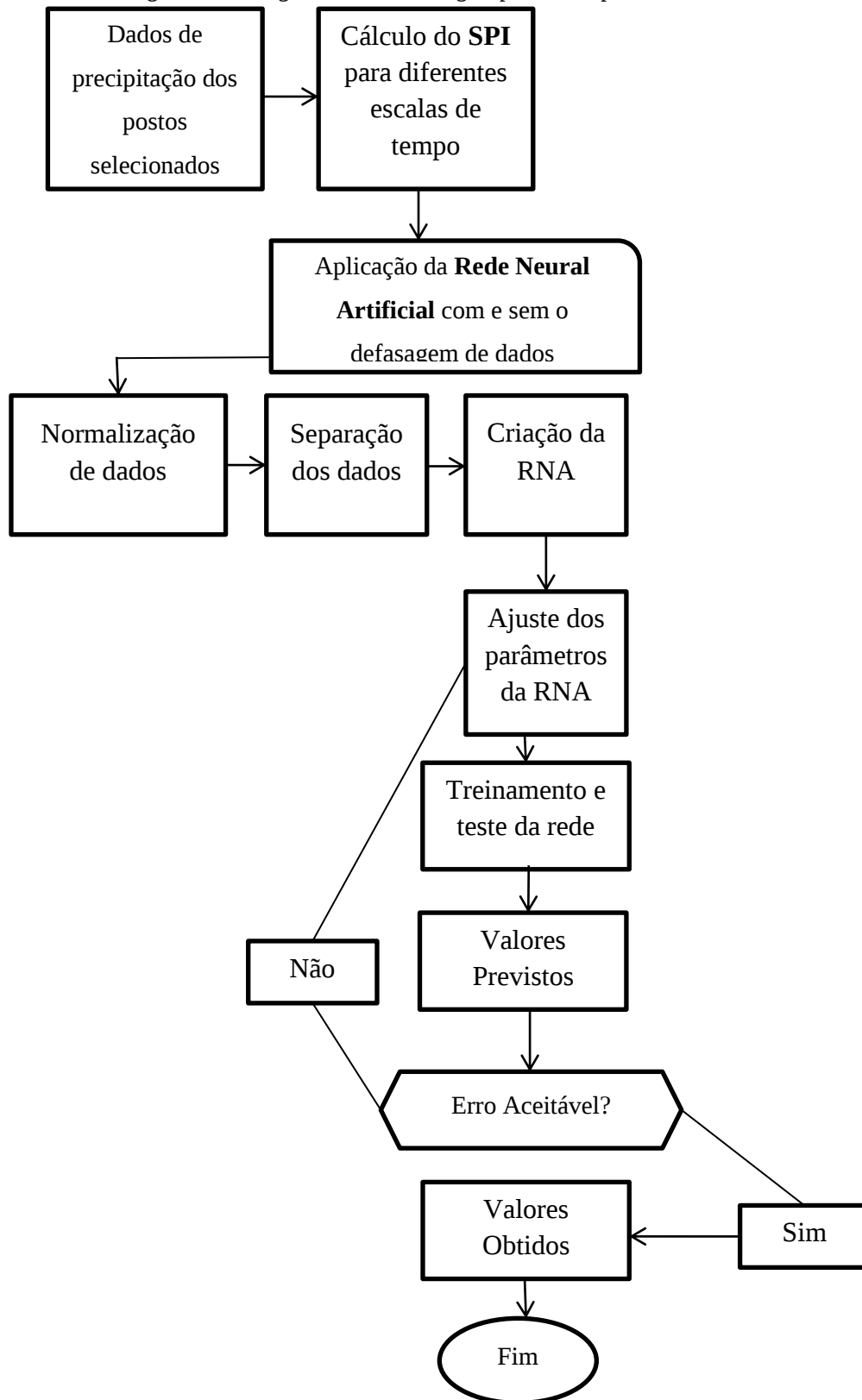
Redes MLP consistem de uma camada de entrada que contém variáveis independentes e são alimentados diretamente, uma única camada de saída, e camadas intermediárias, que podem ser chamadas de camadas escondidas. Cada camada possui neurônios e cada neurônio está totalmente interligado por pesos com todos os neurônios da camada subsequente (HEDDAM, 2016).

3. METODOLOGIA

Para o desenvolvimento do presente estudo foram considerados valores de precipitação média mensal de três postos pluviométricos obtidos junto à Agência Nacional das Águas (ANA, 2017). Para o cálculo do SPI, foram utilizadas como suporte computacional planilhas eletrônicas do Excel. Além disso, a confecção da rede neural artificial foi feita com o auxílio de planilhas eletrônicas do Excel e do software R.

A Figura 4 ilustra o fluxograma da metodologia aplicada nestetrabalho.

Figura 4 - Fluxograma da metodologia aplicada no presente trabalho

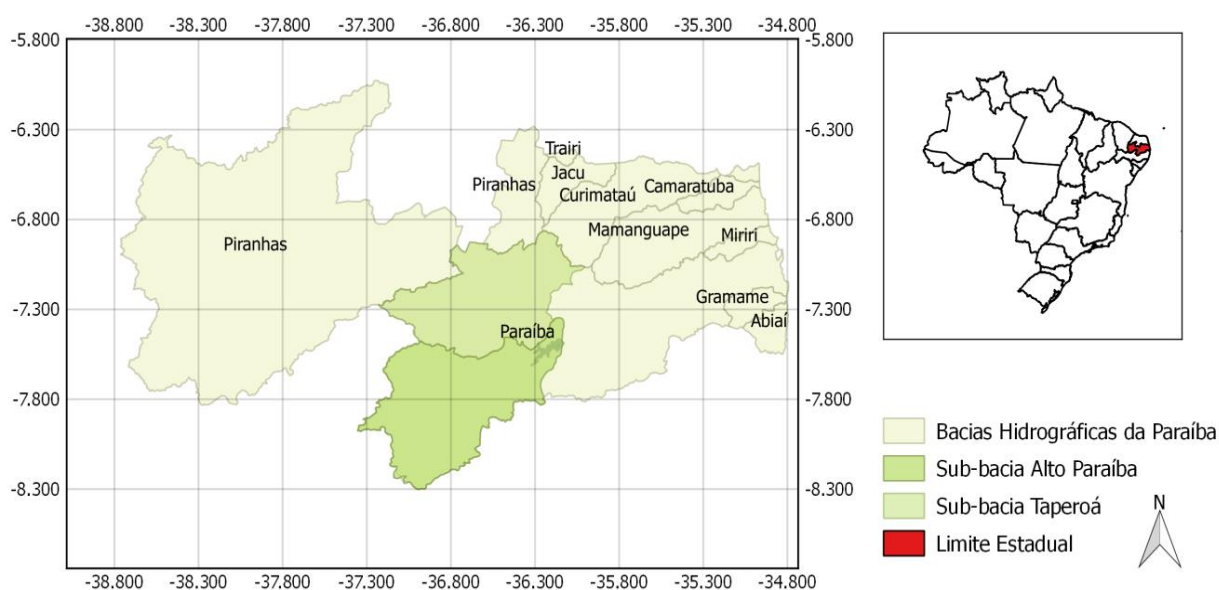


3.1 Área de estudo

A pesquisa foi realizada para a bacia hidrográfica do açude Epitácio Pessoa, também conhecido como açude Boqueirão, situado entre as coordenadas geográficas $07^{\circ} 28' 4''$ e $07^{\circ} 33' 32''$ de latitude Sul e $36^{\circ} 08' 23''$ e $36^{\circ} 16' 51''$ de longitude Oeste. A barragem do açude está localizada no município Boqueirão-PB, a 165 km da capital do estado da Paraíba (João Pessoa) e a 44 km de Campina Grande. O açude em questão tem uma bacia hidráulica que se estende pelos municípios de Boqueirão, Cabaceiras e São Miguel de Taipu, e está a 420m de altitude, especificamente na microrregião do Cariri Oriental paraibano, situado na mesorregião Borborema, no estado da Paraíba.

A bacia hidrográfica do Boqueirão pertence à grande bacia hidrográfica do rio Paraíba, uma dos mais importantes do Estado. A contribuição de vazões para o açude vem, principalmente, das sub-bacias do Alto curso do rio Paraíba e do rio Taperoá (Figura 5).

Figura 5 – Mapa das bacias e sub-bacias hidrográficas do estado da Paraíba



Fonte: elaboração própria (2017).

A sub-bacia hidrográfica do Alto curso do rio Paraíba situa-se na parte sudoeste do Planalto da Borborema, a região do Alto Paraíba localiza-se entre as latitudes $7,3470^{\circ}$ e $8,3030^{\circ}$ Sul e entre as longitudes $36,1280^{\circ}$ e $37,3560^{\circ}$ Oeste de Greenwich. Limita-se ao sul e a oeste com o Estado de Pernambuco, e ao norte com a bacia hidrográfica do rio Taperoá.

A sub-bacia do Rio Taperoá se situa na parte central do Estado da Paraíba, conformando-se sob as latitudes $6,8630^{\circ}$ e $7,5760^{\circ}$ Sul e entre as longitudes $36,1670^{\circ}$ e $37,0230^{\circ}$

Oeste. Limita-se com as sub-bacias hidrográficas do rio Espinharas e do rio Seridó a oeste, com a região do Alto curso do rio Paraíba ao sul, com as bacias do Jacu e Curimataú ao norte e com a região do Médio curso do rio Paraíba a leste. Seu principal rio é o Taperoá, de regime intermitente que nasce na Serra do Teixeira e desemboca no rio Paraíba.

Em termos climatológicos, a bacia hidrográfica do açude Epitácio Pessoa está inserida no clima do tipo Bsh, de acordo com a classificação de Köppen, ou seja, semiárido quente, com estação seca. É considerada a região mais seca do Brasil. As variações de temperatura atingem mínimas mensais de 18°C a 22°C entre os meses de julho e agosto, e máximas mensais de 28°C a 31°C entre os meses de novembro e dezembro. O regime pluviométrico apresenta precipitações médias anuais que variam entre 250 e 750 mm, sendo que os totais anuais se concentram em um período de quatro meses, que, geralmente, corresponde aos meses de fevereiro, março, abril e maio. A umidade relativa do ar alcança uma média mensal de 60 a 75%, observando-se que os valores máximos ocorrem, geralmente, no mês de junho e os mínimos no mês de dezembro. Os totais anuais da evaporação, medidos em tanque Classe A, variam entre 2.500 a 3.000 mm, com valores decrescentes de oeste para leste (BRITO, 2008).

A vegetação divide-se em três níveis de caatinga, caracterizada segundo os aspectos fitossociológicos, delineada quanto ao porte e condição de adensamento: vegetação arbustiva aberta; vegetação arbustiva arbórea fechada e vegetação arbórea fechada. Os solos encontrados em torno do açude Epitácio Pessoa são, na maioria das vezes, pouco profundos, rasos e organizados em dois ambientes edafológicos: o primeiro se refere a uma superfície semi-evoluída, formada pelos solos Bruno Não Cálcicos e, o segundo, por uma superfície não evoluída, formada pelos Litossolos e Afloramentos de Rocha (BRITO, 2008).

3.2 Dados

Neste estudo foram considerados valores de precipitação mensal de três postos pluviométricos pertencentes à bacia hidrográfica do Rio Paraíba: Barra de São Miguel, Salgadinho e Passagem. Os dados foram adquiridos pela ANA (2017) e a escolha dos postos foi feita baseada na qualidade dos dados, analisando-se a consistência e o tamanho das séries disponíveis.

Para o posto Barra de São Miguel, o período de análise compreendeu janeiro de 1963 a dezembro de 1991, totalizando 348 observações. O período de análise para o posto Salgadinho compreendeu janeiro de 1939 a dezembro de 1989, totalizando 612 observações.

No que se refere ao posto pluviométrico Passagem, o período analisado compreendeu janeiro de 1962 a dezembro de 1989, totalizando 336 observações.

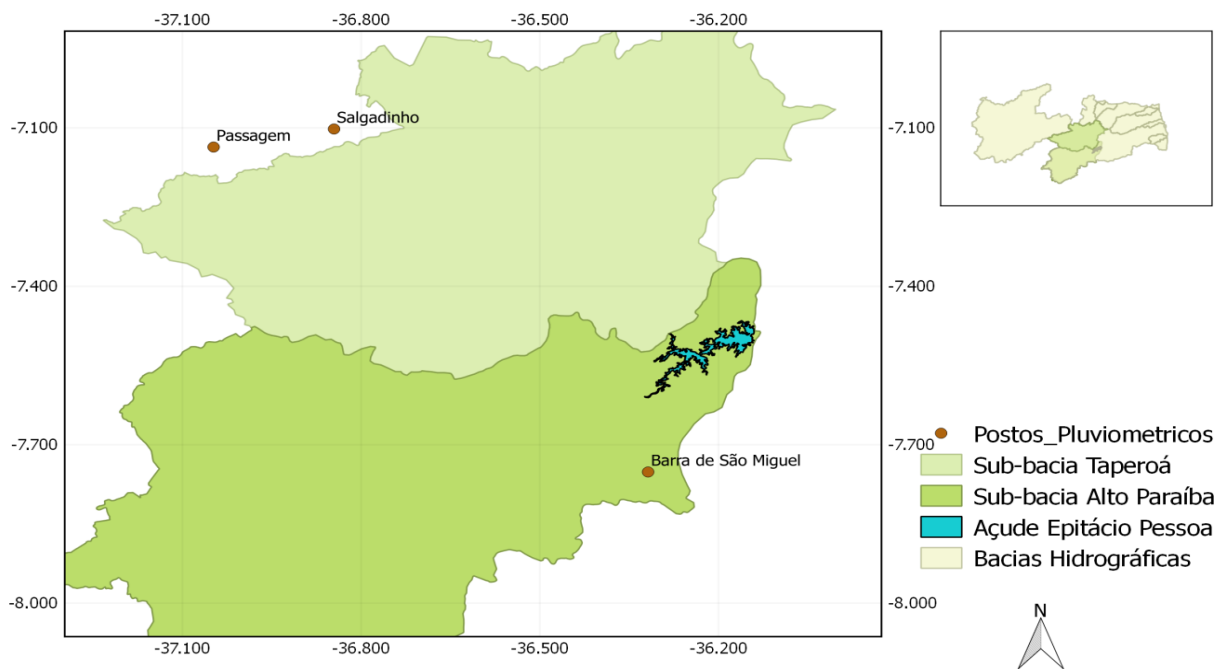
As localizações espaciais dos postos pluviométricos estão fornecidas na Figura 6, e na Tabela 2.

Tabela 2 - Postos pluviométricos selecionados

Número do posto	Código do posto	Longitude (°)	Latitude (°)	Relevo (m)	Município/posto	Período de Dados
1	736025	-36,318	-7,752	520	Barra de São Miguel	1963- 1991
2	737010	-37,048	-7,136	340	Passagem	1962 - 1989
3	736010	-36,845	-7,102	410	Salgadinho	1939 - 1989

Fonte: adaptado de Guedes (2011).

Figura 6 – Localização geográfica dos postos pluviométricos estudados



Fonte: elaboração própria (2017).

3.3 Cálculo do SPI

O cálculo do SPI é determinado pelo ajuste de uma função de densidade de probabilidade que descreve séries históricas de precipitação acumulada para uma escala de tempo de interesse (1 mês, 3 meses, 6 meses, etc). Como a precipitação segue uma distribuição assimétrica, a distribuição gama, definida na Equação (1) é utilizada para representar a distribuição teórica para essa variável.

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0 \quad (1)$$

Onde, α é denominado fator de forma, β fator de escala e x a quantidade de precipitação para o mês de interesse. $\Gamma(\alpha)$ representa a função gama definida como:

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (2)$$

Pelo método de aproximação de Thom (1966), os parâmetros α e β –foram estimados pelo método da máxima verossimilhança, como segue:

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} \right) \quad (3)$$

$$\hat{\beta} = \frac{\bar{x}}{\hat{\alpha}} \quad (4)$$

Sendo :

$$A = \ln(\bar{x}) - x_g \quad (5)$$

Onde, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, é a média aritmética; e $x_g = \frac{1}{n} \sum_{i=1}^n \ln(x_i)$, é a média geométrica das observações.

Após a determinação dos parâmetros α e β , estes foram utilizados para determinar a probabilidade acumulada ($G(x)$) de um valor de precipitação observado para um dado mês e escala de tempo. A função de probabilidade acumulada é dada por:

$$G(x) = \int_0^x g(x) dx = \frac{1}{\hat{\beta} \hat{\alpha} \Gamma(\hat{\alpha})} \int_0^x x^{\hat{\alpha}-1} e^{-x/\hat{\beta}} dx \quad (6)$$

Sabe-se que a função não tem solução imediata, e exige tabelas ou técnicas de integração numérica como expansão em série e a fórmula de *Simpson*, por exemplo. Neste caso, a série que foi utilizada para o cálculo da probabilidade cumulativa é a seguinte:

$$F(\zeta) = \frac{\zeta^{\alpha}}{\alpha \Gamma(\alpha) e^{\zeta}} \left[1 + \frac{\zeta^1}{\alpha + 1} + \frac{\zeta^2}{(\alpha + 1)(\alpha + 2)} + \frac{\zeta^3}{(\alpha + 1)(\alpha + 2)(\alpha + 3)} + \dots \right] \quad (7)$$

Onde:

$$\zeta = \frac{x}{\beta} \quad (8)$$

Visto que a função gama é indefinida para $x = 0$ e uma amostra de dados de precipitação pode conter zeros, a probabilidade acumulada torna-se:

$$H(x) = q + (1 - q)G(x) \quad (9)$$

Onde q é a probabilidade de ocorrência de precipitação zero. Sendo m o número de zeros em uma série de precipitações analisada, Thom (1966) afirma que q pode ser estimado por m/n .

A probabilidade acumulada $H(x)$ é então transformada em variável aleatória normal padronizada Z com média zero e variância um, que é o valor do *SPI*.

$$Z = SPI = + \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \text{ para } 0 < H(x) \leq 0,5 \quad (10)$$

$$Z = SPI = - \left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \text{ para } 0,5 < H(x) < 1,0 \quad (11)$$

Onde:

$$t = \sqrt{\ln \left(\frac{1}{(H(x))^2} \right)} \text{ para } 0 < H(x) \leq 0,5 \quad (12)$$

$$t = \sqrt{\ln \left(\frac{1}{(1,0 - H(x))^2} \right)} \text{ para } 0,5 < H(x) < 1,0 \quad (13)$$

e, $c_0=2,515517$; $c_1=0,802853$; $c_2=0,010328$

$d_1 = 1,432788$; $d_2 = 0,189269$; $d_3 = 0,001308$

O cálculo da função de distribuição de probabilidade das precipitações e do *SPI* foi feito por meio do uso de planilhas do Excel. A função estatística *DISTGAMA* presente no Excel foi utilizada para o cálculo da função de probabilidade acumulada. Além disso, a fim de confrontar os valores obtidos através de tal função, utilizou-se também a fórmula de *Simpson* descrita na Equação (7).

