UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de Jaboticabal - SP

JOÃO TREVIZOLI ESTEVES

CLIMATE AND AGROMETEOROLOGY FORECASTING USING SOFT COMPUTING TECHNIQUES.

Jaboticabal - SP
2018

João Trevizoli Esteves

CLIMATE AND AGROMETEOROLOGY FORECASTING USING SOFT COMPUTING TECHNIQUES.

Dissertação apresentada à Faculdade de Faculdade de Ciências Agrárias e Veterinárias do Câmpus de Jaboticabal - UNESP como parte dos requisitos para obtenção do título de Mestre em Produção Vegetal

Prof. Dr. Glauco Rolim
Orientador

Jaboticabal - SP
2018

E79c

Esteves, João
Climate and agrometeorology dorcasting using soft computing techniques / João Esteves. -- Jaboticabal, 2018
54 p. : il., tabs., mapas

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal
Orientador: Glauco de Souza Rolim
Coorientador: Antonio Sergio Ferraudo

1. Modelos Climáticos. 2. Redes Neurais Artificiais. 3. Multivariada. 4. Precipitação pluviométrica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Jaboticabal

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: CLIMATE AND AGROMETEOROLOGY FORECASTING USING SOFT COMPUTING TECHNIQUES

AUTOR: JOÃO TREVIZOLI ESTEVES
ORIENTADOR: GLAUCO DE SOUZA ROLIM

Aprovado como parte das exigências para obtenção do Título de Mestre em AGRONOMIA (PRODUÇÃO VEGETAL), pela Comissão Examinadora:

Prof. Dr. GLAUCO DE SOUZA ROLIM
Departamento de Ciências Exatas / FCAV / UNESP - Jaboticabal

Prof. Dr. NELSON JOSÉ PERUZZI
Departamento de Ciências Exatas / Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal

Prof. Dr. RONALDO RIBEIRO DE CAMPOS
FATEC / Taquaritinga/SP

Jaboticabal, 07 de dezembro de 2018

To my Family, in special to my wife Natalia and to my dog Tonico, for the trust, support an care that gave me the strength to carry this research.

GREETINGS

My greetings to all family members that supported me trough this research. To all the teachers and employees of the University that direct or indirectly contributed to the achievement of this research. I would like to dedicate special thanks to:

- To my wife Natalia Jorge Esteves.
- To my Mother Cristina de Oliveira Trevizoli;
- To Labmet crew for the support in processing the research data.
- To Prof. Dr. Glauo de Souza Rolim for being my tutor and guide.
- To Prof. Dr. Antonio Sergio Ferraudo who inspired me to be here.
- To all my colleagues in the Group of Agrometeorology Studies.

“conceived it to be the privilege of every citizen, and one of his most essential rights, to bear arms, and to resist every attack upon his liberty or property, by whomsoever made”

Roger Sherman, 1787

RESUMO

Precipitação, em pequenas escalas de tempo, é um fenômeno associado a altos níveis de incerteza e variabilidade. Dada a sua natureza, técnicas tradicionais de previsão são dispendiosas e exigentes em termos computacionais. Este trabalho apresenta um modelo para prever a ocorrência de chuvas em curtos intervalos de tempo por Redes Neurais Artificiais (RNAs) em períodos acumulados de 3 a 7 dias para cada estação climática, mitigando a necessidade de predizer o seu volume. Com essa premissa pretende-se reduzir a variância, aumentar a tendência dos dados diminuindo a responsabilidade do algoritmo que atua como um filtro para modelos quantitativos, removendo ocorrências subsequentes de valores de zero(ausência) de precipitação, o que influencia e reduz seu desempenho. O modelo foi desenvolvido com séries temporais de 10 regiões agricolamente relevantes no Brasil, esses locais são os que apresentam as séries temporais mais longas disponíveis e são mais deficientes em previsões climáticas precisas, com 60 anos de temperatura média diária do ar e precipitação acumulada. foram utilizados para estimar a evapotranspiração potencial e o balanço hídrico; estas foram as variáveis utilizadas como entrada para as RNAs. A precisão média para todos os períodos acumulados foi de 78% no verão, 71% no inverno 62% na primavera e 56% no outono, foi identificado que o efeito da continentalidade, o efeito da altitude e o volume da precipitação normal , tem um impacto direto na precisão das RNAs. Os modelos têm desempenho máximo em estações bem definidas, mas perdem sua precisão em épocas de transição e em locais sob influência de efeitos microclimáticos e mesoclimáticos, o que indica que essa técnica pode ser usada para indicar a eminência da precipitação com algumas limitações.

Palavras-chave: previsão pluviométrica, redes neurais artificiais, redes multicamadas perceptron

ABSTRACT

Precipitation, in short periods of time, is a phenomenon associated with high levels of uncertainty and variability. Given its nature, traditional forecasting techniques are expensive and computationally demanding. This paper presents a model to forecast the occurrence of rainfall in short ranges of time by Artificial Neural Networks(ANNs) in accumulated periods from 3 to 7 days for each climatic season, mitigating the necessity of predicting its amount. With this premise it is intended to reduce the variance, rise the bias of data and lower the responsibility of the model acting as a filter for quantitative models by removing subsequent occurrences of zeros values of rainfall which leads to bias the and reduces its performance. The model were developed with time series from 10 agriculturally relevant regions in Brazil, these places are the ones with the longest available weather time series and and more deficient in accurate climate predictions, it was available 60 years of daily mean air temperature and accumulated precipitation which were used to estimate the potential evapotranspiration and water balance; these were the variables used as inputs for the ANNs models. The mean accuracy of the model for all the accumulated periods were 78% on summer, 71% on winter 62% on spring and 56% on autumn, it was identified that the effect of continentality, the effect of altitude and the volume of normal precipitation, have an direct impact on the accuracy of the ANNs. The models have peak performance in well defined seasons, but looses its accuracy in transitional seasons and places under influence of macro-climatic and mesoclimatic effects, which indicates that this technique can be used to indicate the eminence of rainfall with some limitations.

Keywords: artificial neural networks, multilayer perceptron, rainfall forecasting

FIGURE LIST

TABLE LIST

ABBREVIATION LIST AND ACRONYMS

ANN	Artificial Neural Network
CWS	Conventional Weather Stations
GCM	Global Circulation Model
GDX	Adaptive Learning Rate and Momentum
MSE	Mean Squared Error
NWP	Numerical Weather Prediction Model
PET	Potential Evapotranspiration
SMAPE	Mean Absolute Percentage Error
SWC	Soil Water Content
TWB	Thornthwaite Water Balance
WMO	World Meteorological Organisation

SYMBOL LIST

ρ	Correlation coefficient
σ	Variance
Z_i^p	Normalised value
l_i	Minimum value
u_i	Maximum value
H_O	Extraterrestrial irradiation energy
ND	Number of days in a given period
Tn	is the period mean air temperature
ϕ	Geographic latitude
δ	Solar declination
w_n	ANN connection weights
S_n	Activation function
g_n	Gradient of error
α	Learning rate
μ	Momentum constant

SUMÁRIO

1 INTRODUCTION

Water is essential for all human activities and agriculture is the largest freshwater consumer. Precipitation is a phenomenon highly susceptible to variability and determines its availability (CALZADILLA et al., 2013). Research and apply accurate statistical models to forecast this phenomena has been acknowledged to play a key role for this sector of the human activity (TOTH; BRATH; MONTANARI, 2000). Given the uncertainty and variability that drives its occurrence, it is recognised that is quite difficult to obtain reliable and accurate prediction models that can spatially forecast this element of the hydrological cycle for short periods of time (BRATH, 1997). It is known that due to its behaviour and complex structure, precipitation is an variable harder to forecast than other climate variables, given the processes involved in its generation and nonlinear behaviour (JHA et al., 2018).

The precipitation forecasting problem is commonly approached in different ways. The use of remote sensing observation with radars and satellite images addresses the issue based on the extrapolation of current weather condition, for very short term forecasting (scale of minutes). Unfortunately the use of radar and satellite images do not provide a satisfactory assessment of rain intensities in larger scales of time, in addition, using this technique in mountainous regions is difficult because of the occurrence of soil shading and the altitude effect (TOTH; BRATH; MONTANARI, 2000).

One other mean to obtain rainfall forecasting models is by time series analyses techniques. There are different approaches to time series forecasting, specially for climatic proposes.Traditionally forecasting has long been the domain of linear statistics, usual approaches to time series prediction, such as Box-Jenkins 1976 or ARIMA (autoregressive integrated moving average) method (PANKRATZ, 1983), considers that time series behaves as linear processes. Despite of its easy understanding and applicability it may be totally inappropriate to implement if the ongoing mechanism is subjected to an nonlinear processes (ZHANG, 2003).

In meteorology to deal with non linearity, it is generally used numerical weather prediction models (NWP) in applications such as Global Circulation Models (GCM). NWP is an initial-value problem for which initial data are not available in sufficient quantity and with sufficient accuracy, these models abstract some layers of information by discretizing partial differential equations governing large scale atmospheric flow (GHIL et al., 1981). GCM models are based on highly complex mathematical representations of atmospheric, oceanic, and continental processes being capable to predict climate patterns of different variables such as air temperature, precipitation, atmospheric gases and its behaviour. These models simulates climatic parameters only at grid points requiring downscale of regional models to local models (ALOTAIBI et al., 2018).

NWP can active acceptable accuracy in forecasting some meteorological phenomenas but when dealing with rainfall they yet have not active it (RAMÍREZ; FERREIRA; VELHO, 2006), mainly because of the physical complexity of precipitation processes and the reduced temporal and spacial scale involved in such phenomena that numerical models cannot resolve (KULIGOWSKI; BARROS, 1998b). It is required turbulent parameterizations to accurately represent the planetary boundary layer, shallow convection, subgrid-scale cloud cover, and turbulent fluxes related to deep convective systems which are required to future climate projections. Due to limitation on knowledge about cloud-aerosol interactions which are a major source of uncertainties on NWP models leads to far-reaching consequences on the development and accuracy of precipitation models (PREIN et al., 2015). One other drawback that NWP models such as GCM have is that they are computationally demanding and require powerful and expensive hardware to be implemented in a meteorological prediction center. Limitations in computing power may result in inability to appropriately resolve the important climate processes. Low-resolution models fail to capture many important phenomena on regional and lesser scales such as clouds. Downscaling to higher-resolution models introduces boundary interactions that can contaminate the modeling area and propagate error (ALOTAIBI et al., 2018).