Em seguida calculou-se o índice *SPI* de cada um dos postos pluviométricos para as escalas de tempo de 3 meses, 6 meses, 12 meses e 24 meses. Os valores obtidos foram validados através de comparação com resultados gerados pelo programa computacional *SPI program*, desenvolvido pelo *National drought Mitigation Center* dos Estados Unidos da América. Por fim, foram criadas equações polinomiais baseadas na linha de tendência da relação entre a precipitação e o respectivo *SPI* de cada posto.

3.4 Concepção da rede neural

Os modelos de RNAs apresentados neste estudo foram baseados no uso do SPI pelas seguintes razões:

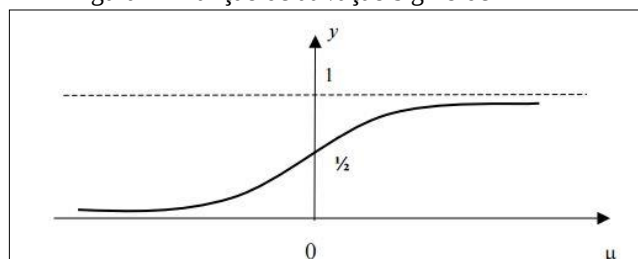
- (a) SPI utiliza apenas valores de precipitação, portanto o cálculo se torna possível mesmo na ausência de outros dados hidrometeorológicos;
- (b) Pode ser computado em diferentes escalas de tempo, o que permite descrever condições de seca para diferentes aplicações seja meteorológica, hidrológica, agrícola, etc;
- (c) Os valores são normalizados, o que garante que a frequência de eventos extremos sejam consistentes em qualquer local e escala de tempo (RAZMKAH; ROSTAMI, 2015).

As RNAs utilizadas no presente estudo foram do tipo MLP com três camadas. A camada de entrada, não faz processamento, portanto, não se atribui a ela funções de ativação. Nessa camada há apenas um neurônio onde é inserido o valor normalizado da precipitação que modificado por um peso envia seu dado de entrada para os neurônios da próxima camada.

Na camada intermediária adotou-se a função de transferência sigmoide e variou-se o número de neurônios a fim de escolher o modelo que melhor prediz os valores de SPI. Na camada de saída, que possui apenas um neurônio, também foi adotada a função de transferência sigmoide.

A função sigmoide (Figura 7) é caracterizada por ser uma função não linear contínua que produz valores entre (0,1). Devido às suas características, as funções do tipo sigmoide são comumente utilizadas para resolver problemas não lineares. Assim, de acordo com Tayfur (2012), muitos problemas de engenharia aplicam tal função para a resolução de problemas.

Figura 7 - Função de ativação sigmóide



Fonte: Guimarães e Chaves (2002).

Em todas as RNAs, o treinamento foi realizado utilizando o algoritmo de aprendizado backpropagation. De acordo com Tayfur (2012), esse método de treinamento é, provavelmente, o mais aplicado e conhecido, além de ser robusto e possuir uma alta capacidade de aprendizagem.

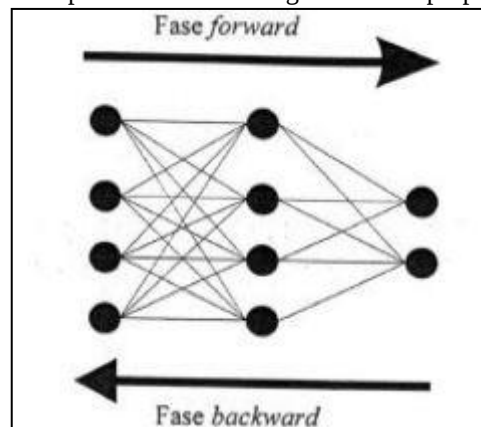
Proposto em 1986 por Rumelhart, Hinton e Willian, o método *backpropagation* é um procedimento de aprendizagem para redes *feedforward* de múltiplas camadas (TISSOT. et al. 2012). Basicamente, o procedimento de aprendizagem consiste em uma atividade de retropropagação do erro.

O aprendizado é realizado por ajuste iterativo dos pesos da rede para minimizar as diferenças entre a saída atual e a saída desejada. A rede é inicializada com pesos aleatórios. Durante o processo de aprendizagem, um vetor de entrada é apresentado para a rede e propagado para determinar os valores da saída. A seguir, o vetor de saída produzido pela rede é comparado com o vetor de saída desejado, resultando num sinal de erro, que é retropropagado através da rede para permitir o ajuste dos pesos. As atividades de retropropagação são as seguintes:

- (1) O erro é computado para cada unidade de saída;
- (2) Camada por camada, o erro de cada unidade intermediária é computado por erros de propagação;
- (3) Finalmente, os pesos são atualizados pelo método da regra do delta generalizado.

Este processo é repetido até que a rede responda, para cada vetor de entrada, um vetor de saída com valores suficientemente próximos dos valores desejados (Figura 8).

Figura 8 – Fluxo de processamento do algoritmo backpropagation



Fonte: Braga, Ludemir e Carvalho (2000).

É importante salientar que para a aplicação às redes neurais artificiais é necessário a realização prévia de duas etapas: (1) a coleta dos dados e a sua separação em dados de treinamento e dados de teste, para verificar o desempenho da rede. Segundo o qual eles são separados de forma aleatória de modo a evitar erros associados à ordem dos dados e, (2) criação de conjunto de validação que verifica a eficiência da rede quanto a sua capacidade de generalização durante o treinamento.

Em seguida, o cálculo de treinamento de uma RNA do tipo *backpropagation* é o feito como se segue.

A conexão dos pesos é ajustada baseada na Equação 14:

$$w_{ij}^{novo} = w_{ij}^{velho} - \eta \frac{\partial E(w_{ij})}{\partial w_{ij}} \quad (14)$$

Onde, w_{ij} é o valor do peso da conexão entre os neurônios da camada intermediária e de saída, η é o fator de aprendizagem e E é o erro entre o valor de saída e o valor esperado.

O erro (E) é calculado como mostrado abaixo (Equação 15).

$$E = \frac{1}{2} \sum_{k=1}^p \sum_{j=1}^q (t_{kj} - z_{kj})^2 \quad (15)$$

Onde, p é o número de padrões de treinamento, q é o número de neurônios de saída; t_{kj} é o valor de saída que se pretende atingir e z_{kj} é o valor de saída na camada de saída e é expresso por (Equação 16):

$$z_j = f\left(\sum_{i=1}^q (w_{ij} y_i)\right) \quad (16)$$

Onde, y_j é o valor de saída da camada intermediária.

Como mencionado anteriormente, a atualização dos pesos é realizada através da computação dos erros de forma retropropagada. Assim, inicialmente, os pesos são computados da camada de saída para a camada intermediária, através da Equação 17.

$$w_{ij}^{novo} = w_{ij}^{velho} - \eta \beta_{oj} y_j \quad (17)$$

Onde, β_{oj} é descrito pela Equação 18 que segue abaixo.

$$\beta_{oj} = \sum_{k=1}^p \sum_{j=1}^q (t_{kj} - z_{kj})(-1)f'(w_{ij}) \quad (18)$$

Observa-se que para a resolução da equação acima é necessário a derivada da função de ativação aplicada. Assim, para o caso da função de ativação sigmoide, a Equação 18 pode ser reescrita como se segue (Equação 19):

$$\frac{\partial E}{\partial w_{ij}} = \sum_{k=1}^p \sum_{j=1}^q (t_{kj} - z_{kj})(-1)f(w_{ij})[1 - f(w_{ij})]y_j \quad (19)$$

Em seguida, o ajuste dos pesos entre a camada de entrada e intermediária é feita por (Equação 20):

$$v_{ij}^{novo} = v_{ij}^{velho} - \eta \beta_{mj} x_i \quad (20)$$

Onde, v_{ij} é o peso entre os neurônios da camada entrada e intermediária; η é o fator de aprendizagem, x_i é o valor de saída da camada de entrada e β_{mj} é definido por (Equação 21):

$$\beta_{mj} = \sum_{k=1}^p \sum_{j=1}^q (t_{kj} - z_{kj}) (-1) f'(y_j) w_{ij} f'(v_{ij}) \quad (21)$$

A Equação 21 pode ser redefinida para a função sigmoide como mostrado abaixo (Equação 22):

$$v_{ij}^{novo} = v_{ij}^{velho} - \eta \sum_{k=1}^p \sum_{j=1}^q (t_{kj} - z_{kj}) (-1) f(w_{ij}) [1 - f(w_{ij})] y_j [1 - y_j] w_{ij} x_i \quad (22)$$

Assim, para esse estudo foram criados dois grupos de RNAs para os valores SPI 24, SPI 12 e SPI6, utilizando o algoritmo acima descrito. As redes neurais desenvolvidas apresentara um neurônio de entrada referente ao valor de precipitação acumulada e um neurônio de saída referente ao valor SPI observado. Variou-se o número de neurônios na camada intermediária para 5, 10, 15 e 20 neurônios, totalizando 32 RNAs distintas.

O primeiro grupo de RNAs foi criado sem a influência da defasagem temporal, ou seja, os valores de entrada de precipitação acumulada e os valores SPI de saída correspondem ao mesmo período de tempo analisado.

O segundo grupo de RNAs foi criado a partir da defasagem temporal da série de dados. Dessa forma, os valores de saída referem-se ao valores de precipitação acumulada do período anterior.

Portanto, neste estudo utilizaram-se diversas arquiteturas de RNA com o algoritmo *backpropagation* a fim de avaliar o uso de redes neurais como método alternativo ao cálculo do SPI – através da construção de RNAs sem defasagem temporal de dados, e sua capacidade de previsão – através do uso de RNAs com defasagem temporal de dados.

3.4.1 Treinamento da RNA

A fase de treinamento das RNAs foi realizada com os dados do posto pluviométrico Salgadinho. O posto pluviométrico Salgadinho foi utilizado por possuir a maior sequência de observações temporais disponíveis dentre os três postos pluviométricos avaliados no presente estudo.

Normalização de dados

A normalização dos dados de entrada foi realizada considerando que a função sigmoide, devido à presença da exponencial, pode causar um treinamento da rede neural pouco adequado para um grande número de valores de entrada. A fim de evitar tal limitação, normalizam-se os valores de entrada da rede. O mesmo acontece para os valores de saída, já que a função sigmoide gera valores entre 0 e 1.

Assim, antes do início do treinamento todos os dados de entrada e saída foram normalizados através da Equação 23 proposta por Tayfur (2012), que comprime os dados em um alcance entre (0,1; 0,9) considerando os limites da função sigmoide.

$$z_i = 0,1 + \frac{0,8(x_i - x_{min})}{x_{max} - x_{min}} \quad (23)$$

Onde, z_i é o valor normalizado, x_i é o valor não normalizado de entrada, x_{min} é o valor mínimo dos dados que se pretende normalizar e x_{max} é o valor máximo dos dados que se pretende normalizar.

A Equação 23 foi adotada para esse estudo, pois, de acordo com Tayfur (2012), pode ser utilizada para a maior parte das aplicações de RNA.

Separação dos dados

Os dados foram separados em valores para treinamento, validação e teste, correspondentes a 60, 20 e 20% dos valores disponíveis, respectivamente.

Adoção de valores de parâmetros

Antes de iniciar o treinamento da rede, deve-se assumir valores aleatórios para os pesos e vieses. Seguindo a regra prática descrita por Tayfur (2012), assumiu-se valores entre $(-3/n; 3/n)$, onde n é o número de neurônios da camada de entrada. Assim, para as RNAs 5, 10, 15 e 20 foram estabelecidos valores iniciais que podem ser visualizados no *Apêndice A*.

Além disso, como mencionado anteriormente, existe um fator η (fator de aprendizagem), no qual seu valor adotado influencia diretamente no ajuste dos pesos. A partir de observações do comportamento das redes neurais trabalhadas, adotou-se $\eta = 0,0001$.

Validação cruzada (K fold)

Podem ocorrer erros metodológicos em processos de aprendizagem de parâmetros em funções de predição. Isso ocorre quando o modelo se ajusta em demasiado ao conjunto de dados/amostra, o que significa que um modelo pode simplesmente repetir os padrões das amostras de forma a memorizar exemplos individuais e falhar na predição de dados ainda não vistos pelo modelo. Essa situação de treinamento excessivo é chamada na literatura de *overtraining*.

Para evitar tal problema, a técnica estatística da validação cruzada pode ser aplicada para o subconjunto de dados de treinamento e validação. Dessa forma, a técnica consiste em dividir os exemplos em subconjuntos de ‘K folds’ disjuntos com o mesmo número de exemplos retirados sem reposição da base original. Cada partição é usada como teste para a avaliação do classificador, onde as k-1 partições restantes são utilizadas como treinamento. Assim, são produzidos k resultados experimentais estatisticamente independentes, onde toda a base de dados é utilizada para teste.

Para o presente trabalho, os dados de treinamento e validação foram divididos em 10 partições a partir da técnica de validação cruzada (K fold). Ao final do treinamento de cada partição, os valores ajustados dos seus respectivos pesos e vieses foram armazenados. Por fim, para a rede treinada os pesos e vieses adotados foram obtidos a partir do cálculo das médias dos valores ajustados das 10 partições treinadas. No final de cada treinamento, com os pesos e vieses ajustados, a RNA foi simulada para gerar resultados.

A rede neural utilizada neste trabalho foi criada no software R - Toolbox *neuralnet*, que através de um processo iterativo aplicando o algoritmo *backpropagation*, treinou redes neurais para os variados números de neurônios na camada intermediária (Apêndice B).

Índices de Desempenho

Para se avaliar o desempenho de todas as RNAs utilizaram-se duas estatísticas: Erro Quadrático Médio (EQM) e Coeficiente de Eficiência (CE)

- i) *Erro Quadrático Médio*: é a medida de erro mais comumente usada para aferir a qualidade do ajuste de um modelo (LOPES; BARBOSA, 2014). É dada pela

raiz do erro médio quadrático da diferença entre a predição e o valor real. Assim:

$$EQM = \sqrt{\frac{\sum_i^N (O_{mi} - O_{pi})^2}{N}} \quad (24)$$

Onde, O_{mi} é o valor de saída observado; O_{pi} é o valor previsto pela rede e N é o número de observações.

- ii) *Coefficiente de Eficiência*: refere-se aos erros do modelo em estimar as médias ou variâncias do conjunto de dados observados. CE varia de menos infinito (pobre modelação) a 1,0 (modelo perfeito).

$$CE = 1,0 - \frac{\sum_i^N (O_{mi} - O_{pi})^2}{\sum_i^N (O_{mi} - \bar{O}_m)^2} \quad (25)$$

Onde, $\bar{O}_m$ é a média dos valores de saída observados.

4. RESULTADOS E DISCUSSÃO

4.1 Valores SPI

A partir da série histórica de precipitação mensal dos postos estudados foi possível calcular seus respectivos valores de índice de precipitação SPI para as diferentes escalas de tempo. De modo a facilitar a leitura dos resultados, adotaram-se dois gráficos distintos. O primeiro fornece a dispersão dos valores de SPI em função da precipitação acumulada em mm. O segundo corresponde às séries de SPI em função do tempo, ou seja, representa a sequência temporal dos SPI resultantes da série de precipitação em função da escala de tempo adotada. Os gráficos podem ser visualizados na Figura 9, Figura 10 e Figura 11.

Figura 9- Valores SPI para o posto pluviométrico Barra de São Miguel, PB

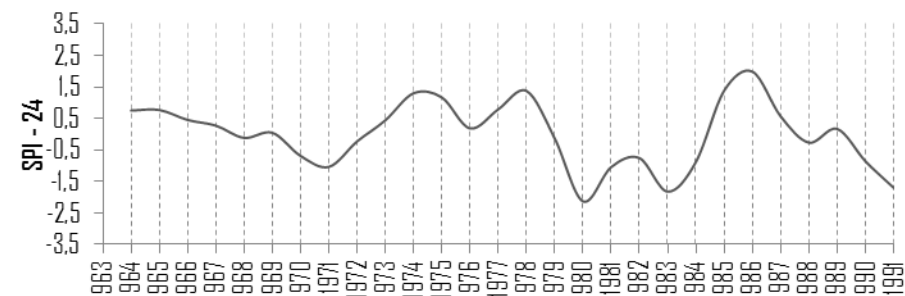
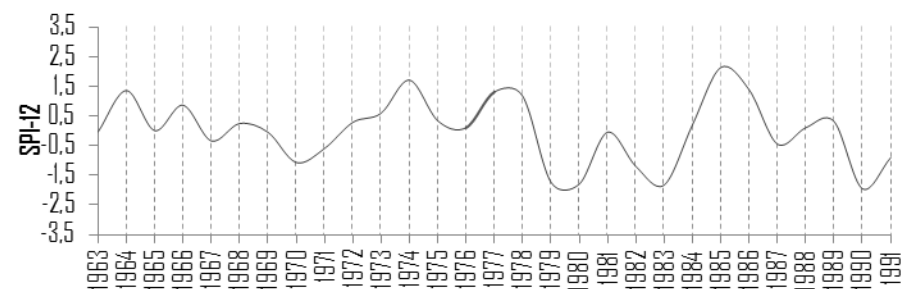
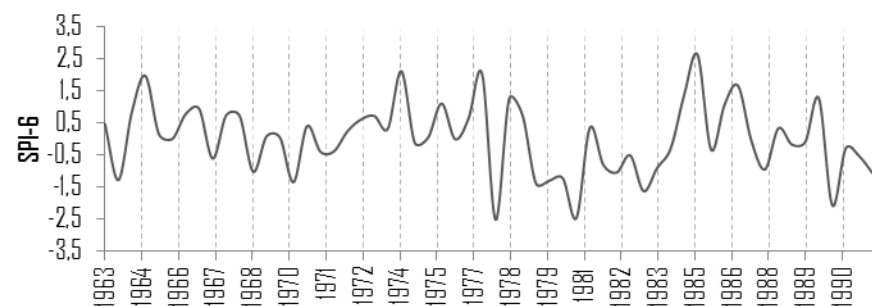
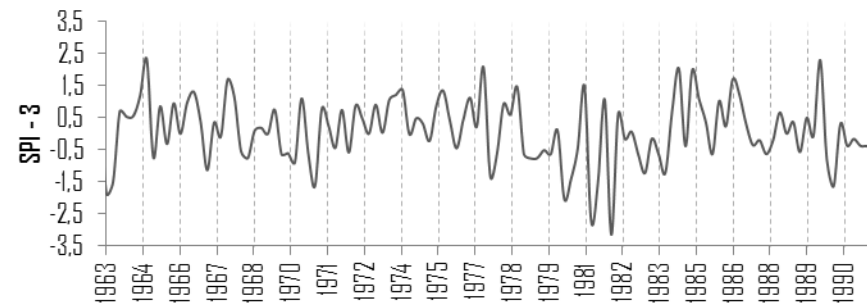
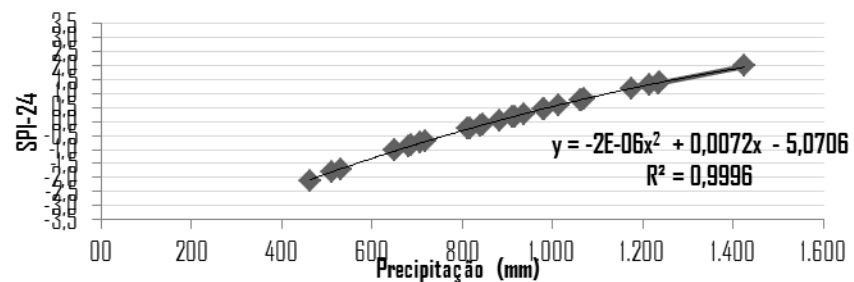
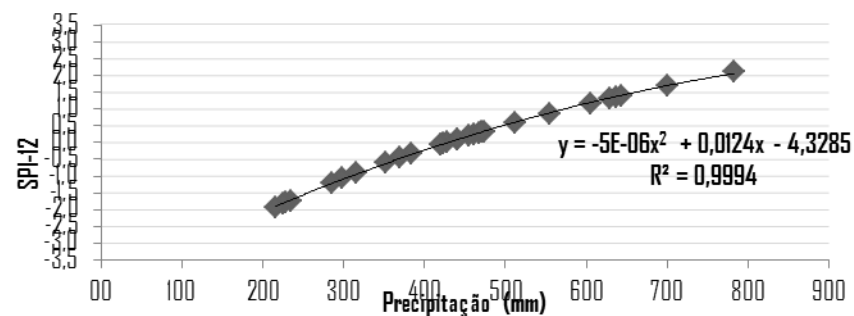
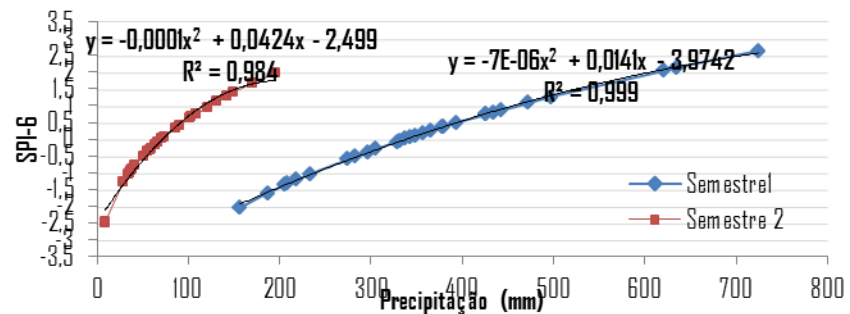
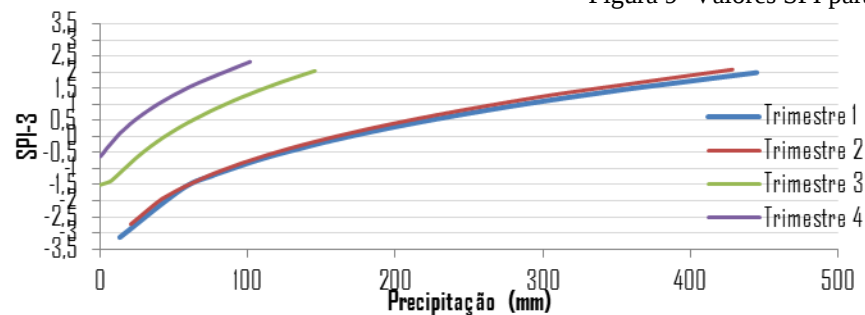


Figura 10 - Valores SPI para o posto pluviométrico Passagem, PB

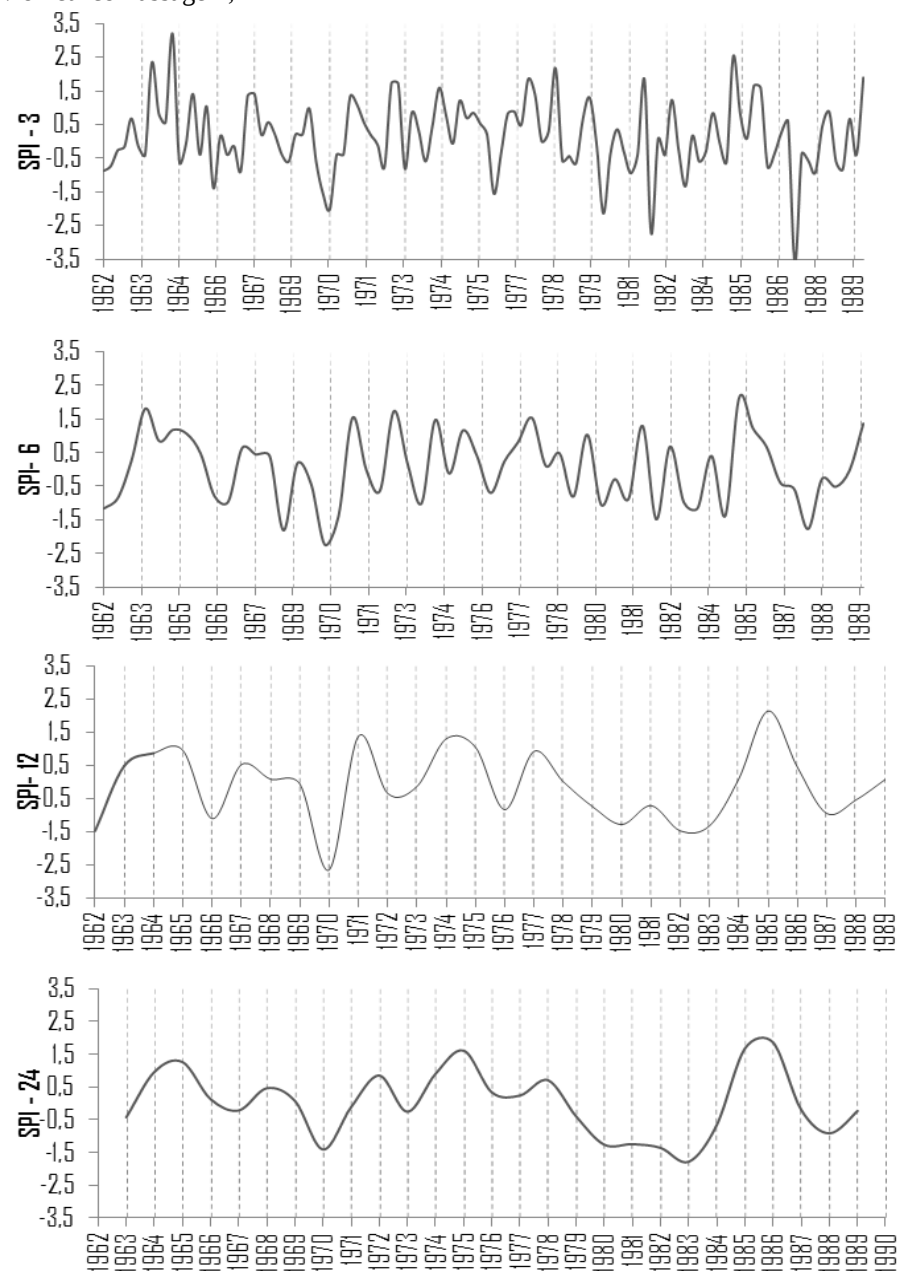
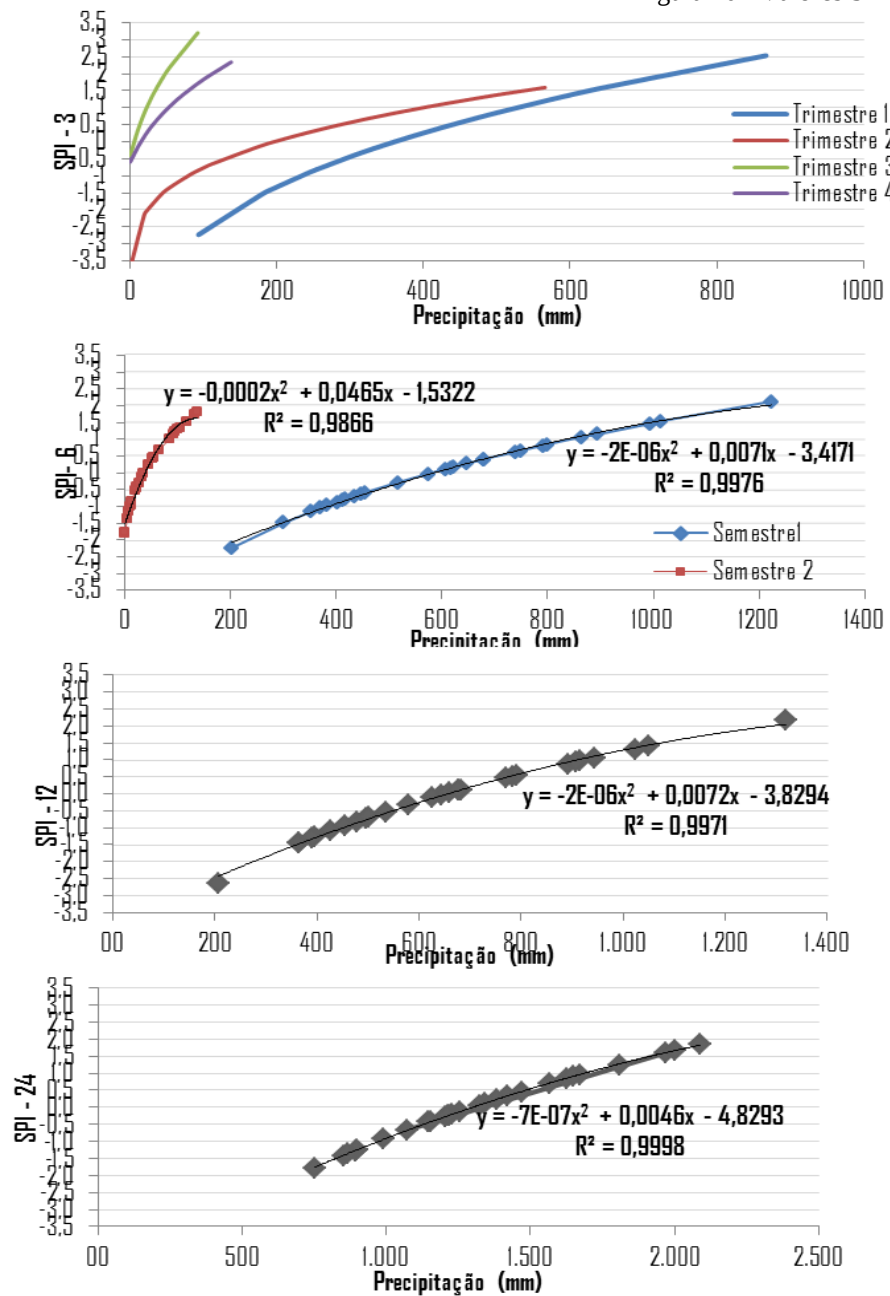
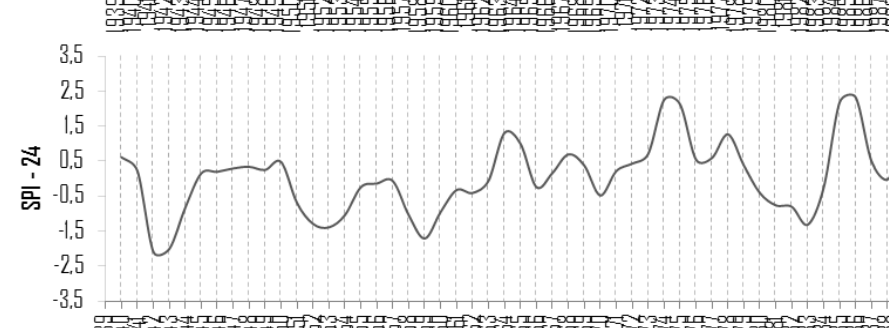
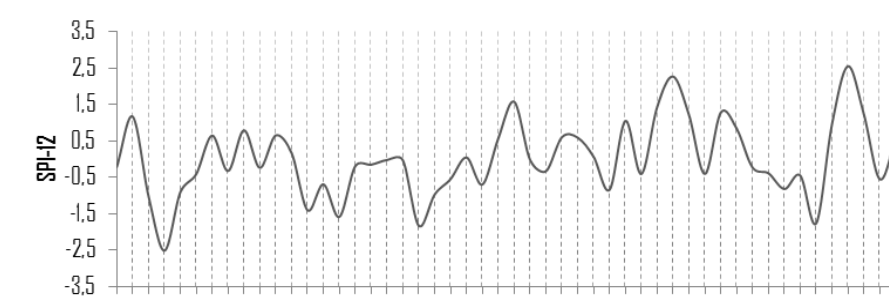
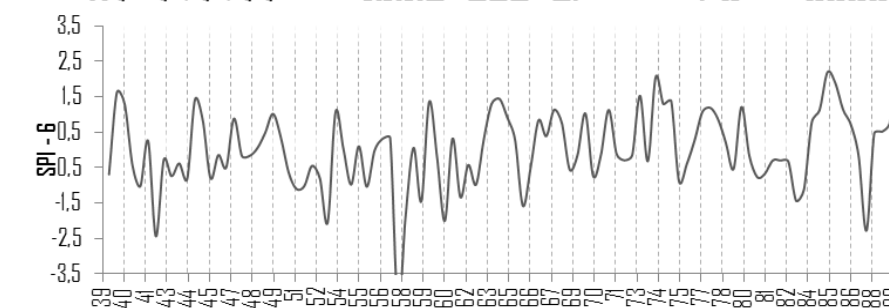
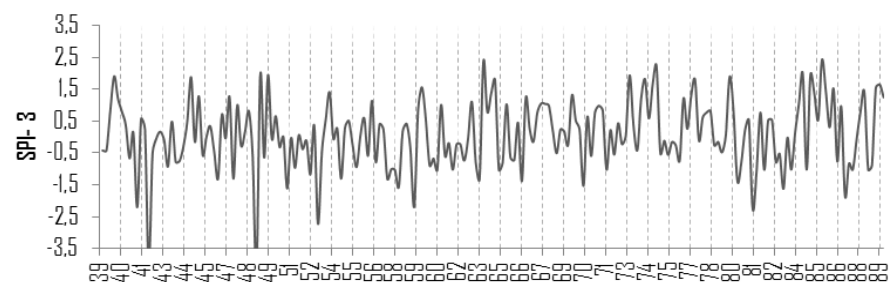
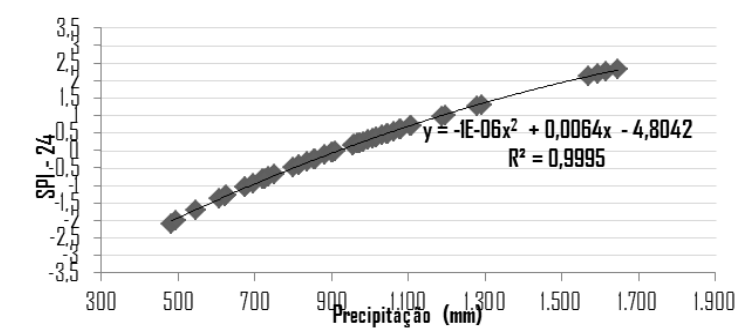
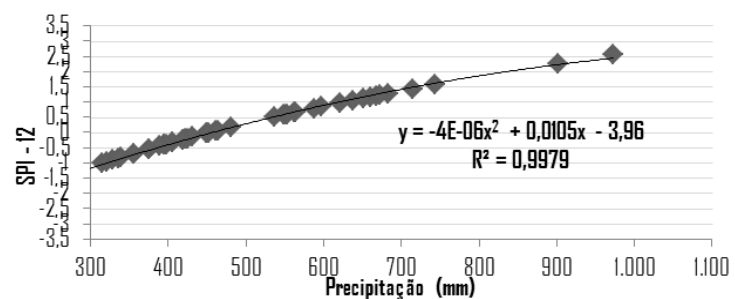
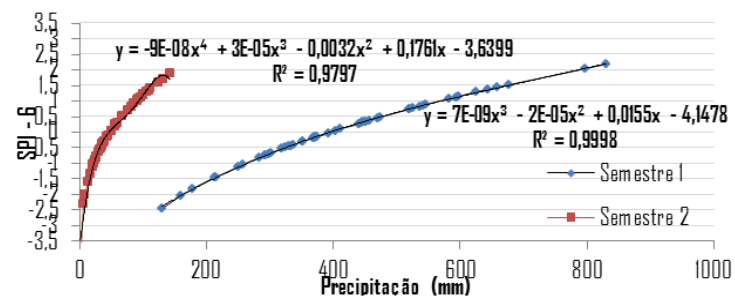
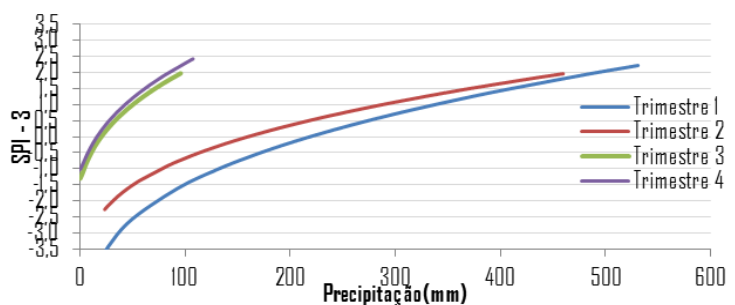


Figura 11 - Valores SPI para o posto pluviométrico Salgadinho, PB



A partir dos resultados obtidos foi possível estabelecer equações polinomiais que relacionam a dispersão entre valores de precipitação acumulada e valores de SPI para cada posto estudado. O cálculo dessas equações permite encontrar o valor do SPI diretamente a partir do valor de precipitação acumulada sem a necessidade de cumprir as etapas descritas na metodologia apresentada na *Seção 3.3*

As equações de regressão para o posto Barra de São Miguel estão demonstradas abaixo.

SPI 3 - Trimestre

1
(mês 3)

$$SPI3 = -9e^{-8}P^3 - 8e^{-5}P^2 + 0,0311P - 3,2721$$

$$R^2 = 0,9935$$
(26)

Onde, P é o valor de precipitação acumulada, e R^2 é o coeficiente de determinação. Para a Equação 26, o intervalo de validade das precipitações acumuladas é: (13,1; 44,9).

SPI 3 - Trimestre

2
(mês 6)

$$SPI3 = -2e^{-5}P^2 + 0,0199P - 2,6757$$

$$R^2 = 0,9867$$
(27)

Para a Equação 27, o intervalo de validade das precipitações acumuladas é: (20,6; 428,5).