More recently researchers have been approaching such problem with artificial neural networks (ANN) which are a powerful alternative to traditional time-series modelling (ZHANG, 1998) as for NWP models. ANNs are data-driven self adaptive methods that are able to understand and solve problems of which there is not enough data or observations to use more traditional statistical models (ZHANG; PATUWO; HU, 1998), rainfall is such a phenomena and ANNs are suited and studied solution.

ANNs are a type of nonlinear model inspired by sophisticated functionalities of human brain. They are universal function approximators that can adaptively discover patterns from data, learn from experience and estimate any complex functional relationship with high accuracy (ZHANG, 1998; WANG, 2003), they mimics the brain functionalities both in knowledge acquisition through a learning process and memory by storing synaptic weights as acquired knowledge (FERRAUDO, 2014). ANNs have an nonparametric nature which enables the development of models without any prior knowlege of the population, its distribution or possible interaction between variables that are commonly used in parametric statistical models (WALCZAK, 2019).

ANNs simulates an reduced set of concepts derived from biological neural systems by emulating the electrical activity of the brain of which each part of the neurone plays a role in the communication of the information throughout its parts. Computations and analysis of the brain are achieved by sending electric signals through its processing units which consists of dendrites, axons, terminal buttons and synapses (KROGH, 2008). Dendrites receives signals from over to the cell body of the neuron. The axon receives signals from cell body and carries them through the sinapses to neighbour neurones dendrites. In math models the processing units, are interconnected in layers or vectors which the output of each neuron serves as input for neighbour neurones. When an electric signal travels from the dendrites to the pre synaptic membrane of the synapse a chemical called neurotransmitter is released in proportional amount to the strength of the signal. The neurotransmitter, diffused within the gap between the synaptic membrane and the neighbour dendrites forces the receiving neurone to generate a new signal that obeys the same set of rules to transmit its impulse (BASHEER; HAJMEER, 2000). The amount of signal passed depends on the intensity of the signal emanated from feeding neurons, its strength and the activation threshold of the receiving neuron which can assist or inhibit the firing neurone. This simplified biological mechanism of signal transferring are the bases of ANN and neurocomputing.

The first artificial neurone model was proposed by McCulloch and Pitts in 1943, it was designed to behaviour as a switch which alter its state depending on its input passing through an weight distribution process. The weight multiplies the inputs corresponding to the strength of the synapses that represents the contact between nerve cells (MCCULLOCH; PITTS, 1943) which can be both positive or excitatory, allowing the electrical pulse to pass, and negative or inhibitory blocking the signals.

In 1958 ROSENBLATT to understand the process of perceptual recognition of higher organisms and to answer three fundamental questions of neural thinking: How the biological system senses and detects information? What is the form that it is stored and remembered? How storage information influences on recognition behaviour? Proposed an hypothetical nervous system called perceptron. The perceptron was designed to illustrate properties of intelligent system without being deeply attached into unknown meshes which are the natural condition for biological organisms. The machine establishes a mapping between the inputs activity and the output signal by passing signals through a linear threshold function and transmitting its signal to other neurons or the environment. By using weights in the connections the signals can be both excited enhancing its strength or inhibited reducing it (BASHEER; HAJMEER, 2000).

The perceptron weights and thresholds can be adjusted in a training processes. This process is equivalent to approximating the output of the neuron to the corresponding desired counterpart or goal by minimising an error function computed by the difference between the goal of a training set and the output of the ANN in a search for minima (RAMCHOUN et al., 2016). The perceptron is a single element ANN that responds correctly to as many patterns as possible, being able to respond correctly with high probability to input patterns that were not included in the training set if the output is binary (WIDROW; LEHR, 1990). In 1969 MINSKY; PAPERT mathematically proved the limitations of the perceptron and other types of ANNs when dealing with non linear separable patterns.

With the rediscovery of the backpropagation algorithm by RUMELHART; HINTON; WILLIAMS (1985) originally proposed by WERBOS (1974) solved the problem of training and implementing non linear solvers that handle non linear groups of variables. To handle non linear problems intermediate layers connected in nodes are added between input and output neurones of the ANN, since this layers of neurons do not connect to the external world they are called hidden layers. This structure is called Multilayer Perceptron (MLP). With the addition of intermediary layers to the perceptron using analogous dynamics and with the implementation of nonlinear training algorithms (backpropagation) the neurons process information and pass over to the output layer with accuracy.

In the field of agriculture and applied math ANNs has been a successful tool to forecast meteorological indexes. KUMARASIRI; SONNADARA (2006) proposed three Neural Network models based on the feedforward backpropagation architecture to forecast rainfall in a short-term or one day ahead, medium-term or one month ahead and long-term or a year in the city of Colombo in Sri Lanka. The researchers obtaining an accuracy from short to long term of 74.25%, 58.33% and 76.67%. According to the author the region had well defined seasons and long strings of observations with rain in the monsoon seasons and no rain observation days which contributed to the performance of the models.

To simulate chaotic rainfall events in the suburbs of Sydney, Australia NASSERI; ASGHARI; ABEDINI (2008) proposed an architecture of ANN that is efficient in events with a scale of minutes. The architecture was based on the feedforward backpropagation coupled with genetic algorithm. The genetic algorithm are a kind of computational models inspired in evolution, they encode problem solution on a chromosome-like structure coupled with operators that recombines structures to preserve critical operations and can be viewed as function optimisers (WHITLEY, 1994). The authors reported that the study led to conclude that associating ANN with genetic algorithm performed with accuracy and given the high variance and turbulence of precipitation events cumulative data leads to increasing statistical performance and when comparing rainfall forecasting to discrete data types.

Some studies used ANN models together with NWP models. RAMÍREZ; FERREIRA; VELHO (2006) to generate accurate rainfall forecasts over southeastern Brazil areas used artificial neural networks to downscale the Eta Model with a resolution of 40 x 40 km to forecast variables at a weather station level. The Eta model is a state of the art atmospheric model (NWP model) used for research and operational purposes. The study were able to conclude that ANNs are effective to adjust rainfall forecast for specific points and that NWP models accuracy are reversibly proportional to ANNs in events of heavy rain because they are more effective in events with higher threshold.

The rainfall due to the complexity of the physical processes involved and its variability in space and time is a difficult variable to forecast. Mapping the effect of temporal and spatial information on short term rainfall forecast is a key component into the development of a successful model. Rainfall is considered a Markovian process (LUK; BALL; SHARMA, 2000) that is a particular case of a stochastic process with discrete estates, which implies that its volume at a given location in a place and time is function of a previous set of observations. Considering this factors knowing the relation between future and past rainfall events is crucial to develop an appropriate ANN architecture that maps this relations and is able to carry over the momentum and accurate to predict. In previous studies, while investigating the effect of temporal and spatial rainfall events in very short periods of time, LUK; BALL; SHARMA (2000) revealed that there is an optimal limit temporal and spacial limit for inclusion into a ANN. The author also demonstrated that too much or too little spacial information can degrade its performance and that for short term rainfall it might not have long therm memory indicating that with lower lags consistently produced smaller prediction errors. Other authors corroborates with this statement and proved that ANN are able to generalize and use previous input to accurate forecast. FRENCH; KRAJEWSKI; CUYKENDALL (1992) demonstrated that by only increasing the number of training iterations the performance can be improved which is not the case on independent data.

In other studies (KUMARASIRI; SONNADARA, 2006; NASSERI; ASGHARI; ABEDINI, 2008; RAMÍREZ; FERREIRA; VELHO, 2006; LUK; BALL; SHARMA, 2000; FRENCH; KRAJEWSKI; CUYKENDALL, 1992; TOTH; BRATH; MONTANARI, 2000; PARTAL; CIGIZOGLU; KAHYA, 2015) the goal was to numerically predict, with a single ANN structure, the accumulative volume of precipitation in a given scale in a future period of time. The performance of these models were very correlated to the time scale of events that ANNs had to handle. In larger scale of time, such as months, the performance of ANNs are vastly superior then in shorter periods of time. This happens because in larger periods of time the probability of some precipitation be recorded is greater, consecutively models are not biased by a big number of observations with zero precipitation (SCHOOF; PRYOR, 2001) and in short scale of time rainfalls are dependent on small scale and unstable physical processes (KULIGOWSKI; BARROS, 1998b).

The objective of this research is to create a methodology to predict the occurrence of rainfall. This is done by constraining the complexity of the predicted events by reducing the variance and rising the bias of the time series. To achieve this objective, a structure of artificial neural networks is being proposed which identifies the signs that lead to the occurrence of rain for each climatic season in short periods of time, letting the ANNs to predict whether or not it is going to rain. The proposed model is intended to filter which days are propitious to rain, so that only the climate variables in the periods that lead to rain are used in quantitative models. With this technique quantitative models can improve its forecasting performance in shorter periods of time and consecutively becoming computationally lighter by reducing the volume of data used in the training stage of the models.

2 MATERIAL AND METHODS

In this section, it is firstly described the dataset with emphasis in its composition, recovery of missing data and data transformation, important factors for the model accuracy. Secondly it is discussed the methodology for estimating potential evapotranspiration (PET), indispensable for calculus of water balance (TWB). Lastly it is described the methodology used in the Artificial Neural Networks to forecast small spacial and temporal scales, that is the goal of this paper.

2.1 Dataset

The raw data used to establish the training set for the forecast model consists basically of the daily mean air temperature and the accumulated precipitation, these indexes were ground measured by conventional weather stations (CWS) and were the one available for this study.

It was chosen the most relevant agriculture production regions distributed in eight Brazilian states, in these locations it was selected ten CWS and its locations are shown in Table 1. The CWS were chosen based on geographical proximity of important agricultural centres and by its operation start date, the Fig.1 illustrates its distribution across Brazilian territory. The optimal range of data chosen for training the prediction model was from 1950 to 2011, the years of 2012 to middle 2015 were not known by algorithm for testing and validation purposes ensuring the learning and generalisation capacities of the artificial neural networks. The cross validation method adopted was the holdout method, which is basically a separation of the dataset in two sets, a training set and a validation set, that the function approximator tests its outputs with unknown data, given the large set of data this is an feasible validation method (FRIEDMAN; HASTIE;

TIBSHIRANI, 2001).