SPI 3 - Trimestre

3
(mês 9)

$$SPI3 = -0,0001P^2 + 0,0393P - 1,5439$$

$$R^2 = 0,998$$
(28)

Para a Equação 28, o intervalo de validade das precipitações acumuladas é: (0; 145,4).

SPI 3 - Trimestre

4
(mês 12)

$$SPI3 = -0,0002P^2 + 0,0493P - 0,5913$$

$$R^2 = 0,9974$$
(29)

Para a Equação 29, o intervalo de validade das precipitações acumuladas é: (0; 101,5).

$$\begin{array}{ll}
 \text{SPI 6 - Semestre 1} & SPI6 = -7e^{-6}P^2 + 0,0141P - 3,9742 \\
 \text{(mês 6)} & R^2 = 0,999
 \end{array} \quad (30)$$

Para a Equação 30, o intervalo de validade das precipitações acumuladas é: (156,1; 724,8).

$$\begin{array}{ll}
 \text{SPI 6 - Semestre 2} & SPI6 = -0,0001P^2 + 0,0424P - 2,499 \\
 \text{(mês 6)} & R^2 = 0,984
 \end{array} \quad (31)$$

Para a Equação 31, o intervalo de validade das precipitações acumuladas é: (9,2; 195,6).

$$\begin{array}{ll}
 \text{SPI 12} & SPI12 = -5e^{-6}P^2 + 0,0124P - 4,3285 \\
 \text{(mês 12)} & R^2 = 0,9994
 \end{array} \quad (32)$$

Para a Equação 32, o intervalo de validade das precipitações acumuladas é: (215,4; 782,5).

$$\begin{array}{ll}
 \text{SPI 24} & SPI24 = -2e^{-6}P^2 + 0,0072P - 5,0706 \\
 \text{(mês 12)} & R^2 = 0,9996
 \end{array} \quad (33)$$

Para a Equação 33, o intervalo de validade das precipitações acumuladas é: (462,4; 1425,3).

As equações de regressão para o posto Passagem estão demonstradas abaixo.

$$\begin{array}{ll}
 \text{SPI 3 - Trimestre} & SPI3 = -5e^{-6}P^2 + 0,0114P - 3,4376 \\
 \text{1} & R^2 = 0,9937 \\
 \text{(mês 3)} &
 \end{array} \quad (34)$$

Para a Equação 34, o intervalo de validade das precipitações acumuladas é: (93,4; 867).

$$\begin{aligned} \text{SPI 3 - Trimestre 2} \quad SPI3 &= -3e^{-10}P^4 + 3e^{-7}P^3 - 0,0002P^2 + 0,0367P - 3,0683 \\ (\text{mês 6}) \quad R^2 &= 0,9829 \end{aligned} \quad (35)$$

Para a Equação 35, o intervalo de validade das precipitações acumuladas é: (1,5; 565,6).

$$\begin{aligned} \text{SPI 3 - Trimestre} \quad SPI3 &= -0,0003P^2 + 0,066P - 0,3815 \\ 3 \quad R^2 &= 0,9977 \\ (\text{mês 9}) \end{aligned} \quad (36)$$

Para a Equação 36, o intervalo de validade das precipitações acumuladas é: (0; 92,3).

$$\begin{aligned} \text{SPI 3 - Trimestre} \quad SPI3 &= -0,0001P^2 + 0,0354P - 0,5594 \\ 4 \quad R^2 &= 0,9972 \\ (\text{mês 12}) \end{aligned} \quad (37)$$

Para a Equação 37, o intervalo de validade das precipitações acumuladas é: (0; 137,8).

$$\begin{aligned} \text{SPI 6 - Semestre 1} \quad SPI6 &= -2e^{-6}P^2 + 0,0071P - 3,4171 \\ (\text{mês 6}) \quad R^2 &= 0,9976 \end{aligned} \quad (38)$$

Para a Equação 38, o intervalo de validade das precipitações acumuladas é: (202,9; 1223).

$$\begin{aligned} \text{SPI 6 - Semestre 2} \quad SPI6 &= -0,0002P^2 + 0,0465P - 1,5322 \\ (\text{mês 6}) \quad R^2 &= 0,9866 \end{aligned} \quad (39)$$

Para a Equação 39, o intervalo de validade das precipitações acumuladas é: (0; 137,8).

$$\begin{aligned} \text{SPI 12} \quad SPI12 &= -2e^{-6}P^2 + 0,0072P - 3,8294 \\ (\text{mês 12}) \quad R^2 &= 0,9971 \end{aligned} \quad (40)$$

Para a Equação 40, o intervalo de validade das precipitações acumuladas é: (208; 1318,3).

$$\begin{aligned}
 \text{SPI 24} \quad & SPI24 = -7e^{-7}P^2 + 0,0046P - 4,8293 \\
 \text{(mês 12)} \quad & R^2 = 0,9998
 \end{aligned} \tag{41}$$

Para a Equação 41, o intervalo de validade das precipitações acumuladas é: (754,0; 2089,3).

As equações de regressão para o posto Salgadinho estão demonstradas abaixo.

$$\begin{aligned}
 \text{SPI 3 - Trimestre} \quad & SPI3 = -2e^{-5}P^2 + 0,022P - 3,7955 \\
 \text{1} \quad & R^2 = 0,9743 \\
 \text{(mês 3)}
 \end{aligned} \tag{42}$$

Para a Equação 42, o intervalo de validade das precipitações acumuladas é: (13,4; 531,0).

$$\begin{aligned}
 \text{SPI 3 - Trimestre} \quad & SPI3 = -2e^{-5}P^2 + 0,0172P - 2,3278 \\
 \text{2} \quad & R^2 = 0,9883 \\
 \text{(mês 6)}
 \end{aligned} \tag{43}$$

Para a Equação 43, o intervalo de validade das precipitações acumuladas é: (23,4; 459,7).

$$\begin{aligned}
 \text{SPI 3 - Trimestre} \quad & SPI3 = -0,0003P^2 + 0,0606P - 1,1912 \\
 \text{3} \quad & R^2 = 0,9943 \\
 \text{(mês 9)}
 \end{aligned} \tag{44}$$

Para a Equação 44, o intervalo de validade das precipitações acumuladas é: (0; 95,7).

$$\begin{aligned}
 \text{SPI 3 - Trimestre} \quad & SPI3 = -0,0003P^2 + 0,0572P - 0,9398 \\
 \text{4} \quad & R^2 = 0,994 \\
 \text{(mês 12)}
 \end{aligned} \tag{45}$$

Para a Equação 45, o intervalo de validade das precipitações acumuladas é: (0; 107,4).

$$\begin{array}{ll} \text{SPI 6 - Semestre 1} & SPI6 = -7e^{-9}P^3 - 2e^{-5}P^2 + 0,0155P - 4,1478 \\ \text{(mês 6)} & R^2 = 0,9998 \end{array} \quad (46)$$

Para a Equação 46, o intervalo de validade das precipitações acumuladas é: (129,1; 830,1).

$$\begin{array}{ll} \text{SPI 6 - Semestre 2} & SPI6 = -9e^{-8}P^4 + 3e^{-5}P^3 - 0,0032P^2 + 0,1761P - 3,6399 \\ \text{(mês 6)} & R^2 = 0,9797 \end{array} \quad (47)$$

Para a Equação 47, o intervalo de validade das precipitações acumuladas é: (0,2; 141,8).

$$\begin{array}{ll} \text{SPI 12} & SPI12 = -4e^{-6}P^2 + 0,0105P - 3,96 \\ \text{(mês 12)} & R^2 = 0,9979 \end{array} \quad (48)$$

Para a Equação 48, o intervalo de validade das precipitações acumuladas é: (168,1; 971,9).

$$\begin{array}{ll} \text{SPI 24} & SPI24 = -1e^{-6}P^2 + 0,0064P - 4,8042 \\ \text{(mês 12)} & R^2 = 0,9995 \end{array} \quad (49)$$

Para a Equação 49, o intervalo de validade das precipitações acumuladas é: (482,5; 1644,6).

A obtenção de equações que relacionam valores de SPI com dados de precipitação acumulada de um determinado sítio de interesse mostra-se como uma alternativa conveniente ao cálculo do índice de precipitação normalizado apresentado no presente trabalho. Dessa forma, é possível abdicar da laboriosa tarefa de ajuste da função de densidade de probabilidade gama que descreve séries históricas de precipitação acumulada para a escala de tempo de interesse e sua respectiva normalização.

Como pode ser visto nas Tabela 3, Tabela 4 e Tabela 5, a curva do SPI tende à suavização à medida que se passa da escala temporal de 3 a 24 meses. Como mencionado por Guedes (2011), quanto maior for a escala temporal, menor será a variação, haja vista que as séries representam a precipitação acumulada para o mesmo período. Por essa razão, os gráficos com escala maior apresentam um atraso no início de suas séries.

4.2 Redes neurais

As mesmas RNAs treinadas para o posto Salgadinho foram aplicadas para os postos restantes. Em relação às redes neurais artificiais, dos resultados obtidos, apenas os melhores foram selecionados para análise. O critério de seleção das melhores arquiteturas foi o de avaliar o desempenho de generalização da RNA por meio de análises estatísticas entre os dados calculados e os observados para o período de teste. Os resultados estatísticos das redes neurais sem defasagem temporal são apresentados na Tabela 3.

Tabela 3 - Resultados estatísticos das redes neurais artificiais (RNAs) sem defasagem temporal de dados para os Índices de Precipitação Padronizado (SPI)

		Treinamento e Validação		Teste	
		Erro Quadrático Médio	Coefficiente de Eficiência	Erro Quadrático Médio	Coefficiente de Eficiência
SPI 24	RNA-5	2,28	-4,63	1,21	-0,04
	RNA-10	1,19	-0,55	1,02	0,75
	RNA-15	1,20	-0,56	1,02	0,75
	RNA-20	1,20	-0,57	1,02	0,75
SPI 12	RNA-5	0,97	1,32	0,00	-1,25
	RNA-10	1,34	-0,92	1,20	0,85
	RNA-15	1,28	-0,75	1,20	0,85
	RNA-20	1,24	-0,65	1,20	0,85
SPI 6 - Semestre 1	RNA-5	0,98	0,00	1,04	-0,05
	RNA-10	0,17	0,99	0,97	0,96
	RNA-15	0,18	0,97	0,99	0,96
	RNA-20	0,18	0,96	0,99	0,96
SPI 6 - Semestre 2	RNA-5	1,15	0,00	1,36	-0,18
	RNA-10	0,42	0,87	1,24	0,98
	RNA-15	0,43	0,86	1,24	0,99
	RNA-20	0,44	0,85	1,25	0,99

Observa-se na Tabela 3, que em todos os casos as RNAs com 5 neurônios na camada intermediária possuem uma baixa capacidade de modelar os dados. As RNAs 10, 15 e 20 possuem valores de coeficiente de eficiência iguais ou próximos. Assim, por meio da análise

estatística de eficiência do modelo e considerando a maior simplicidade possível, as RNAs do tipo 10 neurônios na camada intermediária foram escolhidas para representar todos os casos de SPI trabalhados sem defasagem temporal de dados.

As Figuras 12, 13, 14 e 15 mostram os resultados obtidos através do uso das redes neurais artificiais selecionadas para os três postos pluviométricos do presente estudo.

Figura 12 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 24 sem defasagem temporal de dados

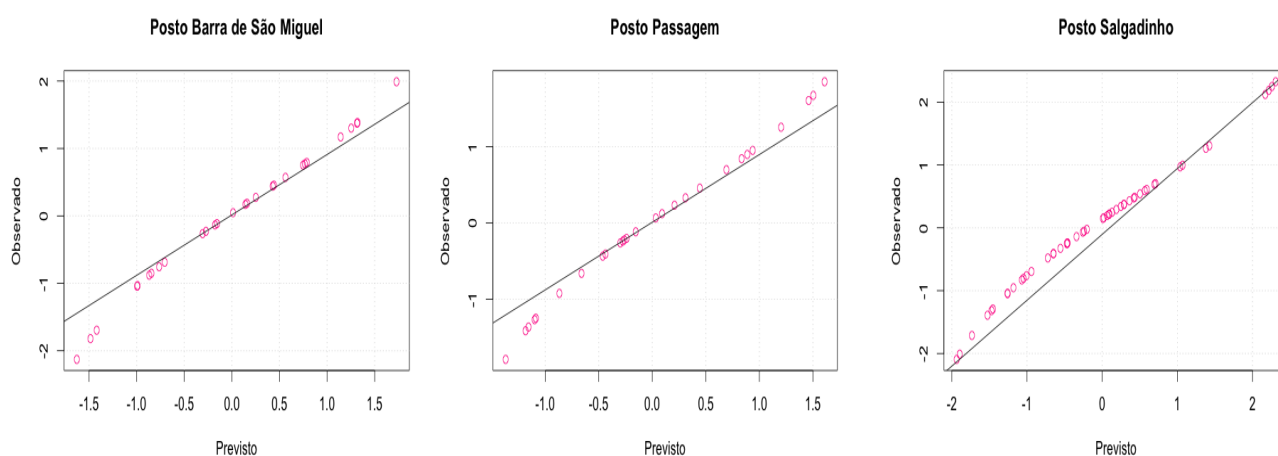


Tabela 4 - Resultados estatísticos SPI 24 sem defasagem temporal de dados

	Índice de Precipitação Padronizado - SPI24		
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	0,14	0,13	0,15
Coefficiente de Eficiência	0,98	0,98	0,98

Figura 13 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 12 sem defasagem temporal de dados

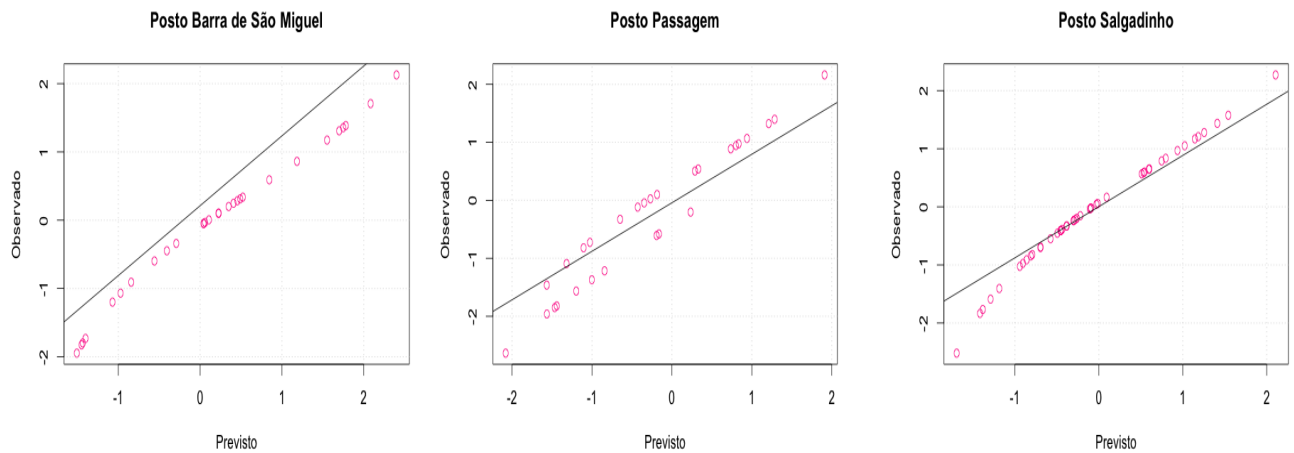


Tabela 5 - Resultados estatísticos SPI 12 sem defasagem temporal de dados

Índice de Precipitação Padronizado - SPI 12			
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	0,25	0,31	0,15
Coefficiente de Eficiência	0,95	0,93	0,98

Figura 14 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 – semestre 1 sem defasagem temporal de dados

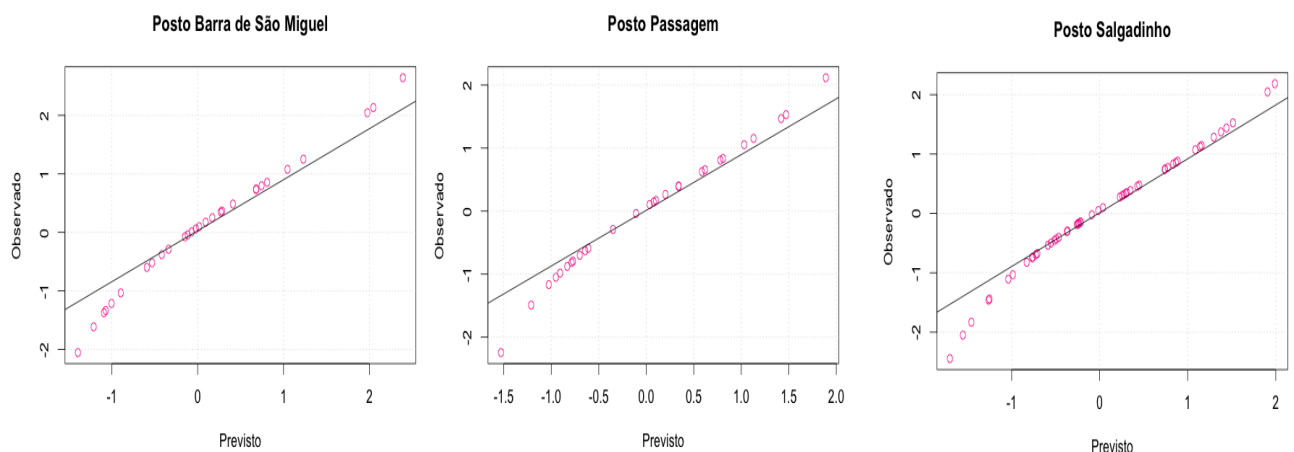


Tabela 6 - Resultados estatísticos SPI 6 – Semestre 1 sem defasagem temporal de dados

Índice de Precipitação Padronizado - SPI 6 - Semestre 1			
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	0,18	0,16	0,15
Coefficiente de Eficiência	0,97	0,97	0,98

Figura 15 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 – semestre 2 sem defasagem temporal de dados

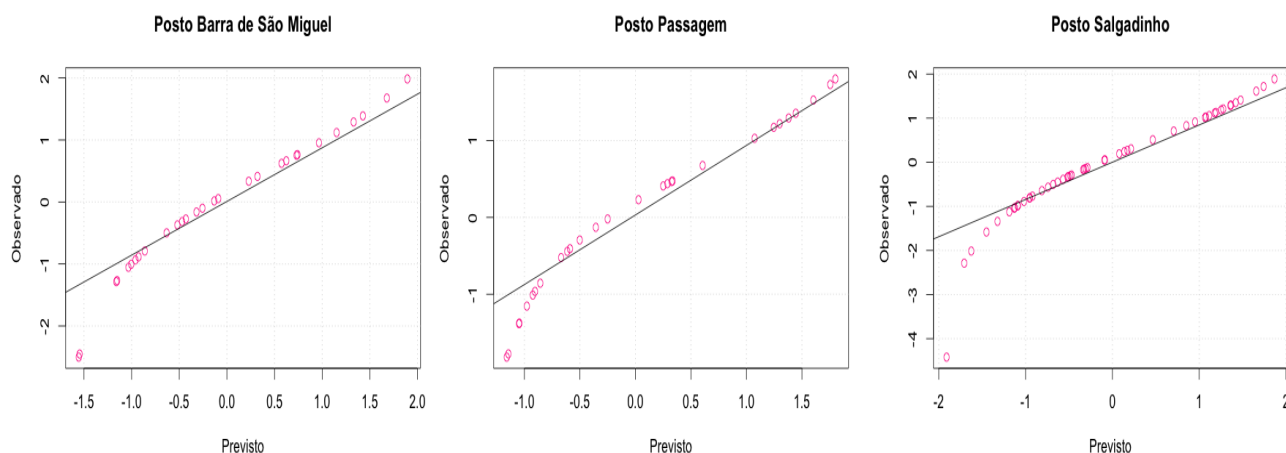


Tabela 7 - Resultados estatísticos SPI 6 – Semestre 2 sem defasagem temporal de dados

Índice de Precipitação Padronizado - SPI 6 - Semestre 2			
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	0,26	0,23	0,38
Coefficiente de Eficiência	0,94	0,95	0,89

Como demonstrado nas figuras e tabelas acima, o coeficiente de eficiência e o erro quadrático médio apresentaram valores satisfatórios para todos os postos pluviométricos em todas as escalas de tempo adotadas. Dessa forma, as redes neurais se apresentam como uma alternativa eficaz no cálculo do SPI, visto que com o treinamento de um único posto pluviométrico, é possível calcular os valores SPI de outros postos sem a necessidade de seguir os passos metodológicos de adequação a uma probabilidade gama e respectiva normalização, como descrito na seção 3.3.

Além disso, o uso da rede neural vence os obstáculos inerentes à variação espaço-temporal que os postos apresentam, já que a rede é treinada para o maior intervalo de precipitação acumulada possível, levando em conta a série histórica de dados pluviométricos da região da bacia hidrográfica estudada.

É importante ressaltar que, as dispersões que por ventura ocorrem entre os valores previstos e observados de SPI, como ilustrado na Figura 13, podem ter como explicação a ausência de uma sequência de dados extensa o suficiente para adequar-se à rede neural adotada.

Com relação aos valores encontrados a partir das RNAS treinadas com a defasagem temporal de dados, os resultados estatísticos são apresentados na Tabela 8.

Tabela 8 - Resultados estatísticos das redes neurais artificiais (RNAs) com defasagem temporal de dados para os índices de precipitação padronizado (SPI)

		Treinamento e Validação		Teste	
		Erro Quadrático Médio	Coefficiente de Eficiência	Erro Quadrático Médio	Coefficiente de Eficiência
SPI 24	RNA-5	0,97	0,00	1,24	-0,02
	RNA-10	-0,28	-0,55	1,17	0,09
	RNA-15	-0,18	-0,56	1,17	0,08
	RNA-20	1,24	-0,64	1,17	0,08
SPI 12	RNA-5	0,98	0,00	1,41	-1,06
	RNA-10	0,98	0,00	1,41	-1,06
	RNA-15	0,98	0,00	1,41	-1,06
	RNA-20	0,97	0,02	1,41	-1,05
SPI 6	RNA-5	1,07	0,00	1,21	-0,16
	RNA-10	1,07	0,00	1,21	-0,16
	RNA-15	1,08	-0,01	1,21	-0,15
	RNA-20	1,08	-0,01	1,21	-0,15

Observa-se, através das análises estatísticas mostradas na Tabela 8 que, em todos os casos as RNAs não foram capazes de apresentar uma boa modelagem. Seus baixos coeficientes de eficiência e altos valores de erro quadrático médio indicam que não é possível prever valores de SPI a partir de dados temporais defasados, utilizando as arquiteturas e algoritmo apresentados nesse trabalho. Observa-se pouca ou nenhuma diferença nos resultados estatísticos quando da variação de neurônios na camada intermediária.

As Figuras 16, 17 e 18 mostram os resultados obtidos através da aplicação das redes neurais artificiais com 10 neurônios na camada intermediária e com defasagem temporal de dados para os três postos pluviométricos em questão.

Figura 16 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 24 com defasagem temporal de dados

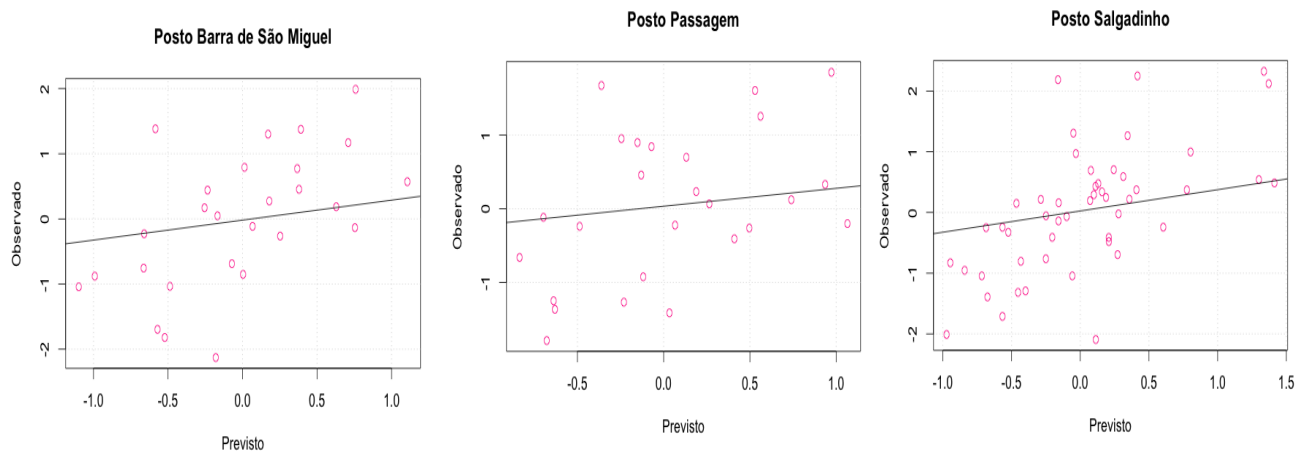


Tabela 9 - Resultados estatísticos SPI 24 com defasagem temporal de dados

SPI 24			
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	0,84	0,89	0,80
Coefficiente de Eficiência	0,32	0,19	0,39

Figura 17 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 12 com defasagem temporal de dados

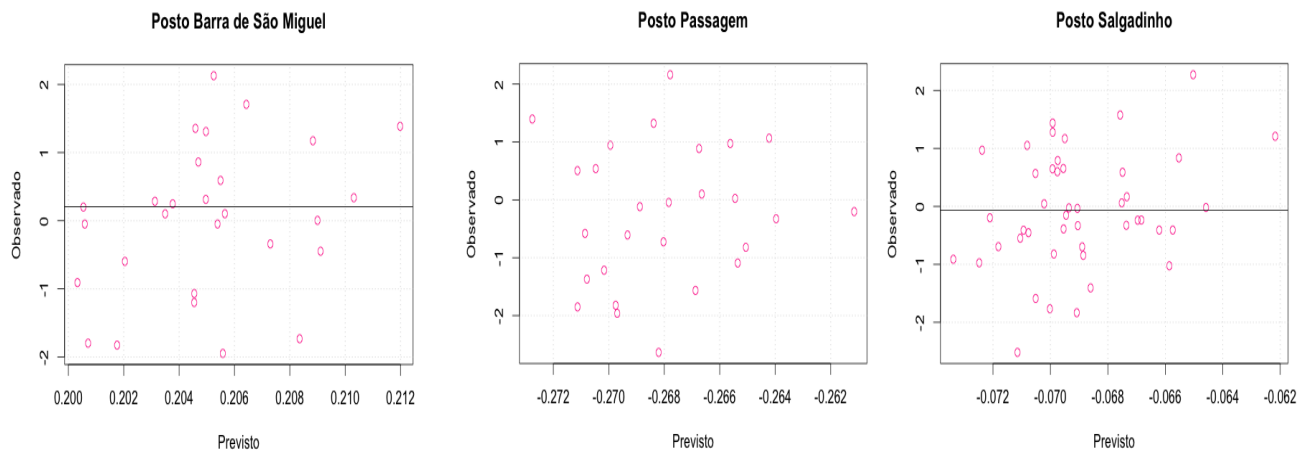


Tabela 10 - Resultados estatísticos SPI 12 com defasagem temporal de dados

SPI 12			
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	1,11	1,18	0,98
Coefficiente de Eficiência	-0,03	0,00	0,00

Figura 18 - Resultados obtidos por meio do uso das redes neurais artificiais (RNAs) selecionadas para o SPI 6 com defasagem temporal de dados

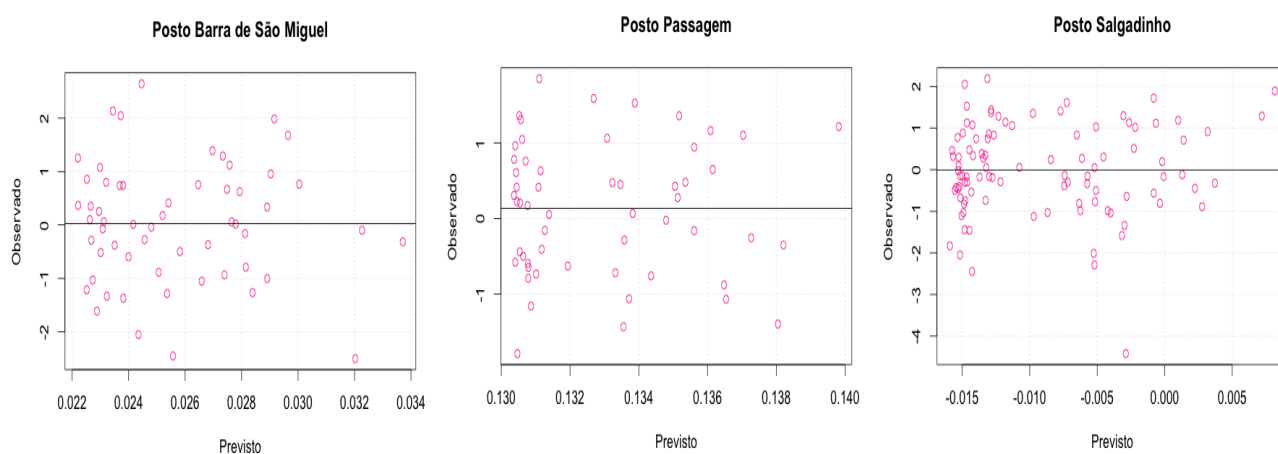


Tabela 11 - Resultados estatísticos SPI 6 com defasagem temporal de dados

	SPI 6		
	Barra de São Miguel	Passagem	Salgadinho
Erro Quadrático Médio	1,11	0,87	1,09
Coefficiente de Eficiência	0,00	0,00	0,00

Como demonstrado nas figuras e tabelas acima, a grande dispersão entre os valores previstos e observados, e os baixos valores dos coeficientes de eficiência corroboram com o discutido acima: não é possível prever através das redes neurais artificiais adotadas nesse estudo o valor SPI futuro a partir de valores de precipitação acumulada de um período temporal anterior ao mesmo. A grande dispersão entre os valores e a incapacidade de previsão, pode ser observada inclusive no Posto Salgadinho, o mesmo usado para o treinamento das RNAs.

Assim, em posse do acumulado de precipitação pluviométrico do primeiro semestre de uma ano qualquer, por exemplo, não é possível prever o SPI do próximo semestre. O mesmo ocorre para valores anuais ou bienais. Isso pode ser explicado pelo fato de que não existe correlação natural entre a precipitação acumulada de um período anterior com o seu próximo.

5. CONCLUSÃO E RECOMENDAÇÕES

Com base nos resultados deste trabalho, pode-se concluir que:

- Durante a escolha dos tipos de arquitetura, a rede neural do tipo *backpropagation* com 10 neurônios na camada intermediária apresentou bom desempenho como método alternativo para o cálculo dos SPI 24, 12 e 6.
- As redes neurais se apresentam como uma alternativa eficaz no cálculo do SPI, visto que com o treinamento de um único posto pluviométrico, é possível calcular os valores SPI de outros postos sem a necessidade de seguir os passos metodológicos de adequação a uma probabilidade gama e respectiva normalização, como descrito por Mckee et al. (1993).
- O uso da rede neural artificial vence os obstáculos inerentes à variação espaço-temporal que os postos apresentam, já que a rede é treinada para o maior intervalo de precipitação acumulada possível da série histórica de dados pluviométricos pertencente à bacia hidrográfica estudada.
- A construção das RNAs com defasagem temporal de dados mostrou-se incapaz de prever secas;
- A grande dispersão entre os valores previstos e esperados, associadas aos valores estatísticos que indicam uma baixa previsibilidade do modelo, mostram que a previsão de secas se torna inadequada com o uso do modelo apresentado. No entanto, vale ressaltar que a imprevisibilidade demonstrada nesse estudo pode ser explicada pelo fato de que não existe correlação natural entre a precipitação acumulada de um período anterior com o seu próximo.

Recomendações para trabalhos futuros:

- Investigar redes neurais artificiais com mais de um dado de entrada, a fim de avaliar o aprimoramento da rede na previsão de secas;
- Investigar a combinação das redes neurais para previsão com outros algoritmos de propagação e/ou número de neurônios diferentes, ou mesmo, outras funções de transferência.
- Investigar a previsibilidade das redes tendo em vista a análise dos valores de SPI analisados mês a mês.

REFERÊNCIAS BIBLIOGRÁFICAS

ANA – Agência Nacional das Águas. **Monitor de Secas** . 2016. Disponível em: < <http://monitordesecas.ana.gov.br/>>. Acesso em: 07/11/2016

ANA – Agência Nacional de Águas. **Hidroweb – Sistema de informações hidrológicas**. Disponível em: < <http://hidroweb.ana.gov.br/default.asp>> Acesso em 10 Janeiro. 2017.

ANDREOLI, R. V.& KAYANO, M. T. **A importância relativa do atlântico tropical sul e pacífico leste na variabilidade de precipitação do nordeste do Brasil**. Revista. Brasileira de Meteorologia, São Paulo, v. 22, n. 1, p. 63-74, <<http://dx.doi.org/10.1590/S0102-77862007000100007>>. 2007

ARCS, 2007. **Drought monitoring using vegetation and LST indices in Nepal and Northeastern India**. 2007. Disponível em: <<http://www.a-a-r-.org/acrs/proceeding/ACRS2007/>>. Acesso em 07/11/2016

ASA. Articulação Semiárido Brasileiro. Disponível em http://www.asabrazil.org.br/Portal/Informacoes.asp?COD_MENU=97. Acesso em 11/11/2016

BAFFA, A. C. E. et al. 2008. **Time series prediction using artificial neural networks**. 5th International Conference on Information Systems and Technology Management. São Paulo. 2008

BATISTA, W. J. **Classificação e avaliação dos fatores de ocorrência de secas na bacia do rio Guandu**. Tese de Doutorado– Espírito Santo. 2012

BLAIN, G. C.& BRUNINI, O. **Avaliação e adaptação do Índice de Severidade de Seca de Palmer (PDSI) e do Índice Padronizado de Precipitação (SPI) às condições climáticas do Estado de São Paulo**. Bragantia, Campinas , v. 64, n. 4, p. 695-705. <http://dx.doi.org/10.1590/S0006-87052005000400020>. 2005

BRAGA, A.; CARVALHO, A.; LUDERMIR, T. **Redes neurais artificiais, teorias e aplicações**. Livros Técnicos e Científicos. Editora S.A. Rio de Janeiro. 2000

BRITO, F. B. de. Conflitos pelo uso da água do açude Epitácio Pessoa (Boqueirão) – PB. Programa de Pós-Graduação em Geografia - PPGG, Dissertação de Mestrado, UFPB, João Pessoa, 2008. 176 p

CANAMARY, E. A. **Avaliação de índices para fins de monitoramento e previsão de secas no Nordeste setentrional**. 2015. Universidade de Brasília (UNB). Dissertação de Mestrado . 2015

CARRÃO, H.; NAUMANN,G.;BARBOSA, P. **Mapping global patterns of drought risk: an empirical framework based on sub-national estimates of hazard, exposure and vulnerability**. Global Environmental Change Journal. Volume 39, July 2016, Pages 108–124. 2016

CEC. Commission to the European Parliament and the Council. **Impact assessment.** Brussels, Belgica (2007)

CRED – Centre for Research on the Epidemiology of Disasters. **The human cost of natural disasters – A Global perspective.** 2015

DANTAS, I. M. A. **Índice de reconhecimento de secas usando a evapotranspiração potencial diária em região semiárida.** Universidade Federal de Campina Grande (UFCG). Dissertação de Mestrado. 2015

DASH, B.K.; RAFIUDDIN, M.; KHANAM, F. et al. Nat Hazards (2012). **Characteristics of meteorological drought in Bangladesh.** Natural Hazards Journal, Volume 64, Issue 2, pp 1461– doi:10.1007/s11069-012-0307-1. 2012

DAYAL, K.; DEO R.; APAN, A.; ARMANDO A. **Drought modelling based on artificial intelligence and neural network algorithms: a case study in Queensland, Australia.** Climate Change Adaptation in Pacific Countries, pp 177-198, Fevereiro de 2017. Springer International Publishing. – doi: 10.1007/978-3-319-50094-2_11. 2017

DOGAN, Selim. **Extreme droughts in Sydney And Melbourne since the 1850s.** EGU General Assembly(2014). Vienna, Austria, id.911. 2014

EUROPEAN COMISSION, CEC - **Impact assessment.** Coordinating European Council, Brussels, Belgium (2010)

FERNANDES, L. G. L.; PORTUGAL, M. S.; NAVAUX, P. O. A. **Previsão de séries de tempo: redes neurais artificiais e modelos estruturais** (1995). Disponível em: < <http://www.inf.pucrs.br/gmap/files/publications/articles/5afca684be812ed8e2a901173d20e6b6.pdf>> Acesso em março de 2017.

FERREIRA, T. R. et al. **Preliminary analysis of droughts in 2012 in semi-arid of Alagoas using indices of vegetation through sensor modis.** Journal of Hyperspectral Remote Sensing, V. 5, n.1 (2015).