Tabela 1 - Geographical locations of Brazilian ground-based conventional weather stations

State	City	Maintainer	Lat(DD)	Long(DD)	Alt(m)
Paraná	Campo Mourão	INMET	- 24.05	- 52.36	616.4
Mato Grosso	Diamantino	INMET	- 14.40	- 56.45	286.3
Mato Grosso do Sul	Ivinhema	INMET	- 22.30	- 53.81	369.2
Ceará	Jaguaruana	INMET	- 4.78	- 37.76	11.7
Alagoas	Maceio	INMET	- 35.70	- 64.50	64.5
São Paulo	Presidente Prudente	INMET	- 22.11	- 51.38	435.5
	Jaboticabal	UNESP	-21.25	-48.32	626.0
	Piracicaba	USP	-22,73	-47.64	547.0
Goiás	Rio Verde	INMET	- 17.8	- 50.91	774.6
Minas Gerais	Uberaba	INMET	- 19.73	- 47.95	737.0

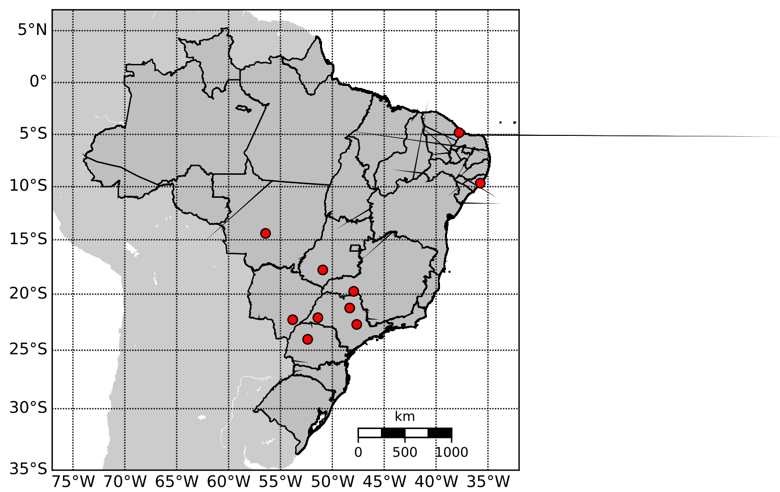


Figura 1 - Weather stations locations

There was air temperature measurements missing within all local datasets, a common problem in long time series. It was necessary to infill the gaps with estimated values to maintain consistency in the training processes.

2.2 Missing Data Recovery

Traditionally the estimation of missing meteorological data are based on measurements of the same location, the reconstruction methods includes simple interpolations using mean values from time series arrays, or even using data from several days before and after the date with no measurement in a non-linear regression (KIM; PACHEPSKY, 2010). ANNs are data-driven, non-linear statistical modelling tools capable to map and understand the relationship between inputs and outputs, this ability renders it possible to simulate large-scale arbitrary complex linear problems (WU; CHAU, 2006) and are often used to forecast time-series (ZHANG, 2003; BOX; JENKINS; REINSEL, 1976; FRENCH; KRAJEWSKI; CUYKENDALL, 1992; ZHANG, 1998).

The ANN implementation chosen was the feed-forward multilayer perceptron with one hidden layer and with 12 neurons, followed by a single neuron output layer. Time-series have a continuous nature and require a transfer function able to output a graded response, to meet this criteria it was chosen Logarithmic transfer function. The best performing transfer function was the logarithmic sigmoid (Eq. 1).

$$y = \frac{1}{1 + e^{-}x} \tag{1}$$

Traditional backpropagation training algorithms are often too slow for practical problems. The performance of these algorithms are improved by allowing the learning rate to change during the training process and keep the learning step size as large as possible, while maintaining learning stable. Gradient search based technics such as backpropagation tend to get trapped at local minima, with enough gain (momentum) it can escape these local minima (MONTANA; DAVIS, 1989). To keep the algorithm responsive to the complexity of the local error surface while getting closer to the local minima, it was adopted the pre. The error function used in the ANN training processes was the mean squared error (MSE) represented by the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}i - Yi)^2 \tag{2}$$

where n is the length of the training array, $\hat{Y}i$ is the predicted value and Yi is the observed value at an given time. All nodes weights were randomly initialised which was a problem, because of the non randomness of computer generated random numbers, this issue will be better discussed further in the *Binary Precipitation ANN* subsection.

Normalising data can improve learning and can impact directly on the computational and classification performance (SHANKER; HU; HUNG, 1996). Prior beginning the training processes every place of the dataset were linearly transformed to the [0, 1] interval, being 0 the minimum value and 1 the maximum value of the dataset, this were done based on the Eq. 3:

$$Z_i^p = \frac{x_i^p - l_1}{u_i - l_i} \tag{3}$$

where Z_i^p is the transformed value, l_i is the minimum and u_i is the maximum value of the time series array.

The dataset used to recover the lost air temperature, was the longest array without any missing air temperature for every location, in each subset it was left untouched around 20 percent of data for algorithm cross validation. It was made an correlation matrix to determine the time-space dependencies of the variable, being considered the interval dependent while the ρ (Eq. 4) value between the time-steps x_{ij} and y_{ij} was greater than 0.5, this was the batch size for the entering layer for each ANN for this reason, it was different for every location.

$$\rho = \frac{\sum_{ij=1}^{n}(x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sqrt{\sum_{ij=1}^{n}(x_{ij} - \bar{x})^2 \sum_{ij=1}^{n}(y_{ij} - \bar{y})^2}} \tag{4}$$

In the validation stage of these ANNs, the variance (σ) between the predicted value and the real one wasn't greater than 1 ^{O}C for every location, this was not the research goal and was considered reasonable to infill usage, no further validation was done and all gaps were filled.

2.3 Weather Indexes Estimation

The proposed model uses as input data based on estimated meteorological indexes which were the soil water content (SWC) in *mm*, the daylight length in *hours* and the extraterrestrial irradiation energy (H_O) in *mm*. These indexes were chosen because of the inertia or carryover processes that they naturally have, these indexes are persistent and tends to have slightly changes from one observation to another unless some event such as precipitation happens. The theory is that the nested information which these indexes inherently carry are a important source of information for the ANN and an positive sign of the rainfall possibility.

To determine the SWC it is necessary to estimate the water balance. This is an practical method developed to quantify the water allocation among watersheds, which calculates its inputs and outputs sequentially, it is usually applied monthly but can be used for monitoring the soil water storage in near-real time (THORNTHWAITE; MATHER, 1957), for the research it was used a daily scale.

In order to determine the water balance of a given place is necessary to estimate the potential evapotranspiration (PET). The PET is the amount of water to be evapotranspirated in a standard grassy surface if there was sufficient water available, this index is considered essential and represents the needed rainfall to supply the vegetation water needs (CAMARGO; CAMARGO, 2000).

The PET values are usually estimated empirically by measured elements in weather stations, there are several methods to estimate its value. The choice of a method for estimating potential evapotranspiration depends on a number of factors. The first one is the availability of meteorological data, complex methods such as Penman–Monteith

(ALLEN et al., 1998), requires a great number of variables which are not aways available. Second is the temporal scale. Usually, empirical methods such as Thornthwaite, estimate the PET well on a monthly scale, whereas methods involving the radiation balance have a better performance in daily scale. Lastly, on empirical methods, it is required to know the climate conditions of which it were developed, some methods like Thornthwaite, are better for humid climates and not capable to perform on arid regions which requires different methods like the one proposed by Hargreaves and Samani (HARGREAVES; SAMANI, 1985).

The Thornthwaite method (THORNTHWAITE, 1948) was the first and widely know to estimate the PET value. It is a empirical method with the drawback of relying on the normal mean air temperature which is not aways available and to be created for humid regions. On 1971 Camargo(CAMARGO; SÃO, 1971) proposed an equation with practically the same results of Thorthwaite original work, without the drawback of needing normal air temperature and has the advantage of computing the extraterrestrial solar irradiation this method was analytically developed specifically for Brazilian conditions. This was the method adopted in this study, follows the Camargo equation:

$$PET = 0.01\, H_O\, Tn\, ND \tag{5}$$

where ND is the number of days contained in the desired period, Tn is the period mean air temperature calculated in ${}^{o}C$ and H_O calculated in $MJm^{-2}day^{-1}$ is the extraterrestrial irradiation energy falling on a plane horizontal to the earths surface throughout a whole day and is represented by the Eq. 6:

$$H_O = 37.6(1+0.033\cos(DOY\frac{360}{365}))[(\frac{\pi}{180^o})N\sin\phi\,\sin\delta+\cos\phi\,\cos\delta\,\sin P] \tag{6}$$

In the H_O equation *DOY* represents the day of year, ϕ is the geographic latitude in *degrees*, δ is solar declination calculated in *degrees* based on Coopers(COOPER, 1969) equation (Eq.7) and *N* is the photoperiod calculated in *hours* by the equation 8. The soil-moisture storage capacities was standardised in 100 *mm* across all locations to simplify the calculus routine.

$$\delta = 23.45 \sin[360 \frac{DOY - 80}{365}] \tag{7}$$

$$N = 2 \frac{\arccos[-\tan\phi \, \tan\delta]}{15^{o}} \tag{8}$$

With these estimated indexes, it was generated an new time-series dataset, for each Brazilian location, that were the estimated data used for the forecasting model with the addition of the the Unix time stamp for each day.

2.4 Binary Precipitation ANN

Traditionally in the field of modelling in climatology and time series, an auto regressive approach is used to solve the index forecasting problem (RAJURKAR; KOTHYARI; CHAUBE, 2002; MISHRA; SHARMA, 2018; RAMÍREZ; FERREIRA; VELHO, 2006). Rainfall is an sparse highly difficult to predict phenomenon that its occurrence depends on a series of complex parameters such as temperature, barometric pressure and wind speed (SUMI; ZAMAN; HIROSE, 2012). Given the nature this phenomenon these approaches relies on historical data that contains high variance, low bias and in short range period of times a great number of very small volumes or a lack of rainfall events. These characteristics make it difficult for traditional models to converge, which leads to a reduction of their potential performance.

The input selection is a key component to develop an accurate rainfall forecast model, many theoretical studies established the relationship between climate indices and rainfall.TULARAM; ILAHEE (2010) showed an strong correlation in trend between rainfall and temperature ranges given the periodic nature of these variables. FENG et al. (2016) correlated water balance components such as PET with rainfall occurrence and proposed an annual rainfall ARIMA model with acceptable accuracy. VALIPOUR (2016a) developed 3 models, for 4 climate conditions based on precipitation volumes capable of estimate monthly rainfall indices. MEDVIGY; BEAULIEU (2012) identified an strong correlation between increments of solar radiation and increases in precipitation variability. Despite the rationality and different exploratory methods on variables selection, studies have been approaching this issue taking in consideration the shortage of available and reliable data.

With this research it was intended forecast the rainfall occurrence in short periods of time with the premise that reducing the variance and rising the bias of the time series could lead to accuracy. To achieve this objective it was firstly determined the ranges of time that the model had to predict, which were from three to seven accumulated days. For each accumulated period it was generated an array containing the time-stamp of the last day of the period, the mean air temperature, the accumulated rainfall, an *boolean* value to determine whether the accumulated precipitation was greater than 5*mm* which is considered the median intensity of a light precipitation (SUN et al., 2006), the mean photoperiod, the soil water content and the average daily H_O, these were the final data that were used as inputs for for the ANNs and are represented by the following array representation:

Tabela 2 - Inputs used for the ANN model

Data Name	Type
Mean air temperature	^{o}C
Unix time stamp	*datetime object*
Rainfall	*mm*
Rainfall success flag	*boolean*
Photoperiod	*hours*
Water content	*mm*
H_O	*mm*

With these new arrays was generated a new data array that were used to create four types of ANNs, one for each year climatic season based on the Unix time-stamp variable of the season change date. Each type of ANN of each place is constituted of 5 ANN, one for every accumulated period($[3,...,7]$*days*) consecutively each ANN had a well established rainfall pattern to predict.

To constraint even more the ANNs task, it was removed the necessity to predict the rainfall volume, by making as target for the model the boolean success flag. The goal with this methodology was to create a filter and in a future research, use as inputs only the time-steps that lead to rain, limiting task of next ANN model, to predict only the amount of rain.

For each ANN the structure used was the multi-layer perceptron (MLP) feed forward, with backpropagation momentum and adaptive learning rate (GDX). The MLP structure usually consists of at least 3 layers, one input layer of which the receptors of the ANN receive external data, one output layer where the problem solution is obtained in this case whether or not it's going to rain. In the middle at least one intermediary layer called hidden layer with undetermined number of neurons, it was used a single hidden layer. To represents the structure a diagram of the ANN is shown in Fig. 2 .

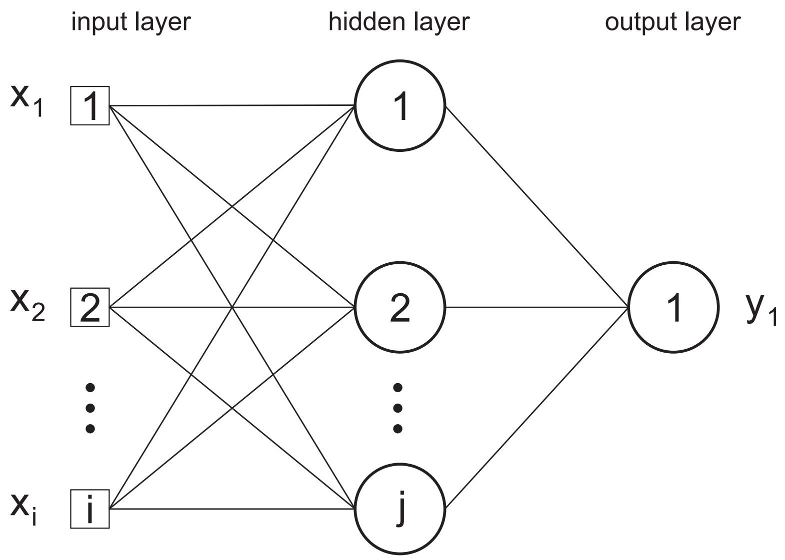


Figura 2 - Feedforward MLP structure

The Mathematical structure of the feed forward multilayer perceptron with one output node can be represented by the following equation (LUK; BALL; SHARMA, 2000):

$$y_1 = S_1(\sum_{j=1}^{Nj} w_j S_2(\sum_{i=1}^{Ni} w_i x_i)) \tag{9}$$

where y_1 is the output($[0,1]$) of the network, x_i is the input array (Fig. 2), w_i the connection weights between the data node and the hidden layer, w_j is the connection weights from the hidden layer to the output layer, S_1 is the activation function from the Input layer to the hidden layer, S_2 is the activation function from the hidden layer to the output layer.

One important decision in designing an backpropagation architecture is the selection of a proper activation function. The activation, or transfer functions are characterised by ruling the behaviour of output for each ANN node. They are a set of equations that have an limited amplitude and are the non linear transformation that is done over input signal (KARLIK; OLGAC, 2011). Sigmoid functions have a nonlinear nature and are widely implemented on backpropagation algorithms, they are easy to distinguish and can interestingly minimize the computation time for training and have an nonlinear output (HECHT-NIELSEN, 1992; KARLIK; OLGAC, 2011). Tangent sigmoid functions are a scaled version of a sigmoid function that solves the problem of values having the same signs. They have an steeper gradient with the advantage that that negative inputs will be mapped strongly negative and the zero inputs will be mapped near zero, this characteristics makes it suited for classification problems.

It was used two different activation functions, one tangent sigmoid (Eq.10) on S_1 and a hard limit (Eq.11) function on S_2. The reason for a hard limit transfer function was the definition of a binary target or an boolean value, in which the ANN would have only two forecasting possibilities.

$$f(x) = \frac{\sinh x}{\cosh x} \tag{10}$$

$$f(x) = \begin{cases} 1 & if\ x > 0 \\ 0 & else \end{cases} \tag{11}$$

Determining the number of neurons in the hidden layer for a time-series problem is not an easy task (ZHANG; PATUWO; HU, 1998), firstly the hidden layer of each ANN had 200 neurons, then it was observed its forecasting accuracy and processing cost, then it was lowered to 50 without noticeable performance lost. With the results this was the number of neurons used and this parameter was not change in any of the ANNs in order to facilitate performance analysis and comparison.

The backpropagation method is a technique used to update the nodes weights in supervised training ANNs. It is consisted of two passes throughout the different layers of the network, a forward pass and a backward pass. In the forward pass all the connection synaptic weights are fixed and a activity patterns is applied to the input nodes, then it propagates layer by layer, node by node producing a output signal as the network response. During the backward pass all the weights are corrected by an error-correction rule, that tries to minimize an error function, it was used the MSE (Eq. 2), this is done by subtracting the actual ANN response by the desired response producing an error signal. All the network weights are backwardly adjusted to make the output closer to the desired one in a statistical sense (DAO; VEMURI, 2002).

At the first learning epoch of the ANNs the first weights has to be randomly distributed within the $[0,1]$ limits, this first random distribution was a problem. The computer is a deterministic machine and to generate random numbers by a deterministic machine a pseudo random number generator is needed. A random generator is an algorithm that produces numbers or vectors that its properties approximates of truly random numbers, this algorithm usually has a seed parameter that uses the computer clock, which can lead to an normal distribution of the random numbers, for this reason sometimes it was required to run the training processes several times. After the first weights distribution, the equation that defines the weights adjustment for each iteration w_{n+1} of the algorithm was:

$$w_{n+1} = w_n - \alpha_{n+1}\, g_n + \mu\, w_{k-1} \tag{12}$$

where g_n is the gradient of the error to the weight vector, α is the learning rate and μ is the momentum constant. The momentum term is used to avoid the weight adjustment to be stuck in the local minima and reduce the algorithm instability (HAYKIN;

NETWORK, 2004), the μ value must be variate between 0 and 1 but it is recommended to use values between 0.4 and 0.9 (WYTHOFF, 1993). An low μ value increases the risk to the ANN get stuck in the local minima and a excessively high value might make the model surpass the problem solution, it was used for all the ANNs an value of 0.9.

Other particularity of the model, despite the back propagation and the μ constant, was the use of variable learning rate, the learning rate is a parameter used in the back propagation stage to define the conversion speed to the minimum solution. Setting the *lr* too high the algorithm would converge too fast making it unstable, setting too low would make it to take too long to find the minimum solution or even never find it. To optimize the forecasting problem, the ANN uses an larger α when it is far from the solution and progressively decreases it while it gets closer by the use of the Eq. 13 and Eq. 14.

$$\alpha_{n+1} = \beta\,\alpha \tag{13}$$

$$\beta = \begin{cases} 0.7 & if\ \frac{error_n}{error_{n-1}} > 1.04 \\ 1.05 & if\ \frac{error_n}{error_{n-1}} < 1.04 \end{cases} \tag{14}$$

Having been determined the basic ANNs structure, we had to choose how many steps before should be appended in each input array to be computed by the ANN to forecast one step further or $t+1$, which were call by time-steps, these time-steps are the amount of lagged arrays (Fig. 2) that should be used as inputs for the ANNs, this concept is shown in the Fig. 3 which represents one time-step array, two time-steps array, up to the time-series length (*n time-steps*). To optimize the lag determination it was made an correlation analysis, for each place and accumulated period, just as in the time-series missing data recovery, that was done autonomously by the algorithm and was set to select only an number of time-steps that had an ρ value bigger than 0.5. This time-step parameter ranged from $t-1$ in the least correlated vectors up to

$t-4$ in the most correlated vectors.