FREITAS, M. A. S. **Previsão de secas por meio de métodos estatísticos e redes neurais e análise de suas características através de diversos índices.** Institut für Wasserwirtschaft, Hydrologie und landwirtschaftlichen Wasserbau, Universität Hannover, Deutschland. 2000. Disponível em: < <http://documentslide.com/documents/previsao-de-secas-redes-neurais.html>> Acesso em:11/03/2017

FUNK, C.; HOELL, A.; SHUKLA, S.; BLADÉ, I.; LIEBMANN, B.; ROBERTS, J. B.; ROBERTSON, F. R.; and HUSAK, G.: **Predicting East African spring droughts using Pacific and Indian Ocean sea surface temperature indices,** Hydrology and Earth System Journal . Science., 18, 4965-4978, doi:10.5194/hess-18-4965-2014, 2014.

GIANNINI, A.; SARAVANAN, R.; CHANG, P. **The preconditioning role of Tropical Atlantic Variability in the development of the ENSO teleconnection: implications for the prediction of Nordeste rainfall.** Climate. Dynamics Journal., v.22, p. 839-855, doi: 10.1007/s00382- 004-0420-2, 2004

GUEDES, R. V. S. **Previsão de seca na bacia hidrográfica do Açude Epitácio Pessoa com base em redes neurais artificiais**. Campina Grande. Dissertação de Mestrado – Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais. 2011

GUIMARAES, I. A. & CHAVES NETO, A. **Reconhecimento de padrões: metodologias estatísticas em crédito ao consumidor**. RAE electron., São Paulo, v. 1, n. 2, p. 02-14, 2002.

HAYES, M. J.; SVOBODA, M. D.; WILHITE, D. A.; VANYARKHO, O. V. **Monitoring the 1996 drought using the standardized precipitation index**. Bulletin of the American Meteorological Society, 80(3), 429–438. doi:10.1175/1520-0477.0802.0.CO;2. 1999

HEDDAM, S. **Multilayer perceptron neural network-based approach for modeling phycocyanin pigment concentrations: case study from lower Charles River buoy, USA**. Environmental Science and Pollution Research 23: 17210. doi:10.1007/s11356-016-6905-9. 2016

HEIM JUNIOR, R. R. **A review of twentieth - century drought indices used in the United States**. Bulletin of American Meteorological Society Journal. 83(8): p.1149-1163, 2002

HOELL, A.; FUNK, C. ; BARLOW, M. **La Niña diversity and Northwest Indian Ocean Rim teleconnections** Climate Dynamics Volume 43, Issue 9, pp 2707–2724 doi:10.1007/s00382-013-1799-4. 2014

IBGE – Instituto Brasileiro de Geografia e Estatística, **Base de Dados Demográficos**. Acesso em dezembro de 2016.

LOPES, H. & BARBOSA, S.. **Data science com R**. Departamento de Informática – PUC-Rio. 2014. Disponível em: <http://www-di.inf.puc-rio.br/~lopes/inf2391/Data_Science_Learning2.pdf>. Acesso em: 02/04/2017

MARATHE, K. & SEIGFRIED, D. **National drought policy guidebook**. Journal of Water Research and Management, Vol. 3, No. 3 pages 3-12. (2013)

MARTINEZ, P. H. **Vida e morte no sertão: história das secas no Nordeste nos séculos XIX e XX**. Revista Brasileira de História, São Paulo , v. 22, n. 43, p. 251-254, doi: 10.1590/S0102-01882002000100016. 2002 .

MCKEE, T. B.; DOESKEN, N. J.; KLEIST, J. **The relationship of drought frequency and duration of time scales**. In Eighth Conference on Applied Climatology (p. 179–186). Anaheim CA: American Meteorological Society. 1993

MORAIS, B. S; SILVA, G. B. L; SANTOS, C. A. G.. **Previsão de seca usando redes neurais artificiais**. IX Simpósio de Recursos Hídricos do Nordeste. Salvador 2012

NOAA National Centers for Environmental Information (NCEI) **U.S. Billion dollar weather and climate disasters (2017)**. Disponível em: <<https://www.ncdc.noaa.gov/billions/>>. Acesso em Janeiro de 2017

NOBREGA, R. S. & SANTIAGO, G. A. C. F. **Tendência de temperatura na superfície do mar nos Oceanos Atlântico e Pacífico e variabilidade de precipitação e Pernambuco**.

Revista Mercator (Fortaleza), Fortaleza , v. 13,n. 1,p. 107-118, Apr. 2014 Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-22012014000100107&lng=en&nrm=iso>. Acesso em 18 de abril de 2017.

NÓBREGA,R.S; SANTIAGO,G.A. C.F.; SOARES D. B. **Tendências do controle climático oceânico sob a variabilidade temporal da precipitação no Nordeste do Brasil**. Revista Brasileira de Climatologia. Ano 12 –Vol. 18 –JAN/JUN 2016

PALMER, W.C. **Meteorological drought**. US Weather Bureau Res. Nº 45 . Washington, 58p, 1965

PARTAL, T.; CIGIZOGLU, H.K. ; KAHYA. **Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data**. Stochastic Environmental Research and Risk Assessment Journal. Volume 29, Issue 5, pp 1317–1329. doi:10.1007/s00477-015-1061-1. 2015

PAULO, A. M. A. **Metodologias de caracterização e previsão das secas – Modelação espacial e temporal**. Tese de Doutorado. Instituto Superior de Agronomia para a obtenção do grau de Doutor em Engenharia Agronómica. 254 p. 2004

PONTES, E. M & MACHADO, T, A. **Desenvolvimento sustentável e convivência com o semi-Árido: o caso do programa de um milhão de cisternas rurais no Nordeste brasileiro**. RAS – Rede de Agricultura Sustentável. 2008. Disponível em: <<http://www.agrisustentavel.com/doc/p1mc.pdf>>. Acesso em : 07/11/2016

PONTES, E. T. M. **A convivência com o semiárido no contexto sulamericano: segurança hídrica em Afogados da ingazeira (Pernambuco, Brasil) e Graneros (Tucumán, Argentina)**. Tese de doutorado – Universidade Federal de Pernambuco. 2014

PORTELA, M. M. et al. **Application of the standardized precipitation index (SPI) to the establishment of surfaces of precipitation thresholds for drought recognition**. International Journal of Environmental Engineering and Natural Resources. Volume 1, 1, pp. 19-35. 2014

RAZMKHAH, H.; ROSTAMI E.; FATTTAHI M.H. **Drought forecasting using artificial neural network: case study: Kohgiloooye and Boyer Ahmad**. Arabian Journal of Geosciences. DOI: 10.13140/RG.2.1.1378.7605. 2015

ROSA, R. G. T. D. **Índices de seca. Aplicação ao continente português**. Dissertação de Mestrado. Instituto Superior de Agronomia. Lisboa: ISA, 2011

SANTOS, J. F. & PORTELA, M. M. **Caracterização de secas em bacias hidrográficas de Portugal Continental: aplicação do índice de precipitação padronizada, SPI, a séries de precipitação e de escoamento**. 10º Congresso da Água, Associação Portuguesa dos recursos Hídricos, APRH, Alvor (Portugal). 2010

SILVA FGCS & PACHECO JS. **Processo de desertificação: estudo de caso em Irauçuba-CE**. Revista Eletrônica Teccen. 09 (1): 47-51, Jan/Jun 2016

SILVA, V. A.; MARCELINO P. M.; DE A. RIBEIRO, V.; MAINAR M. R. **O desastre seca no Nordeste Brasileiro**. Revista Polêmica. 12, jun. 2013. Disponível em: <<http://www.e-publicacoes.uerj.br/index.php/polemica/article/view/6431/4857>>. Acesso em: 11/11/2016

SYOBODA. M; Hayes M.;WOOD D. **Standardized precipitation index user guide**. World Meteorological Organization (WMO) No. 1090), Geneva. 2012

TAYFUR, G. **Soft computing in water resources engineering**. WIT Press. ISBN: 978-1-84564-636-3. Southampton, Boston, 2012

THOM, H. C. S. **Some methods of climatological analyses**. World Meteorological Organization.(WMO), Geneva, 53p, 1966; TP, 103; Technical note.81. 1966

TISSOT, H. C.; CAMARGO, L. C.; POZO, A. T. R. **Treinamento de redes neurais feedforward: comparativo dos algoritmos backpropagation e differential evolution**. IX Encontro Nacional de Inteligência Artificial (ENIA). 2012

VALIPOUR M.; BANIHABIB M. E.; SEYYED M. R. B. **Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir**, Journal of Hydrology, Volume 476, 7 January 2013, Pages 433-441, ISSN 0022-1694, <http://doi.org/10.1016/j.jhydrol.2012.11.017>. 2013

VILLA, M. A. **Vida e morte no sertão. História das secas no Nordeste nos séculos XIX e XX**. São Paulo. Editora Ática, 2001.

VILLANUEVA, W. J. P. **Síntese automática de redes neurais artificiais com conexões à frente arbitrárias** . Tese de Doutorado. Universidade Estadual de Campinas-Campinas, SP: 2011.

WILHITE, D. A. & GLANTZ, M. H. **Understanding the drought phenomenon: the role of definitions**. Drought Mitigation Center Faculty Publications. Paper 20. Water International Journal. 10:3 (1985), pp. 111–120. 1985

WMO - World Meteorological Organization. **High-level meeting on national drought policy - meeting minutes**., Geneva, Switzerland. 2013

WMO, World Meteorological Organization. **Drought monitoring and early warning: concepts, progress and future challenges**. WMO no. 1006, Geneva, Switzerland. 2006

APÊNDICE A – Pesos e vieses iniciais

Tabela 12.1 - Pesos e vieses iniciais para RNAs 5

RNA 5	
b1	-0,01
v1	0,01
b2	-0,02
v2	0,02
b3	-0,03
v3	0,03
b4	-0,04
v4	0,04
b5	-0,05
v5	0,05
b6	-0,01
w1	0,01
w2	0,02
w3	0,03
w4	0,04
w5	0,05

Tabela 13.2 - Pesos e vieses iniciais para RNAs 10

RNA 10					
b1	-0,01	b8	-0,02	w8	0,2
v1	0,01	v8	0,02	w9	0,3
b2	-0,02	b9	-0,03	w10	0,4
v2	0,02	v9	0,03		
b3	-0,03	b10	-0,04		
v3	0,03	v10	0,04		
b4	-0,04	b11	-0,01		
v4	0,04	w1	0,1		
b5	-0,05	w2	0,2		
v5	0,05	w3	0,3		
b6	-0,06	w4	0,4		
v6	0,06	w5	0,5		
b7	-0,01	w6	0,6		
v7	0,01	w7	0,1		

Tabela 14.3 - Pesos e vieses iniciais para RNAs 15

RNA 15							
b1	-0,01	b8	-0,08	b15	-0,35	w12	0,65
v1	0,01	v8	0,08	v15	0,35	w13	0,7
b2	-0,02	b9	-0,09	b16	-0,1	w14	0,75
v2	0,02	v9	0,09	w1	0,1	w15	0,8
b3	-0,03	b10	-0,1	w2	0,15		
v3	0,03	v10	0,1	w3	0,2		
b4	-0,04	b11	-0,15	w4	0,25		
v4	0,04	v11	0,15	w5	0,3		
b5	-0,05	b12	-0,2	w6	0,35		
v5	0,05	v12	0,2	w7	0,4		
b6	-0,06	b13	-0,25	w8	0,45		
v6	0,06	v13	0,25	w9	0,5		
b7	-0,07	b14	-0,3	w10	0,55		
v7	0,07	v14	0,3	w11	0,6		

Tabela 15.4 - Pesos e vieses iniciais para RNAs 20

RNA 20									
b1	-0,01	b8	-0,08	b15	-0,35	w2	0,15	w16	0,85
v1	0,01	v8	0,08	v15	0,35	w3	0,2	w17	0,9
b2	-0,02	b9	-0,09	b16	-0,4	w4	0,25	w18	0,1
v2	0,02	v9	0,09	v16	0,4	w5	0,3	w19	0,15
b3	-0,03	b10	-0,1	b17	-0,45	w6	0,35	w20	0,2
v3	0,03	v10	0,1	v17	0,45	w7	0,4		
b4	-0,04	b11	-0,15	b18	-0,5	w8	0,45		
v4	0,04	v11	0,15	v18	0,5	w9	0,5		
b5	-0,05	b12	-0,2	b19	-0,55	w10	0,55		
v5	0,05	v12	0,2	v19	0,55	w11	0,6		
b6	-0,06	b13	-0,25	b20	-0,6	w12	0,65		
v6	0,06	v13	0,25	v20	0,6	w13	0,7		
b7	-0,07	b14	-0,3	b21	-0,1	w14	0,75		
v7	0,07	v14	0,3	w1	0,1	w15	0,8		

APÊNDICE B – Código de construção da rede neural artificial no software R

```
#Criando uma rede neural para o SPI 12 sem defasagem de dados
# posto Salgadinho
```

```
#SPi 12 - Sem defasagem
```

```
install.packages('neuralnet')
library("neuralnet")
library("ggplot2")
```

```
# ?neuralnet para mais informacoes
```

```
#Gerar Vetores
```

```
#E armazena-los como dataframes
```

```
#Vetor a1 refere-se às entradas de precipitação normalizadas para treinamento e validação
```

```
#Vetor b1 refere-se aos valores SPI normalizados esperados
```

```
#vetor c1 refere-se aos valores SPI nao normalizados
```

```
a1<- c(0.352998249,    0.590072147,    0.245608335,    0.1,    0.259343108,    0.325130612,
      0.49253541,    0.336974354,    0.51831303,    0.349713828,    0.493630236,    0.411520271,
      0.202313997,    0.286912161,    0.183105241,    0.355884532,    0.362353806,    0.381562579,
      0.380169181,    0.158820593,    0.251878571,    0.306120917,    0.392411022,    0.286414526,
      0.478601642,    0.673177371, 0.382557836,    0.336178136,    0.483976103,    0.482184632,
      0.394899209,    0.267603882,    0.567180841,    0.3254292,    0.644115438,    0.830131863,
      0.597835279,    0.325528742,    0.611769071,    0.526673275,    0.350211486)
b1<- c(0.463066388,    0.681566569,    0.335521226,    0.1,    0.353364817,    0.432349244,
      0.599169461,    0.445582099,    0.621803189,    0.459520681,    0.600144452,    0.523275286,
      0.275469072,    0.387682262,    0.246661763,    0.466166457,    0.47306192,    0.49311792,
      0.491683566,    0.207932171,    0.343732712,    0.410524084,    0.504182986,    0.387079348,
      0.586650569,    0.745693956,    0.494140508,    0.444700928,    0.591503738,    0.589889475,
      0.506697224, 0.363849659,    0.66298553, 0.432686205,    0.723824819,    0.85517991,
      0.6877722,    0.4327985,    0.698793453,    0.629003674,    0.460059169)
c1<- (((b1-0.1) *(5.073048))/0.8)+(-2.51862)
```

```
#vetor a2 refere-se as entradas de precipitacao normalizadas para a fase de teste
```

```
#vetor b2 refere-se aos valores de saída normalizados esperados - fase de teste
```

```
#vetor c2 refere-se aos valores de saida nao noralizados esperados - fase de teste
```

```
a2 <- c(0.328215972,    0.270589681,    0.319158982,    0.16528986,    0.551455543)
b2<-c(0.435822656,    0.367595629,    0.425572876,    0.218524967,    0.649966587)
c2<- (((b2-0.1) *(5.073048))/0.8)+(-2.51862)
```

```
#vetor weights refere-se aos pesos iniciais adotados
```

```
weights.5<- c(-0.01,    0.01,    -0.02,    0.02,    -0.03,    0.03,    -0.04,    0.04    ,-0.05    ,0.05    ,-0.01
      ,0.01,    0.02    ,0.03    ,0.04    ,0.05)
```

```
#Criando tabelas a partir dos vetores criados
```

```
traininginput.5<- as.data.frame(a1)
```

```
trainingoutput.5 <- as.data.frame(b1)
```

```
SaidaSPI.5<- as.data.frame(c1)
```

```
testdata.5<- as.data.frame(a2)
```

```
testoutput.5<- as.data.frame(b2)
```

```
testSPI.5<-as.data.frame(c2)
```

```
testSPI.5<-as.data.frame(c2)
```

```
#Combinando as colunas em uma variavel
```

```
trainingdata.5 <- cbind(traininginput.5,trainingoutput.5)
```

```
colnames(trainingdata.5) <- c("Entrada", "Saida")
```

```

#Usar validacao cruzada K-fold nos dados de treinamento e validacao a fim de evitar over-training
#randomizar os dados de forma aleatória
trainingdata.5<-trainingdata.5[sample(nrow(trainingdata.5)),]

#Criar 10 partições de tamanhos iguais
folds.5 <- cut(seq(1,nrow(trainingdata.5)),breaks=10,labels=FALSE)

#Para um learning rate de 0.0001
#10 fold cross validation

for(i in 1:10){
  #segmentar os dados por fold usando a funcao which()
  testIndexes.5 <- which(folds.5==i,arr.ind=TRUE)
  dadovalidacao.5<- trainingdata.5[testIndexes.5, ]
  dadotreinamento.5 <- trainingdata.5[-testIndexes.5, ]

  #Treinando a rede neural
  #Com 5 neuronios na camada intermediaria

  #*Threshold is a numeric value specifying the threshold for the partial derivatives of the error function as
  #stopping criteria.*
  nn.5 <- neuralnet(Saida~Entrada,dadotreinamento.5, hidden=5, stepmax = 1000000, startweights = weights.5,
  learningrate = 0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
  threshold = 0.01,linear.output = FALSE, constant.weights = NULL)
  print(nn.5)
  #Plotar a Rede Neural
  plot(nn.5)
  #Criar uma tabela com os valores dos pesos de cada K-fold
  valores.5<- as.data.frame(nn.5$result.matrix)
  assign(paste("X",i,sep=""),valores.5)

}
#Combinar as tabelas em uma so
vf.5 <- cbind(X1,X2,X3,X4,X5,X6,X7,X8,X9,X10)
colnames(vf.5) <- c("K-fold1","K-fold2","K-fold3","K-fold4","K-fold5","K-fold6","K-fold7","K-fold8","K-
fold9","K-fold10")
#Calcular a media dos pesos
mean.5 <- as.data.frame(apply(vf.5,1,mean))

valoresfinais.5<- cbind(vf.5,mean.5)
pesos.5<-as.vector(valoresfinais.5[4:19,11])
print(pesos.5)

#Rede neural com o erro minimizado e utilizando a media dos pesos dos k-folds
NN.5<-neuralnet(Saida~Entrada, trainingdata.5, hidden=5, 1000000, startweights = pesos.5, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.5)
print(NN.5)

print(NN.5$result.matrix)
plot (NN.5)

saidaredeneural.5 <- as.data.frame(NN.5$net.result)
Saidaredeneural.5<- as.data.frame((((saidaredeneural.5 -0.1) *(5.073048))/0.8)+(-2.51862))
saida.5 <- cbind (SaidaSPI.5, Saidaredeneural.5)
colnames(saida.5)<- c("Observado", "Previsto")

```

```

#tabela comparando os valores observados e previstos

treinamento.5<-cbind(traininginput.5, SaidaSPI.5, Saidaredeneural.5)
colnames(treinamento.5) <- c("Entrada","Observado", "Previsto")

#Plotar os valores observados e previstos e calcular seu coeficiente de correlacao

plot(saida.5$Previsto,saida.5$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Previsto x Observado")
grid()
R2.5<-lm(Previsto~Observado,data=saida.5)
summary(R2.5)
summary(R2.5)$r.squared
abline(R2.5)

# Funcao que retorna o erro medio quadraticoFunction
rmse <- function(error.5)
{
  sqrt(mean(error.5 ^2))
}

# Calculate error
error.5 <- SaidaSPI.5 - Saidaredeneural.5
rmse(error.5)
rmsetreino.5<- as.data.frame(rmse(error.5))

#Coeficiente de eficiencia (E)

x.5<- (error.5)^2
xm.5<- mean(c1)
y.5<-(SaidaSPI.10 - xm.5)^2
print(x.5)
E.5 = 1 - (sum(x.5)/sum(y.5))
print(E.5)
Etreino.5<-as.data.frame(E.5)

#Plotar valores para visualizacao
ggplot(treinamento.5, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#Testar a rede neural nos dados de teste
nnteste.5<- compute(NN.5, testdata.5)
#Observar suas propriedades
ls(nnteste.5)

#Observando os resultados...
print(nnteste.5$net.result)

SaidadeTeste.5 <- as.data.frame(nnteste.5$net.result)

# Calculando RMSE para o teste
error.teste.5<- testSPI.5- SaidadeTeste.5
rmse(error.teste.5)
rmseteste.5<- as.data.frame(rmse(error.teste.5))
(etc...)