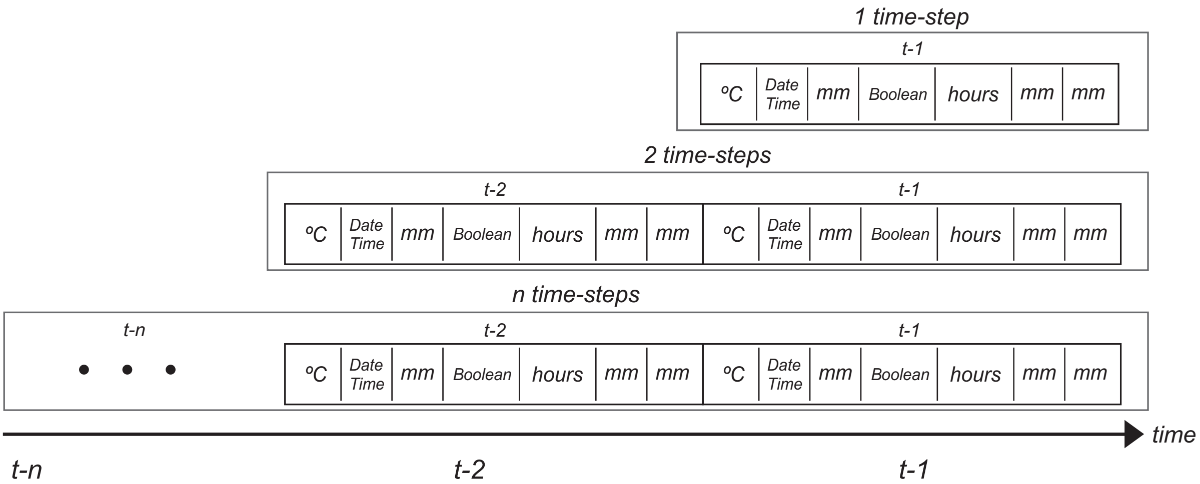


Figura 3 - Diagram of the time-steps concept

ANNs training become more efficient if certain preprocessing steps are made on data. Normalisation is crucial to prepare data to made it suitable for training, without this step training would be slow and ineffective. In order to minimize bias into each input feature that have widely different scales, this process is made to scale down data into a similar range (YALDI et al., 2009). There are many types of normalisation procedures such as statistical normalisation, that produces data where each feature has a zero mean and a unit variance and Min-Max normalisation that rescales features from one range to a new one depending on the type of activation function

To keep data inside the constraints of the tangent sigmoid transfer, fitting it into the $[-1,1]$ interval and making it proper for training, the dataset was normalised by the Min-Max normalisation method demonstrated by the Eq. 15. Then the algorithm was set to run and train all the ANNs models. It was generated 200 individual rainfall forecasting ANNs based on the described methodology, the results of this research are the accuracy of each individual ANN.

$$Z_i^p = -\left(\left(\frac{-2(u_i - x_i^p)}{u_i - l_i}\right) + 1\right) \tag{15}$$

3 RESULTS AND DISCUSSION

Brazil is a country of continental dimensions with contrasting climates, which were represented by the chosen locations. The World Meteorological Organisation (WMO) establishes the general procedures to calculate the monthly 30 year standard normals and averages (WMO, 1989), which are important climatological variables that describes the climatic conditions of a given location. This index were used to contradistinguish the high variability of climate conditions that the ANN structures had to handle. The two opposite climate conditions were Campo Mourão and Jaguaruana, the normals of both locations are represented by Fig. 4.

(a) Campo Mourão (b) Jaguaruana

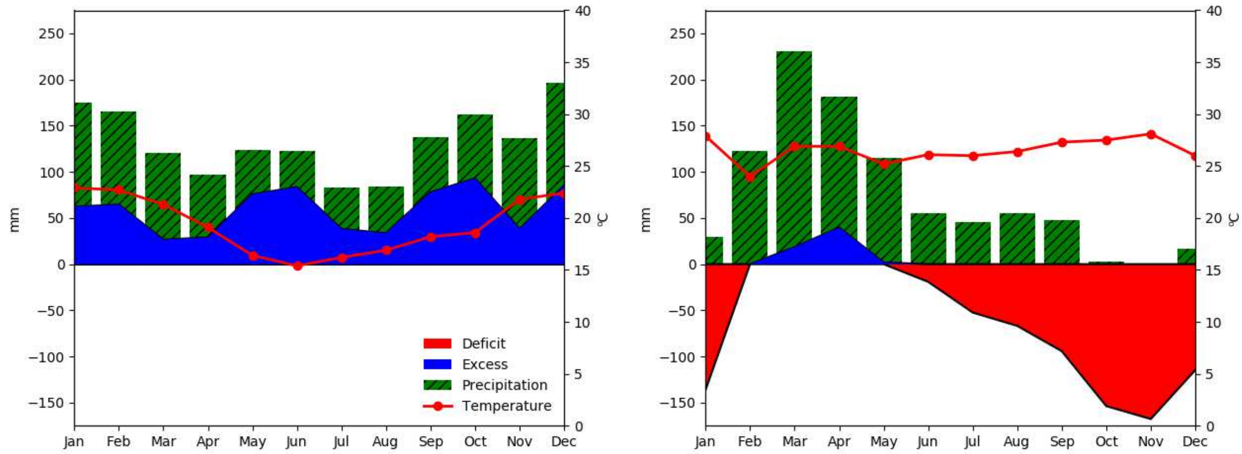


Figura 4 - Thornthwaite Water Balance and Normal Temperature

Campo Mourão have an subtropical humid mesothermal weather with hot summers and not frequent frosts, the precipitation is well distributed with an accumulated volume of 1603 *mm* there is allegedly no water deficit throughout the year, in contrast Jaguaruana have an tropical savanna climate with water deficit across the year with exception of the months from February to May with an accumulated precipitation of 906 *mm*, the location is an good representation of the Brazilian semi-arid region. Between these two contrasting climates there are the climes of all the other locations used in this paper, the climate of each location lays among the Jaguaruana tropical savanna and the Campo Mourão humid mesothermal weather. Given the conditions if it were used only one ANN structure for all locations and seasons the noise would be high, and both the accuracy and precision would decay.

To summarise all the the ANNs assertiveness or the capacity to retrieve information in an general perspective, it was computed the mean accuracy percentage average for all locations for each time range ($[3,...,7]$*days*) and season. In the Fig. 5 it is shown the summarisation both for the *autoassociative* (a) and *heteroassociative*(b) capabilities of the ANNs structures, the auto-association is the phenomenon of associating the input vector with itself as the output as called by estimation capacity, and the hetero-association is that of recalling a related vector given an input vector or the forecasting capacity (RAO; RAO, 2016).

(a) autoassociative (b) heteroassociative

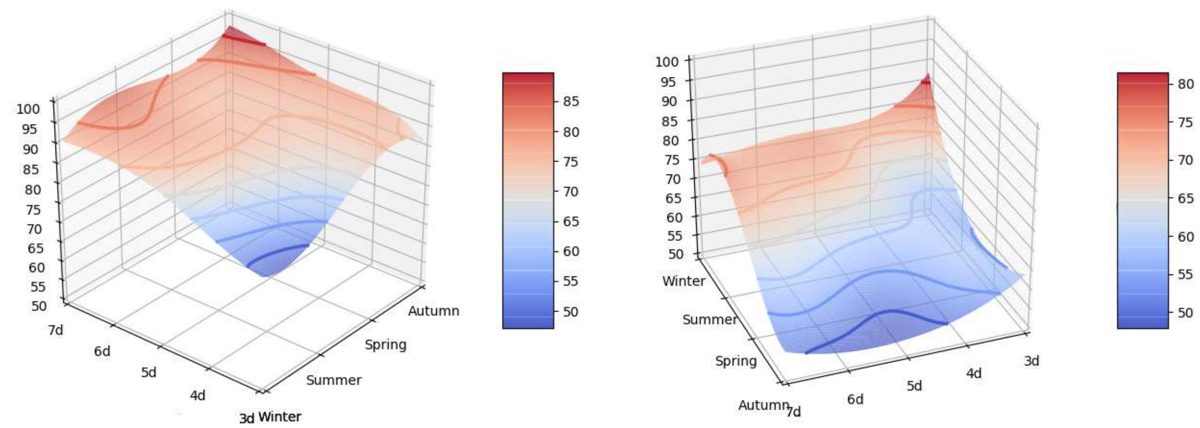


Figura 5 - Summary of ANNs autoassociative (a) and heteroassociative(b) performance

When trying to self associate, the less accurate ANN was the one trained with winter data for a cumulative period of three days and had an accuracy of 77.15% for all regions, The most accurate were trained with autumn data and an period of seven days achieving a accuracy of 97.61% , between all ANNs and cumulative periods the estimation performance average was of 89.18% indicating that the ANN was able to recall the input variables and associate it with the desired output, which is an quality indicative of the chosen inputs variables.

Different from estimation, the forecasting performance had an increased performance variance, in its least accurate point which were in autumn with an accumulated period of 6 days, the ANN structure had an accuracy of 53.14%, when most accurate it had an forecasting success of 87.14% and were in winter with an 3 accumulated days. The Artificial Neural Networks that had the best performance were the ones that had their weights adjusted with data from winter and summer. Relatively to other studies the performance was acceptable, VALIPOUR 2016b while detecting drought and wet years obtained an average correlation of 0.90, in the prediction stage the model was mostly accurate and dependant on the levels of deforestation. RIVERO; PATIÑO; PUCHETA 2015 developed a methodology based on ANNs to forecast rainfall on a monthly period with incomplete datasets, the author utilised the symmetric mean absolute percentage error (SMAPE) as a performance metric, the best performing model had an score of 0.51 which the author classifies as almost acceptable.

In Brazil on latitudes near the equator line like the city of Jaguaruana, winter is the time of year that the rainfall index is usually higher. In summer this index tend to fall, however at lower latitudes this indices reverse and winter happens to be the dry season of the year, in both situations the climate is well defined making it easier for the neural network to generalize its knowledge and accurately forecast the rainfall occurrence, winter in all the accumulated periods was the most predictable season.

Autumn and spring are transitional epochs and there is a mix of climate characteristics both from winter to summer, as from summer to winter. In these seasons the artificial neural networks notably had greater difficulty in forecasting clearly whether or not there would be rainfall, autumn was the least predictable season. Despite the forecast accuracy being smaller in both seasons this is an important result, it indicates that it was wise to create an model for each climate season, if this were not done and the general model have been divided into only 2 times of the year, this effect would have been diluted in the results vector, so that the shape of the forecast chart in Fig 5 would become flattened.

In the first half of the year of southern hemisphere are contained the summer and fall, the inability of the network to generalise its knowledge of autumn would have negatively impacted the summer forecast capacity, in the second half of the year the effect would have been the same with the difference that the forecasting ability of winter would be impacted by the spring. In the Fig. 6 is shown the detailed performance of each ANN with an colour scale that visually represents the relative accuracy of the model, by this figure is clear the predominance of the ANNs models being more accurate both on summer and winter, the combination of location and season with the most notable performance was Maceio on winter and the worst was Campo Mourão on autumn.