```

#Coeficiente de eficiencia (E) para o teste

```
x.5.t<- (error.teste.5)^2
xm.5.t<- mean(c2)
y.5.t<-(testSPI.5 - xm.5.t)^2
print(x.5.t)
E.5.t = 1 - (sum(x.5.t)/sum(y.5.t))
print(E.5.t)
Eteste.5<-as.data.frame(E.5.t)
```

#Criar uma tabela com o RMSE e E para fases de treinamento e teste

```
desempenho.5<- cbind(rmseetreino.5,rmseteste.5, Etreino.5, Eteste.5)
colnames(desempenho.5)<- c("rmse-5/ treinamento", "rmse-5/teste", "E -5/ treinamento", "E - 5/ teste")
```

#Criando uma versao melhor para visualizacao

```
previsao.5<- cbind(testdata.5,testoutput.5,SaidadeTeste.5)
colnames(previsao.5) <- c("Entrada","Observado","Previsto")
print(previsao.5)
```

#Criando uma versao melhor para visualizacao

```
previsao.5<- cbind(testdata.5,testoutput.5,SaidadeTeste.5)
colnames(previsao.5) <- c("Entrada","Observado","Previsto")
print(previsao.5)
```

```
ggplot(previsao.5, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")
#-----
```

#SPi 12 - Sem defasagem

#10 neuronios

?neuralnet para mais informacoes

#vetor weights refere-se aos pesos iniciais adotados

```
weights.10<- c(-0.01      ,0.01      ,      -0.02      ,0.02      , -0.03      ,0.03      , -0.04      ,0.04
               ,-0.05      ,0.05      , -0.06      ,0.06      , -0.01      ,0.01      , -0.02      ,0.02      , -0.03      ,0.03      , -0.04      ,0.04
               ,-0.01      ,0.1       ,0.2       ,0.3       ,0.4       ,0.5       ,0.6       ,0.1       ,0.2       ,0.3       ,0.4
               )
```

#Criando tabelas a partir dos vetores criados

```
traininginput.10<- as.data.frame(a1)
trainingoutput.10 <- as.data.frame(b1)
SaidaSPI.10<- as.data.frame (c1)
```

```
testdata.10<- as.data.frame(a2)
testoutput.10<- as.data.frame (b2)
testSPI.10<-as.data.frame(c2)
```

#Combinando as colunas em uma variavel

```
trainingdata.10 <- cbind(traininginput.10,trainingoutput.10)
colnames(trainingdata.10) <- c("Entrada","Saida")
```

#Usar validacao cruzada K-fold nos dados de treinamento e validacao a fim de evitar over-training

```

#randomizar os dados de forma aleatória
trainingdata.10<-trainingdata.10[sample(nrow(trainingdata.10)),]

#Criar 10 partições de tamanhos iguais
folds.10 <- cut(seq(1,nrow(trainingdata.10)),breaks=10,labels=FALSE)

#Para um learning rate de 0.0001
#10 fold cross validation

for(i in 1:10){
  #segmentar os dados por fold usando a funcao which()
  testIndexes.10 <- which(folds.10==i,arr.ind=TRUE)
  dadovalidacao.10<- trainingdata.10[testIndexes.10, ]
  dadotreinamento.10 <- trainingdata.10[-testIndexes.10, ]

  #Treinando a rede neural
  #Com 10 neuronios na camada intermediaria

  #*Threshold is a numeric value specifying the threshold for the partial derivatives of the error function as
  #stopping criteria.*
  nn.10 <- neuralnet(Saida~Entrada,dadotreinamento.10, hidden=10, stepmax = 1000000, startweights =
  weights.10, learningrate = 0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
  threshold = 0.01,linear.output = FALSE, constant.weights = NULL)
  print(nn.10)
  #Plotar a Rede Neural
  plot(nn.10)
  #Criar uma tabela com os valores dos pesos de cada K-fold
  valores.10<- as.data.frame(nn.10$result.matrix)
  assign(paste("Y",i,sep=""),valores.10)
}
#Combinar as tabelas em uma so
vf.10 <- cbind(Y1,Y2,Y3,Y4,Y5,Y6,Y7,Y8,Y9,Y10)
colnames(vf.10) <- c("K-fold1","K-fold2","K-fold3","K-fold4","K-fold5","K-fold6","K-fold7","K-fold8","K-
fold9","K-fold10")
#Calcular a media dos pesos
mean.10 <- as.data.frame(apply(vf.10,1,mean))

valoresfinais.10<- cbind(vf.10,mean.10)
pesos.10<-as.vector(valoresfinais.10[4:34,11])
print(pesos.10)

#Rede neural com o erro minimizado e utilizando a media dos pesos dos k-folds
NN.10<-neuralnet(Saida~Entrada, trainingdata.10, hidden=10, 1000000, startweights = pesos.10, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.10)
print(NN.10)

print(NN.10$result.matrix)
plot (NN.10)

saidaredeneural.10 <- as.data.frame(NN.10$net.result)
Saidaredeneural.10<- as.data.frame((((saidaredeneural.10 -0.1) *(5.073048))/0.8)+(-2.51862))
saida.10 <- cbind (SaidaSPI.10, Saidaredeneural.10)
colnames(saida.10)<- c("Observado", "Previsto")

#tabela comparando os valores observados e previstos

```

```

treinamento.10<-cbind(traininginput.10, SaidaSPI.10, Saidaredeneural.10)
colnames(treinamento.10) <- c("Entrada", "Observado", "Previsto")

#Plotar os valores observados e previstos e calcular seu coeficiente de correlacao

plot(saida.10$Previsto,saida.10$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Previsto x Observado")
grid()
R2.10<-lm(Previsto~Observado,data=saida.10)
summary(R2.10)
summary(R2.10)$r.squared
abline(R2.10)

# Funcao que retorna o erro medio quadraticoFunction
rmse <- function(error.10)
{
  sqrt(mean(error.10 ^2))
}

# Calculate error
error.10 <- SaidaSPI.10 - Saidaredeneural.10
rmse(error.10)
rmsetreino.10<- as.data.frame(rmse(error.10))

#Coeficiente de eficiencia (E)

x.10<- (error.10)^2
xm.10<- mean(x.10)
y.10<-(SaidaSPI.10 - xm.10)^2
print(x.10)
E.10 = 1 - (sum(x.10)/sum(y.10))
print(E.10)
Etreino.10<-as.data.frame(E.10)

#Plotar valores para visualizacao
ggplot(treinamento.10, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#Testar a rede neural nos dados de teste
nnteste.10<- compute(NN.10, testdata.10)
#Observar suas propriedades
ls(nnteste.10)

#Observando os resultados...
print(nnteste.10$net.result)

SaidadeTeste.10 <- as.data.frame(nnteste.10$net.result)

# Calculando RMSE para o teste
error.teste.10 <- testSPI.10 - SaidadeTeste.10
rmse(error.teste.10)
rmseteste.10<- as.data.frame(rmse(error.teste.10))

#Coeficiente de eficiencia (E) para o teste

```

```

x.10.t<- (error.teste.10)^2
xm.10.t<- mean(c2)
y.10.t<-(testSPI.10 - xm.10.t)^2
print(x.10.t)
E.10.t = 1 - (sum(x.10.t)/sum(y.10.t))
print(E.10.t)
Eteste.10<-as.data.frame(E.10.t)

#Criar uma tabela com o RMSE e E para fases de treinamento e teste

desempenho.10<- cbind(rmseetreino.10,rmseteste.10, Etreino.10, Eteste.10)
colnames(desempenho.10)<- c("rmse-10/ treinamento", "rmse-10/teste", "E - 10/ treinamento", "E - 10/ teste")

#Criando uma versao melhor para visualizacao
previsao.10<- cbind(testdata.10,testoutput.10,SaidadeTeste.10)
colnames(previsao.10) <- c("Entrada","Observado","Previsto")
print(previsao.10)

ggplot(previsao.10, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#-----

#SPi 12 - Sem defasamento
#15 neuronios

#vetor weights refere-se aos pesos iniciais adotados

weights.15<- c(-0.01, 0.01 , -0.02 ,0.02 , -0.03 ,0.03 , -0.04 ,0.04 , -0.05 ,0.05 , -0.06
,0.06 , -0.07 ,0.07 , -0.08 ,0.08 , -0.09 ,0.09 , -0.1 ,0.1 , -0.15 ,0.15 , -0.2
,0.2 , -0.25 ,0.25 , -0.3 ,0.3 , -0.35 ,0.35 , -0.1 ,0.1 , 0.15 ,0.2 , 0.25
,0.3 , 0.35 ,0.4 , 0.45 ,0.5 , 0.55 ,0.6 , 0.65 ,0.7 , 0.75 ,0.8
)

#Criando tabelas a partir dos vetores criados
traininginput.15<- as.data.frame(a1)
trainingoutput.15 <- as.data.frame(b1)
SaidaSPI.15<- as.data.frame (c1)

testdata.15<- as.data.frame(a2)
testoutput.15<- as.data.frame (b2)
testSPI.15<-as.data.frame(c2)
#Combinando as colunas em uma variavel
trainingdata.15 <- cbind(traininginput.15,trainingoutput.15)
colnames(trainingdata.15) <- c("Entrada","Saida")

#Usar validacao cruzada K-fold nos dados de treinamento e validacao a fim de evitar over-training
#randomizar os dados de forma aleatória
trainingdata.15<-trainingdata.15[sample(nrow(trainingdata.15)),]

#Criar 10 partições de tamanhos iguais
folds.15 <- cut(seq(1,nrow(trainingdata.15)),breaks=10,labels=FALSE)

#Para um learning rate de 0.0001

```

#10 fold cross validation

```

for(i in 1:10){
  #segmentar os dados por fold usando a funcao which()
  testIndexes.15 <- which(folds.15==i,arr.ind=TRUE)
  dadovalidacao.15<- trainingdata.15[testIndexes.15, ]
  dadotreinamento.15 <- trainingdata.15[-testIndexes.15, ]

  #Treinando a rede neural
  #Com 15 neuronios na camada intermediaria

  #*Threshold is a numeric value specifying the threshold for the partial derivatives of the error function as
  #stopping criteria.*
  nn.15 <- neuralnet(Saida~Entrada,dadotreinamento.15, hidden=15, stepmax = 1000000, startweights =
  weights.15, learningrate = 0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
  threshold = 0.01,linear.output = FALSE, constant.weights = NULL)
  print(nn.15)
  #Plotar a Rede Neural
  plot(nn.15)
  #Criar uma tabela com os valores dos pesos de cada K-fold
  valores.15<- as.data.frame(nn.15$result.matrix)
  assign(paste("Z",i,sep=""),valores.15)
}
#Combinar as tabelas em uma so
vf.15 <- cbind(Z1,Z2,Z3,Z4,Z5,Z6,Z7,Z8,Z9,Z10)
colnames(vf.15) <- c("K-fold1","K-fold2","K-fold3","K-fold4","K-fold5","K-fold6","K-fold7","K-fold8","K-
fold9","K-fold10")
#Calcular a media dos pesos
mean.15 <- as.data.frame(apply(vf.15,1,mean))

valoresfinais.15<- cbind(vf.15,mean.15)
pesos.15<-as.vector(valoresfinais.15[4:49,11])
print(pesos.15)

#Rede neural com o erro minimizado e utilizando a media dos pesos dos k-folds
NN.15<-neuralnet(Saida~Entrada, trainingdata.15, hidden=15, 1000000, startweights = pesos.15, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.15)
print(NN.15)

print(NN.15$result.matrix)
plot (NN.15)

saidaredeneural.15 <- as.data.frame(NN.15$net.result)
Saidaredeneural.15<- as.data.frame((((saidaredeneural.15 -0.1) *(5.073048))/0.8)+(-2.51862))
saida.15 <- cbind (SaidaSPI.15, Saidaredeneural.15)
colnames(saida.15)<- c("Observado", "Previsto")

#tabela comparando os valores observados e previstos

treinamento.15<-cbind(traininginput.15, SaidaSPI.15, Saidaredeneural.15)
colnames(treinamento.15) <- c("Entrada","Observado", "Previsto")

#Plotar os valores observados e previstos e calcular seu coeficiente de correlacao

plot(saida.15$Previsto,saida.15$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Previsto x Observado")

```

```

grid()
R2.15<-lm(Previsto~Observado,data=saida.15)
summary(R2.15)
summary(R2.15)$r.squared
abline(R2.15)

# Funcao que retorna o erro medio quadraticoFunction
rmse <- function(error.15)
{
  sqrt(mean(error.15 ^2))
}

# Calculate error
error.15 <- SaidaSPI.15 - Saidaredeneural.15
rmse(error.15)
rmsetreino.15<- as.data.frame(rmse(error.15))

#Coeficiente de eficiencia (E)

x.15<- (error.15)^2
xm.15<- mean(c1)
y.15<-(SaidaSPI.15 - xm.15)^2
print(x.15)
E.15 = 1 - (sum(x.15)/sum(y.15))
print(E.15)
Etreino.15<-as.data.frame(E.15)

#Plotar valores para visualizacao
ggplot(treinamento.15, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#Testar a rede neural nos dados de teste
nnteste.15<- compute(NN.15, testdata.15)
#Observar suas propriedades
ls(nnteste.15)

#Observando os resultados...
print(nnteste.15$net.result)

SaidadeTeste.15 <- as.data.frame(nnteste.15$net.result)

# Calculando RMSE para o teste
error.teste.15 <- testSPI.15 - SaidadeTeste.15
rmse(error.teste.15)
rmseteste.15<- as.data.frame(rmse(error.teste.15))

#Coeficiente de eficiencia (E) para o teste

x.15.t<- (error.teste.15)^2
xm.15.t<- mean(c2)
y.15.t<-(testSPI.15 - xm.15.t)^2
print(x.15.t)
E.15.t = 1 - (sum(x.15.t)/sum(y.15.t))
print(E.15.t)
Eteste.15<-as.data.frame(E.15.t)