Figura 6 - ANNs accuracy percentage performance

Accumulated Period	3	4	5	6	7	3	4	5	6	7
Location			**Spring**					**Summer**		
Jaguaruana	97.22	100.00	100.00	94.44	94.44	68.29	44.44	38.89	52.78	63.89
Maceio	46.15	64.00	52.00	44.83	59.26	68.57	54.29	70.00	58.33	50.00
Diamantino	63.89	75.00	66.67	66.67	61.11	77.78	88.89	86.11	86.11	88.89
Rio Verde	52.78	61.11	55.56	69.44	55.56	69.44	83.33	94.44	77.78	88.89
Uberaba	56.10	56.10	60.98	60.98	60.98	68.29	68.29	78.05	92.68	85.37
Jaboticabal	46.34	51.22	60.98	63.41	60.98	46.34	68.29	80.49	80.49	92.68
Presidente Prudente	60.98	56.10	51.22	60.98	63.41	63.41	70.73	70.73	85.37	87.80
Ivinhema	52.78	52.78	41.67	61.11	66.67	66.67	61.11	63.89	63.89	72.22
Piracicaba	60.98	60.98	51.22	60.98	70.73	56.10	60.98	73.17	65.85	78.05
Campo Mourão	47.22	58.33	61.11	66.67	63.89	58.33	61.11	77.78	90.63	90.63
			Autumn					**Winter**		
Jaguaruana	72.22	66.67	63.89	63.89	58.33	85.37	80.49	70.73	73.17	65.85
Maceio	61.29	51.61	54.84	70.97	68.75	93.55	96.67	100.00	96.77	100.00
Diamantino	58.33	55.56	38.89	44.44	55.56	88.89	88.89	83.33	83.33	80.56
Rio Verde	69.44	63.89	63.89	47.22	55.56	88.89	86.11	80.56	80.56	77.78
Uberaba	53.66	56.10	51.22	46.34	41.46	92.68	82.93	82.93	82.93	78.05
Jaboticabal	80.49	48.78	43.90	51.22	60.98	85.37	78.05	80.49	75.61	73.17
Presidente Prudente	53.66	56.10	51.22	51.22	60.98	70.73	65.85	63.41	63.41	63.41
Ivinhema	63.89	52.78	63.89	58.33	61.11	88.89	58.33	55.56	63.89	63.89
Piracicaba	65.85	56.10	58.54	56.10	63.41	85.37	70.73	75.61	63.41	58.54
Campo Mourão	58.33	55.56	41.67	41.67	38.89	91.67	72.22	81.08	83.33	75.00

When oceanic air masses moves to continent inland they loose water through precipitation and the remaining of this masses become progressively depleted in water vapour, this phenomena can be called continentality effect. By reaching orographic obstacles, the condensation and rainfall associated with the adiabatic cooling of these raising air masses and further deplete the vapor of it, which is called altitude effect (VUILLE et al., 2003). The continentality and altitude effect therefore can be important as sources of rainfall variability over the years and as shown by Fig.7 and Fig. 8 impact on the ANNs prediction performance.

Figura 7 - The effect of continentality on the ANNs accuracy at all seasons.

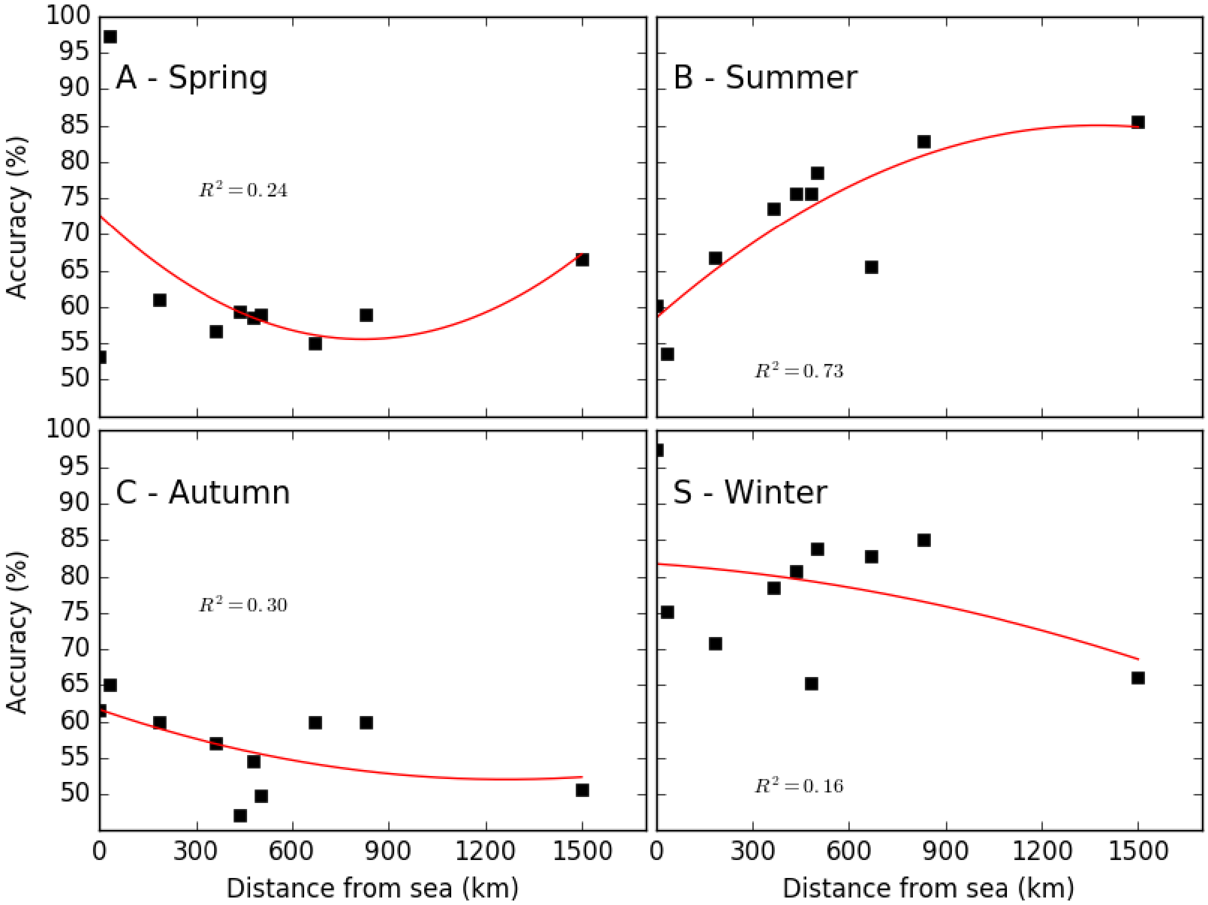
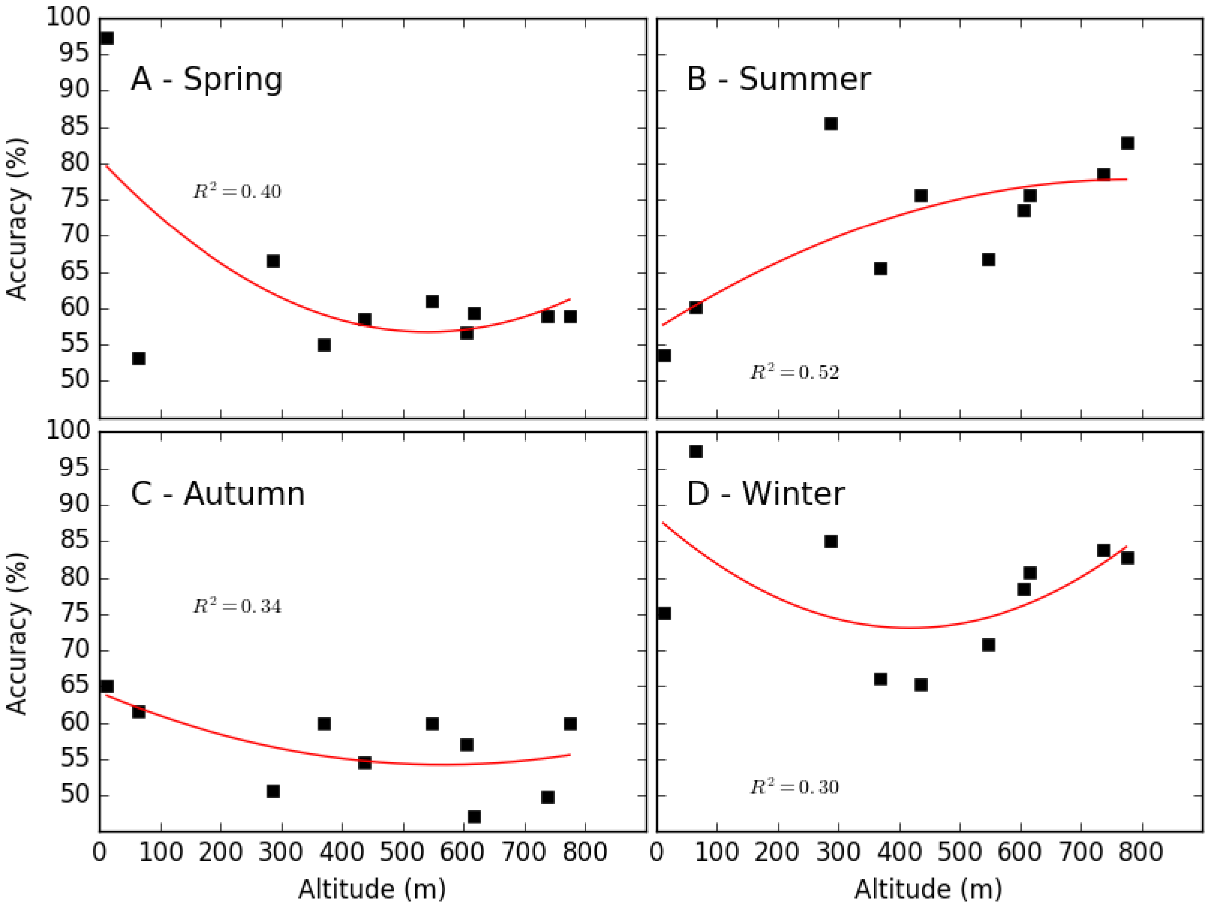


Figura 8 - The effect of altitude on the ANNs accuracy at all seasons.



The Continentality effect was a more prominent source of noise to the ANNs at summer, the closer to the sea the bigger it was the impact on the ANNs accuracy which is actually coherent. Summer is the season that the amount of solar radiation and energy in the atmosphere are higher and consequently the amount of oceanic air masses coming inward are greater. These air masses are highly unstable closer to the sea and tend to loose its strength and stabilise as they move into the continent, the impact of this phenomena on the ANNs forecasting accuracy is represented in the Fig. 7 graph B.

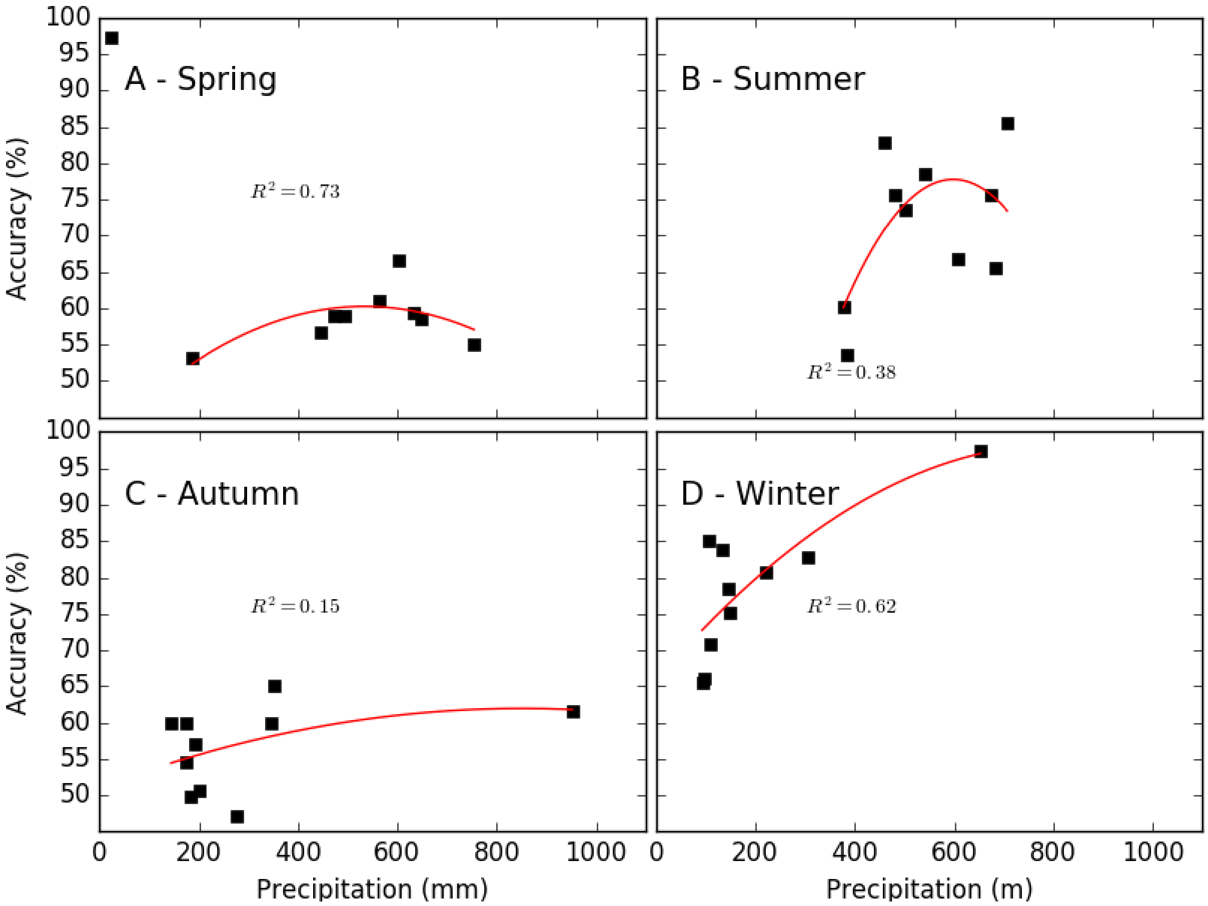
On winter the effect of continentality on the ANNs is quite the opposite of what happens on summer, mainly because the amount of maritime air masses is lower than summer which reduces the climatic variability and reverses accuracy tendency of the ANNs as shown in the graph D of Fig. 7. As spring and autumn are transitional seasons the effect of continentality is not quite well defined, on the first half of spring the oceanic air masses behaves more like the ones of winter and on the second half it starts to behaves as the masses of summer, inverting this behaviour on autumn, the effect of continentality on this seasons are shown on graphs A and C of Fig. 7. Alvares et al. (ALVARES et al., 2013) described an strong correlation of temperature and the effect of continentality during summer and the opposite on winter which corroborate with the result obtained.

The Altitude effect on summer behaves closely as the effect of continentality, there is a large amount of steam loaded air masses coming from sea and an increased amount of orographic rainfalls (SALATI et al., 1979) which is apparently a type of precipitation that the ANNs were able of correctly predict as shown by graph B of Fig 8. The altitude effect is not as prominent on the accuracy of the ANNs on winter (Fig.9 graph D) of which the amount of orographic rainfalls is quite reduced in comparison of summer, and both on spring and autumn it behaves as the continentality effect and by the same reasons. Other studies (GONFIANTINI et al., 2001) on tropical rains described an seasonal variation on rainfall volumes duo to altitude effect being more positive on summer with respects to winter. This seasoned influence is explained duo to the lowering of temperatures and consequent increase of the condensation rate of atmospheric vapour and a greater availability of air moisture on summer when compared to winter .

One factor that can affect directly the ANNs accuracy is the rainfall frequency and volume by itself of which lack of exposure to a significant number of rainfall events can make the ANN underforecast and miss its occurrence (KULIGOWSKI; BARROS, 1998a). As show by Fig. 9 graphs B and D, on winter and summer the ANNs forecasting accuracy is correlated with the amount of rainfall in the period of which the accuracy of the model increases with the precipitation amount. In spring and autumn the precipitation volume do not affect the accuracy of the model.

Comparing the results of this paper with previous studies (KUMARASIRI; SONNADARA, 2006; KULIGOWSKI; BARROS, 1998a; NASSERI; ASGHARI; ABEDINI, 2008; HALL; BROOKS; III, 1999; RAMÍREZ; FERREIRA; VELHO, 2006; RAJURKAR;

Figura 9 - The effect of precipitation volumes on ANN models



KOTHYARI; CHAUBE, 2002) was difficult firstly because of the disparity of ranges of these studies, some have used short forecasting periods on the scale of hours and others used a monthly scale, secondly these studies were focused on predicting not only the rainfall event but also the precipitation volume, which wasn't the goal of this research.

4 CONCLUSION

The objective of this dissertation was to develop an automatic ANN modelling strategy that through the analysis of time series predicts the occurrence of rainfall greater than 5mm for each climatic season for accumulated periods from 3 to 7 days. The study has led to the conclusion that the ANNs can forecast these events with an average accuracy for all the accumulated periods of 78% on summer, 71% on winter, 62% on spring and 56% on autumn. Despite the results, the performance of these models could be improved in future studies, by using training algorithms that are capable of converging on results closer to the global optimum such as training feedforward ANNs with genetic algorithms.

Macroclimatic and mesoclimatic effects, such as the effect of continentality and the effect of altitude as well as the normal precipitation volume, has an direct impact on the forecasting accuracy of the ANNs in well defined seasons. Furthermore despite the relatively lower forecasting performance of transitional seasons, the most important seasons for Brazilian crop production are the summer and winter that are those that the model had best accuracy, nevertheless the results of autumn and spring are still applicable with some limitations. To improve this technic different classificatory algorithms could be implemented, in addition exploratory multivariate statistical procedures, such as principal component analysis or correspondence analysis, would better select input variables. This work can also be applied and improved on other fields of human activities such as transportation and traffic control and being employed to create alert systems given its binary nature. However this type of ANNs structures are suited as an indicative of rainfall eminence and in future studies separate models can be developed to forecast its volume.

BIBIOGRAPHY

ALLEN, R. G.; PEREIRA, L. S.; RAES, D.; SMITH, M. et al. Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *FAO, Rome*, v. 300, n. 9, p. D05109, 1998.

ALOTAIBI, K.; GHUMMAN, A.; HAIDER, H.; GHAZAW, Y.; SHAFIQUZZAMAN, M. Future predictions of rainfall and temperature using gcm and ann for arid regions: A case study for the qassim region, saudi arabia. *Water*, Multidisciplinary Digital Publishing Institute, v. 10, n. 9, p. 1260, 2018.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. de M. Modeling monthly mean air temperature for brazil. *Theoretical and applied climatology*, Springer, v. 113, n. 3-4, p. 407–427, 2013.

BASHEER, I. A.; HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, Elsevier, v. 43, n. 1, p. 3–31, 2000.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C. Time series analysis: Forecastingand control. *San Francisco: Holdenday*, 1976.

BRATH, A. On the role of numerical weather prediction models in real-time flood forecasting. In: *Proceedings of the International Workshop on River Basin Modeling: Management and Flood Mitigation*. [S.l.: s.n.], 1997. p. 249–259.

CALZADILLA, A.; REHDANZ, K.; BETTS, R.; FALLOON, P.; WILTSHIRE, A.; TOL, R. S. Climate change impacts on global agriculture. *Climatic change*, Springer, v. 120, n. 1-2, p. 357–374, 2013.

CAMARGO, A.; SÃO, B. hídrico no estado de. Paulo. *IAC-Boletim Técnico*, n. 116, 1971.

CAMARGO, Â. P. D.; CAMARGO, M. B. P. D. Uma revisão analítica da evapotranspiração potencial. *Bragantia*, SciELO Brasil, v. 59, n. 2, p. 125–137, 2000.

COOPER, P. The absorption of radiation in solar stills. *Solar energy*, Elsevier, v. 12, n. 3, p. 333–346, 1969.

DAO, V. N.; VEMURI, V. A performance comparison of different back propagation neural networks methods in computer network intrusion detection. *Differential equations and dynamical systems*, v. 10, n. 1&2, p. 201–214, 2002.