```

```
#Criar uma tabela com o RMSE e E para fases de treinamento e teste
```

```
desempenho.15<- cbind(rmseetreino.15,rmseteste.15, Etreino.15, Eteste.15)
colnames(desempenho.15)<- c("rmse-15/ treinamento", "rmse-15/teste", "E - 15/ treinamento", "E - 15/ teste")
```

```
#Criando uma versao melhor para visualizacao
previsao.15<- cbind(testdata.15,testoutput.15,SaidadeTeste.15)
colnames(previsao.15) <- c("Entrada","Observado","Previsto")
print(previsao.15)
```

```
ggplot(previsao.15, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")
```

```
#-----
```

```
#SPi 12 - Sem defasamento
#20 neuronios
# ?neuralnet para mais informacoes
```

```
#vetor weights refere-se aos pesos iniciais adotados
```

```
weights.20<- c(-0.01      ,0.01      ,-0.02      ,0.02      ,-0.03      ,0.03      ,-0.04      ,0.04      ,-0.05      ,0.05      ,-0.06
,0.06      ,-0.07      ,0.07      ,-0.08      ,0.08      ,-0.09      ,0.09      ,-0.1       ,0.1       ,-0.15      ,0.15      ,-0.2
,0.2       ,-0.25      ,0.25      ,-0.3       ,0.3       ,-0.35      ,0.35      ,-0.4       ,0.4       ,-0.45      ,0.45      ,-0.5
,0.5       ,-0.55      ,0.55      ,-0.6       ,0.6       ,-0.1       ,0.1       ,0.15      ,0.2       ,0.25      ,0.3       ,0.35
,0.4       ,0.45      ,0.5       ,0.55      ,0.6       ,0.65      ,0.7       ,0.75      ,0.8       ,0.85      ,0.9       ,0.1
,0.15      ,0.2)
```

```
#Criando tabelas a partir dos vetores criados
```

```
traininginput.20<- as.data.frame(a1)
trainingoutput.20 <- as.data.frame(b1)
SaidaSPI.20<- as.data.frame (c1)
```

```
testdata.20<- as.data.frame(a2)
testoutput.20<- as.data.frame (b2)
testSPI.20<-as.data.frame(c2)
```

```
#Combinando as colunas em uma variavel
trainingdata.20 <- cbind(traininginput.20,trainingoutput.20)
colnames(trainingdata.20) <- c("Entrada","Saida")
```

```
#Usar validacao cruzada K-fold nos dados de treinamento e validacao a fim de evitar over-training
#randomizar os dados de forma aleatória
trainingdata.20<-trainingdata.20[sample(nrow(trainingdata.20)),]
```

```
#Criar 10 partições de tamanhos iguais
folds.20 <- cut(seq(1,nrow(trainingdata.20)),breaks=10,labels=FALSE)
```

```
#Para um learning rate de 0.0001
#10 fold cross validation
```

```
for(i in 1:10){
  #segmentar os dados por fold usando a funcao which()
```

```
testIndexes.20 <- which(folds.20==i,arr.ind=TRUE)
dadovalidacao.20<- trainingdata.20[testIndexes.20, ]
dadotreinamento.20 <- trainingdata.20[-testIndexes.20, ]
```

```
#Treinando a rede neural
#Com 20 neuronios na camada intermediaria
```

```
##*Threshold is a numeric value specifying the threshold for the partial derivatives of the error function as
stopping criteria.*
```

```
nn.20 <- neuralnet(Saida~Entrada,dadotreinamento.20, hidden=20, stepmax = 10000000, startweights =
weights.20, learningrate = 0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = NULL)
```

```
print(nn.20)
```

```
#Plotar a Rede Neural
```

```
plot(nn.20)
```

```
#Criar uma tabela com os valores dos pesos de cada K-fold
```

```
valores.20<- as.data.frame(nn.20$result.matrix)
```

```
assign(paste("W",i,sep=""),valores.20)
```

```
}
```

```
#Combinar as tabelas em uma so
```

```
vf.20 <- cbind(W1,W2,W3,W4,W5,W6,W7,W8,W9,W10)
```

```
colnames(vf.20) <- c("K-fold1","K-fold2","K-fold3","K-fold4","K-fold5","K-fold6","K-fold7","K-fold8","K-
fold9","K-fold10")
```

```
#Calcular a media dos pesos
```

```
mean.20 <- as.data.frame(apply(vf.20,1,mean))
```

```
valoresfinais.20<- cbind(vf.20,mean.20)
```

```
pesos.20<-as.vector(valoresfinais.20[4:64,11])
```

```
print(pesos.20)
```

```
#Rede neural com o erro minimizado e utilizando a media dos pesos dos k-folds
```

```
NN.20<-neuralnet(Saida~Entrada, trainingdata.20, hidden=20, 1000000, startweights = pesos.20, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
```

```
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.20)
```

```
print(NN.20)
```

```
print(NN.20$result.matrix)
```

```
plot (NN.20)
```

```
saidaredeneural.20 <- as.data.frame(NN.20$net.result)
```

```
Saidaredeneural.20<- as.data.frame((((saidaredeneural.20 -0.1) *(5.073048))/0.8)+(-2.51862))
```

```
saida.20 <- cbind (SaidaSPI.20, Saidaredeneural.20)
```

```
colnames(saida.20)<- c("Observado", "Previsto")
```

```
#tabela comparando os valores observados e previstos
```

```
treinamento.20<-cbind(traininginput.20, SaidaSPI.20, Saidaredeneural.20)
```

```
colnames(treinamento.20) <- c("Entrada","Observado", "Previsto")
```

```
#Plotar os valores observados e previstos e calcular seu coeficiente de correlacao
```

```
plot(saida.20$Previsto,saida.20$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
```

```
title(main="Previsto x Observado")
```

```
grid()
```

```
R2.20<-lm(Previsto~Observado,data=saida.20)
```

```

summary(R2.20)
summary(R2.20)$r.squared
abline(R2.20)

# Funcao que retorna o erro medio quadraticoFunction
rmse <- function(error.20)
{
  sqrt(mean(error.20 ^2))
}

# Calculate error
error.20 <- SaidaSPl.20 - Saidaredeneural.20
rmse(error.20)
rmsetreino.20<- as.data.frame(rmse(error.20))

#Coeficiente de eficiencia (E)

x.20<- (error.20)^2
xm.20<- mean(c1)
y.20<-(SaidaSPl.20 - xm.20)^2
print(x.20)
E.20 = 1 - (sum(x.20)/sum(y.20))
print(E.20)
Etreino.20<-as.data.frame(E.20)

#Plotar valores para visualizacao
ggplot(treinamento.20, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")
#Testar a rede neural nos dados de teste
nnteste.20<- compute(NN.20, testdata.20)
#Observar suas propriedades
ls(nnteste.20)

#Observando os resultados...
print(nnteste.20$net.result)

SaidadeTeste.20 <- as.data.frame(nnteste.20$net.result)

# Calculando RMSE para o teste
error.teste.20 <- testSPl.20 - SaidadeTeste.20
rmse(error.teste.20)
rmseteste.20<- as.data.frame(rmse(error.teste.20))

#Coeficiente de eficiencia (E) para o teste

x.20.t<- (error.teste.20)^2
xm.20.t<- mean(c2)
y.20.t<-(testSPl.20 - xm.20.t)^2
print(x.20.t)
E.20.t = 1 - (sum(x.20.t)/sum(y.20.t))
print(E.20.t)
Eteste.20<-as.data.frame(E.20.t)

#Criar uma tabela com o RMSE e E para fases de treinamento e teste

desempenho.20<- cbind(rmsetreino.20,rmseteste.20, Etreino.20, Eteste.20)
colnames(desempenho.20)<- c("rmse-20/ treinamento", "rmse-20/teste", "E - 20/ treinamento", "E - 20/ teste")

```

```

#Criando uma versao melhor para visualizacao
previsao.20<- cbind(testdata.20,testoutput.20,SaidadeTeste.20)
colnames(previsao.20) <- c("Entrada","Observado","Previsto")
print(previsao.20)

ggplot(previsao.20, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#-----
#Unindo as tabelas dos indices de desempenho a fim de comparacao
indices_desempenho<- cbind(desempenho.5,desempenho.10,desempenho.15,desempenho.20)
print(indices_desempenho)
# Atraves da analise dos indices de desempenho para a fase de treino, a rede neural
#escolhida foi a que apresentou na fase de teste menor erro medio quadratico e maior eficiencia
# A RNA com dez neuronios na camada intermediaria foi a escolhida
#-----

#SPI 12 - Sem defasamento
#Posto Salgadinho

Asalgadinho<- c(0.352998249, 0.590072147, 0.245608335, 0.1, 0.259343108, 0.325130612,
0.49253541, 0.336974354, 0.51831303, 0.349713828, 0.493630236, 0.411520271,
0.202313997, 0.286912161, 0.183105241, 0.355884532, 0.362353806, 0.381562579,
0.380169181, 0.158820593, 0.251878571, 0.306120917, 0.392411022, 0.286414526,
0.478601642, 0.673177371, 0.382557836, 0.336178136, 0.483976103, 0.482184632,
0.394899209, 0.267603882, 0.567180841, 0.3254292, 0.644115438, 0.830131863,
0.597835279, 0.325528742, 0.611769071, 0.526673275, 0.350211486,
0.328215972, 0.270589681, 0.319158982, 0.16528986, 0.551455543)

Bsalgadinho<- c(0.463066388, 0.681566569, 0.335521226, 0.1, 0.353364817, 0.432349244,
0.599169461, 0.445582099, 0.621803189, 0.459520681, 0.600144452, 0.523275286,
0.275469072, 0.387682262, 0.246661763, 0.466166457, 0.47306192, 0.49311792,
0.491683566, 0.207932171, 0.343732712, 0.410524084, 0.504182986, 0.387079348,
0.586650569, 0.745693956, 0.494140508, 0.444700928, 0.591503738, 0.589889475,
0.506697224, 0.363849659, 0.66298553, 0.432686205, 0.723824819, 0.85517991,
0.6877722, 0.4327985, 0.698793453, 0.629003674, 0.460059169,
0.435822656, 0.367595629, 0.425572876, 0.218524967, 0.649966587)

Csalgadinho<- (((Bsalgadinho-0.1) *(5.073048))/0.8)+(-2.51862)

EntradaS <- as.data.frame(Asalgadinho)

SaidaS<- as.data.frame(Bsalgadinho)

SaidaNS<- as.data.frame(Csalgadinho)

DadosS<- cbind(EntradaS, SaidaS, SaidaNS)

colnames(DadosS)<- c("Entrada", "Observado", "SaidaSPI")

NN.10S<-neuralnet(Observado~Entrada, DadosS, hidden=10, 1000000, startweights = pesos.10, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.10)

print(NN.10S$result.matrix)

```

```
plot (NN.10S)
```

```
saidaredeneural.S <- as.data.frame(NN.10S$net.result)
Saidaredeneural.S<- as.data.frame((((saidaredeneural.S -0.1) *(5.073048))/0.8)+(-2.51862))
```

```
#Criando tabela
```

```
RedeNeuralS<-cbind(EntradaS,SaidaNS, Saidaredeneural.S)
colnames(RedeNeuralS) <- c("Entrada","Observado", "Previsto")
```

```
# Calculando Erro medioquadratico
erroS <- SaidaNS - Saidaredeneural.S
rmse(erroS)
```

```
EQMS<- as.data.frame(rmse(erroS))
```

```
#Coeficiente de eficiencia (E)
```

```
x.S<- (erroS)^2
xm.S<- mean(Csalgadinho)
y.S<-(SaidaNS - xm.S)^2
```

```
E.S = 1 - (sum(x.S)/sum(y.S))
print(E.S)
```

```
Ef.S<-as.data.frame(E.S)
#Calculando o raio de discrepancia
```

```
#Unindo os indices
```

```
indices_desempenho_Salgadinho<- cbind(EQMS, Ef.S,DR.S)
colnames(indices_desempenho_Salgadinho)<- c("Erro_quadratico_medio", "Coeficiente_eficiencia")
```

```
plot(RedeNeuralS$Previsto,RedeNeuralS$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Posto Salgadinho")
grid()
R2.S<-lm(Previsto~Observado,data=RedeNeuralS)
```

```
summary(R2.S)$r.squared
abline(R2.S)
ggplot(RedeNeuralS, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")
```

```
#-----
```

```
#SPI 12 - Sem defasamento
#Posto Barra de Sao Miguel
```

```
Abarra<- c(0.392576284, 0.695309471, 0.399911841, 0.578363605, 0.336289894, 0.447451947,
0.389754911, 0.21765121, 0.292135429, 0.454928577, 0.518832668, 0.785311252,
0.465649799, 0.418673961, 0.683882909, 0.65101395, 0.126379839, 0.118197851,
0.389613828, 0.198324823, 0.114671141, 0.438000345, 0.9, 0.702927193,
0.317527763, 0.41839183, 0.460571334, 0.1, 0.242056099)
```

```
Bbarra<- c(0.475775465, 0.748381019, 0.483335743, 0.651450071, 0.415479274, 0.530959759,
0.472850141, 0.272548142, 0.365064379, 0.538245067, 0.598431769, 0.81752685,
0.548599403, 0.502406303, 0.73928618, 0.712692378, 0.142286541, 0.129433289,
0.472703602, 0.2467291, 0.123822344, 0.521672601, 0.9, 0.754402626,
0.394416046, 0.502122245, 0.543708189, 0.1, 0.30401436)
```

```

Cbarra<- (((Bbarra-0.1) *(4.074313))/0.8)+(-1.94696)

EntradaB <- as.data.frame(Abarra)

SaidaB<- as.data.frame(Bbarra)

SaidaNB<- as.data.frame(Cbarra)

DadosB<- cbind(EntradaB, SaidaB, SaidaNB)

colnames(DadosB)<- c("Entrada", "Observado", "SaidaSPI")

NN.10B<-neuralnet(Observado~Entrada, DadosB, hidden=10, 1000000, startweights = pesos.10, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.10)

print(NN.10B$result.matrix)
plot (NN.10B)

saidaredeneural.B <- as.data.frame(NN.10B$net.result)
Saidaredeneural.B<- as.data.frame((((saidaredeneural.B -0.1) *(4.074313))/0.8)+(-1.94696))

#Criando tabela

RedeNeuralB<-cbind(EntradaB,SaidaNB, Saidaredeneural.B)
colnames(RedeNeuralB) <- c("Entrada","Observado", "Previsto")

# Calculando Erro medioquadratico
erroB <- SaidaNB - Saidaredeneural.B
rmse(erroB)

EQMB<- as.data.frame(rmse(erroB))

#Coeficiente de eficiencia (E)

x.B<- (erroB)^2
xm.B<- mean(Cbarra)
y.B<-(SaidaNB - xm.B)^2

E.B = 1 - (sum(x.B)/sum(y.B))
print(E.B)

Ef.B<-as.data.frame(E.B)
#Calculando o raio de discrepancia

#Unindo os indices
indices_desempenho_Barra<- cbind(EQMB, Ef.S)
colnames(indices_desempenho_Barra)<- c("Erro_quadratico_medio", "Coeficiente_eficiencia")

plot(RedeNeuralB$Previsto,RedeNeuralB$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Previsto x Observedo")
grid()

ggplot(RedeNeuralB, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

```

```
#-----
```

```
#SPI 12 - Sem defasamento
```

```
#Posto Barra de Sao Miguel
```

```
Abarra<- c(0.392576284, 0.695309471, 0.399911841, 0.578363605, 0.336289894, 0.447451947,
0.389754911, 0.21765121, 0.292135429, 0.454928577, 0.518832668, 0.785311252,
0.465649799, 0.418673961, 0.683882909, 0.65101395, 0.126379839, 0.118197851,
0.389613828, 0.198324823, 0.114671141, 0.438000345, 0.9, 0.702927193,
0.317527763, 0.41839183, 0.460571334, 0.1, 0.242056099)
```

```
Bbarra<- c(0.475775465, 0.748381019, 0.483335743, 0.651450071, 0.415479274, 0.530959759,
0.472850141, 0.272548142, 0.365064379, 0.538245067, 0.598431769, 0.81752685,
0.548599403, 0.502406303, 0.73928618, 0.712692378, 0.142286541, 0.129433289,
0.472703602, 0.2467291, 0.123822344, 0.521672601, 0.9,
0.754402626,0.394416046, 0.502122245, 0.543708189, 0.1, 0.30401436)
```

```
Cbarra<- (((Bbarra-0.1) *(4.074312554))/0.8)+(-1.94696)
```

```
EntradaB <- as.data.frame(Abarra)
```

```
SaidaB<- as.data.frame(Bbarra)
```

```
SaidaNB<- as.data.frame(Cbarra)
```

```
DadosB<- cbind(EntradaB, SaidaB, SaidaNB)
```

```
colnames(DadosB)<- c("Entrada", "Observado", "SaidaSPI")
```

```
NN.10B<-neuralnet(Observado~Entrada, DadosB, hidden=10, 1000000, startweights = pesos.10, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
threshold = 0.01,linear.output = FALSE, constant.weights = pesos.10)
```

```
print(NN.10B$result.matrix)
```

```
plot (NN.10B)
```

```
saidaredeneural.B <- as.data.frame(NN.10B$net.result)
```

```
Saidaredeneural.B<- as.data.frame((((saidaredeneural.B -0.1) *(4.696224))/0.8)+(-2.05341))
```

```
#Criando tabela
```

```
RedeNeuralB<-cbind(EntradaB,SaidaNB, Saidaredeneural.B)
```

```
colnames(RedeNeuralB) <- c("Entrada", "Observado", "Previsto")
```

```
# Calculando Erro medioquadratico
```

```
erroB <- SaidaNB - Saidaredeneural.B
```

```
rmse(erroB)
```

```
EQMB<- as.data.frame(rmse(erroB))
```

```
#Coeficiente de eficiencia (E)
```

```
x.B<- (erroB)^2
```

```
xm.B<- mean(Cbarra)
```

```
y.B<-(SaidaNB - xm.B)^2
```

```
E.B = 1 - (sum(x.B)/sum(y.B))
```

```
print(E.B)
```

```
Ef.B<-as.data.frame(E.B)
```

```
#Unindo os indices
```

```

indices_desempenho_Barra<- cbind(EQMB, Ef.B)
colnames(indices_desempenho_Barra)<- c("Erro_quadratico_medio", "Coeficiente_eficiencia")

plot(RedeNeuralB$Previsto,RedeNeuralB$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Posto Barra de São Miguel")
grid()

R2.B<-lm(Previsto~Observado,data=RedeNeuralB)

summary(R2.B)$r.squared
abline(R2.B)

ggplot(RedeNeuralB, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

#-----
#SPI 12 - Sem defasamento

#Posto Passagem

Apassagem<- c(0.212257962,    0.51459067,    0.592119259,    0.610636741,    0.257651075,
              0.521291497,    0.439223631,    0.414149325,    0.1,    0.705962349,    0.367459252,
              0.40168422,    0.688669735,    0.630811495,    0.294037637,    0.604152018,    0.426182118,
              0.307295327,0.23574708,0.311114108,    0.212546173,    0.231061195,    0.44282628,    0.9,
              0.505656126,    0.278546323,    0.336764844,    0.438284466)

Bpassagem<- c(0.295614186,    0.623736064,    0.687421329,    0.701880377,    0.357601465,
              0.629457931,    0.556105014,    0.532121749,    0.1,    0.772421462,    0.484972243,
              0.519869379,    0.760071886,    0.717335945,    0.402848401,    0.69684732,    0.543733264,
              0.418512845,0.23574708,0.311114108,    0.212546173,    0.231061195,    0.44282628,    0.9,
              0.505656126,    0.278546323,    0.336764844,    0.438284466)

Cpassagem<- (((Bpassagem-0.1) *(4.797912))/0.8)+(-2.63593)
EntradaP <- as.data.frame(Apassagem)

SaidaP<- as.data.frame(Bpassagem)

SaidaNP<- as.data.frame(Cpassagem)

DadosP<- cbind(EntradaP, SaidaP, SaidaNP)

colnames(DadosP)<- c("Entrada", "Observado", "SaidaSPI")

NN.10P<-neuralnet(Observado~Entrada, DadosP, hidden=10, 1000000, startweights = pesos.10, learningrate =
0.0001, lifesign = "full", algorithm = "backprop",err.fct = "sse",
                threshold = 0.01,linear.output = FALSE, constant.weights = pesos.10)

print(NN.10P$result.matrix)
plot (NN.10P)

saidaredeneural.P <- as.data.frame(NN.10P$net.result)
Saidaredeneural.P<- as.data.frame((((saidaredeneural.P -0.1) *(4.797912))/0.8)+(-2.63593))
#Criando tabela

RedeNeuralP<-cbind(EntradaP,SaidaNP, Saidaredeneural.P)
colnames(RedeNeuralP) <- c("Entrada","Observado", "Previsto")

```

```

# Calculando Erro medioquadratico
erroP <- SaidaNP - Saidaredeneural.P
rmse(erroP)

EQMP<- as.data.frame(rmse(erroP))

#Coeficiente de eficiencia (E)

x.P<- (erroP)^2
xm.P<- mean(Cpassagem)
y.P<-(SaidaNP - xm.P)^2

E.P = 1 - (sum(x.P)/sum(y.P))
print(E.P)

Ef.P<-as.data.frame(E.P)

#Unindo os indices
indices_desempenho_passagem<- cbind(EQMP, Ef.P)
colnames(indices_desempenho_passagem)<- c("Erro_quadratico_medio", "Coeficiente_eficiencia")

plot(RedeNeuralP$Previsto,RedeNeuralP$Observado,col="deeppink", xlab="Previsto",ylab="Observado")
title(main="Posto Passagem")
grid()

R2.P<-lm(Previsto~Observado,data=RedeNeuralP)

summary(R2.P)$r.squared
abline(R2.P)

ggplot(RedeNeuralP, aes(Entrada)) +
  geom_line(aes(y = Observado, colour = "Observado")) +
  geom_line(aes(y = Previsto, colour = "Previsto"))+
  ggtitle("Séries Temporais")

```