FENG, G.; COBB, S.; ABDO, Z.; FISHER, D. K.; OUYANG, Y.; ADELI, A.; JENKINS, J. N. Trend analysis and forecast of precipitation, reference evapotranspiration, and rainfall deficit in the blackland prairie of eastern mississippi. *Journal of Applied Meteorology and Climatology*, v. 55, n. 7, p. 1425–1439, 2016.

FERRAUDO, A. S. Artificial neural networks. In: *Nutritional Modelling for Pigs and Poultry*. [S.l.]: CABI, 2014. p. 88–95.

FRENCH, M. N.; KRAJEWSKI, W. F.; CUYKENDALL, R. R. Rainfall forecasting in space and time using a neural network. *Journal of hydrology*, Elsevier, v. 137, n. 1, p. 1–31, 1992.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. *The elements of statistical learning*. [S.l.]: Springer series in statistics New York, 2001.

GHIL, M.; COHN, S.; TAVANTZIS, J.; BUBE, K.; ISAACSON, E. Applications of estimation theory to numerical weather prediction. In: *Dynamic meteorology: Data assimilation methods*. [S.l.]: Springer, 1981. p. 139–224.

GONFIANTINI, R.; ROCHE, M.-A.; OLIVRY, J.-C.; FONTES, J.-C.; ZUPPI, G. M. The altitude effect on the isotopic composition of tropical rains. *Chemical Geology*, Elsevier, v. 181, n. 1-4, p. 147–167, 2001.

HALL, T.; BROOKS, H. E.; III, C. A. D. Precipitation forecasting using a neural network. *Weather and forecasting*, v. 14, n. 3, p. 338–345, 1999.

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

HAYKIN, S.; NETWORK, N. A comprehensive foundation. *Neural Networks*, v. 2, n. 2004, 2004.

HECHT-NIELSEN, R. Theory of the backpropagation neural network. In: *Neural networks for perception*. [S.l.]: Elsevier, 1992. p. 65–93.

JHA, S. K.; SHRESTHA, D. L.; STADNYK, T. A.; COULIBALY, P. Evaluation of ensemble precipitation forecasts generated through post-processing in a canadian catchment. *Hydrology and Earth System Sciences*, Copernicus GmbH, v. 22, n. 3, p. 1957–1969, 2018.

KARLIK, B.; OLGAC, A. V. Performance analysis of various activation functions in generalized mlp architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, v. 1, n. 4, p. 111–122, 2011.

KIM, J.-W.; PACHEPSKY, Y. A. Reconstructing missing daily precipitation data using regression trees and artificial neural networks for swat streamflow simulation. *Journal of hydrology*, Elsevier, v. 394, n. 3, p. 305–314, 2010.

KROGH, A. What are artificial neural networks? *Nature biotechnology*, Nature Publishing Group, v. 26, n. 2, p. 195, 2008.

KULIGOWSKI, R. J.; BARROS, A. P. Experiments in short-term precipitation forecasting using artificial neural networks. *Monthly weather review*, v. 126, n. 2, p. 470–482, 1998.

KULIGOWSKI, R. J.; BARROS, A. P. Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Weather and forecasting*, v. 13, n. 4, p. 1194–1204, 1998.

KUMARASIRI, A.; SONNADARA, D. Rainfall forecasting: an artificial neural network approach. In: *Proceedings of the Technical Sessions*. [S.l.: s.n.], 2006. v. 22, p. 1–13.

LUK, K.; BALL, J.; SHARMA, A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology*, Elsevier, v. 227, n. 1, p. 56–65, 2000.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943.

MEDVIGY, D.; BEAULIEU, C. Trends in daily solar radiation and precipitation coefficients of variation since 1984. *Journal of Climate*, v. 25, n. 4, p. 1330–1339, 2012.

MINSKY, M.; PAPERT, S. *Perceptrons: An essay in computational geometry*. [S.l.]: Cambridge, MA: MIT Press, 1969.

MISHRA, S. K.; SHARMA, N. Rainfall forecasting using backpropagation neural network. In: *Innovations in Computational Intelligence*. [S.l.]: Springer, 2018. p. 277–288.

MONTANA, D. J.; DAVIS, L. Training feedforward neural networks using genetic algorithms. In: *IJCAI*. [S.l.: s.n.], 1989. v. 89, p. 762–767.

NASSERI, M.; ASGHARI, K.; ABEDINI, M. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Systems with Applications*, Elsevier, v. 35, n. 3, p. 1415–1421, 2008.

PANKRATZ, A. Forecasting with univariate box-jenkins method. *NY: Wiley*, 1983.

PARTAL, T.; CIGIZOGLU, H. K.; KAHYA, E. Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 29, n. 5, p. 1317–1329, 2015.

PREIN, A. F.; LANGHANS, W.; FOSSER, G.; FERRONE, A.; BAN, N.; GOERGEN, K.; KELLER, M.; TÖLLE, M.; GUTJAHR, O.; FESER, F. et al. A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges.

Reviews of geophysics, Wiley Online Library, v. 53, n. 2, p. 323–361, 2015.

RAJURKAR, M.; KOTHYARI, U.; CHAUBE, U. Artificial neural networks for daily rainfall—runoff modelling. *Hydrological Sciences Journal*, Taylor & Francis, v. 47, n. 6, p. 865–877, 2002.

RAMCHOUN, H.; AMINE, M.; IDRISSI, J.; GHANOU, Y.; ETTAOUIL, M. Multilayer perceptron: Architecture optimization and training. *IJIMAI*, v. 4, n. 1, p. 26–30, 2016.

RAMÍREZ, M. C.; FERREIRA, N. J.; VELHO, H. F. C. Linear and nonlinear statistical downscaling for rainfall forecasting over southeastern brazil. *Weather and forecasting*, v. 21, n. 6, p. 969–989, 2006.

RAO, V.; RAO, H. *Learn C++ Neural Networks and Fuzzy Logic*. [S.l.]: Amazon Digital Services LLC, 2016.

RIVERO, C. R.; PATIÑO, H. D.; PUCHETA, J. A. Short-term rainfall time series prediction with incomplete data. In: IEEE. *Neural Networks (IJCNN), 2015 International Joint Conference on*. [S.l.], 2015. p. 1–6.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. *Learning internal representations by error propagation*. [S.l.], 1985.

SALATI, E.; DALL'OLIO, A.; MATSUI, E.; GAT, J. R. Recycling of water in the amazon basin: an isotopic study. *Water resources research*, Wiley Online Library, v. 15, n. 5, p. 1250–1258, 1979.

SCHOOF, J. T.; PRYOR, S. Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of climatology*, Wiley Online Library, v. 21, n. 7, p. 773–790, 2001.

SHANKER, M.; HU, M. Y.; HUNG, M. S. Effect of data standardization on neural network training. *Omega*, Elsevier, v. 24, n. 4, p. 385–397, 1996.

SUMI, S. M.; ZAMAN, M. F.; HIROSE, H. A rainfall forecasting method using machine learning models and its application to the fukuoka city case. *International Journal of Applied Mathematics and Computer Science*, Versita, v. 22, n. 4, p. 841–854, 2012.

SUN, Y.; SOLOMON, S.; DAI, A.; PORTMANN, R. W. How often does it rain? *Journal of Climate*, v. 19, n. 6, p. 916–934, 2006.

THORNTHWAITE, C.; MATHER, J. Instructions and tables for computing potential evapotranspiration and the water balance, 5th printing. *CW Thornthwaite Associates, Laboratory of Climatology, Elmer, NJ, USA*, v. 10, n. 3, 1957.

THORNTHWAITE, C. W. An approach toward a rational classification of climate.

Geographical review, JSTOR, v. 38, n. 1, p. 55–94, 1948.

TOTH, E.; BRATH, A.; MONTANARI, A. Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology*, Elsevier, v. 239, n. 1, p. 132–147, 2000.

TULARAM, G. A.; ILAHEE, M. Time series analysis of rainfall and temperature interactions in coastal catchments. *Journal of Mathematics and Statistics*, v. 6, n. 3, p. 372–380, 2010.

VALIPOUR, M. How much meteorological information is necessary to achieve reliable accuracy for rainfall estimations? *Agriculture*, Multidisciplinary Digital Publishing Institute, v. 6, n. 4, p. 53, 2016.

VALIPOUR, M. Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Meteorological Applications*, Wiley Online Library, v. 23, n. 1, p. 91–100, 2016.

VUILLE, M.; BRADLEY, R. S.; WERNER, M.; HEALY, R.; KEIMIG, F. Modeling δ18o in precipitation over the tropical americas: 1. interannual variability and climatic controls. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 108, n. D6, 2003.

WALCZAK, S. Artificial neural networks. In: *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction*. [S.l.]: IGI Global, 2019. p. 40–53.

WANG, S.-C. Artificial neural network. In: *Interdisciplinary Computing in Java Programming*. [S.l.]: Springer, 2003. p. 81–100.

WERBOS, P. J. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Tese (Doutorado) — Harvard University, 1974.

WHITLEY, D. A genetic algorithm tutorial. *Statistics and computing*, Springer, v. 4, n. 2, p. 65–85, 1994.

WIDROW, B.; LEHR, M. A. 30 years of adaptive neural networks: perceptron, madaline, and backpropagation. *Proceedings of the IEEE*, IEEE, v. 78, n. 9, p. 1415–1442, 1990.

WMO. Calculation of monthly and annual 30-year standard normals. 1989.

WU, C.; CHAU, K. A flood forecasting neural network model with genetic algorithm. *International journal of environment and pollution*, Inderscience Publishers, v. 28, n. 3-4, p. 261–273, 2006.

WYTHOFF, B. J. Backpropagation neural networks: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 18, n. 2, p. 115–155, 1993.

YALDI, G.; TAYLOR, M. A. P.; YUE, W. L. et al. Improving artificial neural network performance in calibrating doubly-constrained work trip distribution by using a simple data normalization and linear activation function. Ministry of Transport, 2009.

ZHANG, G. *Linear and nonlinear time series forecasting with artificial neural networks*. [S.l.]: Kent State University, 1998.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, Elsevier, v. 14, n. 1, p. 35–62, 1998.

ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, Elsevier, v. 50, p. 159–175, 2003.