

J. Comp. Int. Sci. (2015) 6(2):71-78  
http://epacis.net/jcis/PDF\_JCIS/JCIS-0096.pdf  
jcis@epacis.net  
©2015 PACIS (http://epacis.net)



PAN-AMERICAN  
ASSOCIACION OF  
COMPUTATIONAL  
INTERDISCIPLINARY  
SCIENCES

## Meteorological data mining for climate precipitation prediction using neural networks

Juliana Aparecida Anochi<sup>a1</sup>, Haroldo Fraga de Campos Velho <sup>a</sup>

<sup>a</sup> Laboratório Associado de Computação e Matemática Aplicada (LAC),  
Instituto Nacional de Pesquisas Espaciais (INPE),  
São José dos Campos, SP, 12.227-010, Brazil

Received on June 25, 2015 / Accepted on August 25, 2015

---

### Abstract

Climate precipitation prediction field is a key aspect in meteorology. The precipitation is a variable associated with natural disasters and can cause impacts in the several sectors of society. However, this is a meteorological variable that is difficult to predict because of large spatial and temporal variability. A method based on Artificial Neural Network (ANN) is applied to seasonal climate prediction precipitation in the Central-West region of Brazil, using the method of data reduction by the Rough Sets Theory. The dimensionality reduction process was used to extract relevant information from meteorological data to perform climate prediction. The results obtained in the conducted experiments show the effectiveness of the proposed methodology, presenting prediction similar to climatological situations considered as available observations in the database.

**Keywords:** *climate precipitation prediction; artificial neural networks; data mining; rough sets theory; computational mathematics.*

---

### 1. Introduction

The climate prediction is defined as a set of natural conditions that dominates a certain region obtained by the average behaviour of the atmosphere, such as variability and extreme weather conditions (precipitation, temperature, wind, etc.) in a time interval sufficiently long, which are essential for characterizing a geographic region.

Due to high spatial and temporal variability, precipitation requires greater challenge to climate prediction and weather. New techniques are used to estimate precipitation. In this work, the seasonal prediction of precipitation field, was performed using three types of networks: multilayer perceptron (MLP), Elman recurrent network, and Jordan recurrent network, with supervised training conducted by the error backpropagation algorithm.

Rough sets theory (RST) has been used as a reduction of data size technique. The prediction of the occurrence of extreme weather with reasonable notice, can reduce and, in some cases, eliminate many of the impacts associated with climatic events.

Research on the implementation of data mining techniques for meteorological data has been performed [1, 11, 10]. These research works are based on rough sets theory. Many research on the weather are also under development.

Currently, the volume of data generated in almost all disciplines, particularly in meteorology and climate science, is dramatically increasing. In meteorology, from different data sources are used in weather and climate prediction. Therefore, prediction of weather events is a complex challenge, even more when it include the need for analysis of large

---

<sup>1</sup>E-mail Corresponding Author: juliana.anochi@lac.inpe.br

volumes of data. The size reduction without loss of information observation data is an important subject of research.

The paper is organized as follows: Section 2 addresses meteorological data mining by RST; Section 3 deals with brief review on neural networks; Section 4 describes the study case adopted in this work (climate prediction); In Section 5, the results and discussions are expressed; some final remarks are presented in Section 6.

## 2. Meteorological data mining

Currently, the massive size of data stored in the database exceeds the capacity of traditional data analysis methods. Therefore data mining has become an important research topic, because it provides a set of techniques and tools capable of intelligently and automatically support for analyze of enormous amounts of data, in search of significant knowledge hidden by other irrelevant data.

Data mining is the extraction of hidden predictive information from large databases. It is a powerful technology with great potential to analyze important information in big databases [5].

### 2.1 - Rough Sets Theory

Rough Sets Theory (RST) was proposed in 1982 by Zdzislaw Pawlak [9] as a mathematical theory to treat uncertain and imprecise information, by deriving approximations of a data set. The RST is based on the similarities among objects measured by an indiscernibility relation, which establish that a set of objects are similar (indiscernible) if they hold the same values for all of their attributes. Rough sets have been proposed for a very wide variety of applications. In particular, the rough set approach seems to be important for Artificial Intelligence, especially in machine learning, knowledge discovery, data mining, expert systems, approximate reasoning and pattern recognition.

### 2.2 - Information and Decision System

Rough Sets Theory uses the concept of Information Systems (IS) in which the available data are represented in a table in which the objects are displayed in the rows and the attributes in the columns. An information system may be extended by the inclusion of decision attributes. Such a system is termed a decision system [8].

Formally, an information system is composed of a finite non-empty set  $U$  (Universe) of objects and a finite non-empty set  $A$  of attributes,  $IS = (U, A)$ , so

that, for each  $a \in A$ ,  $a : U \rightarrow V_a$ . The set  $V_a$  is the set of values of  $a$ , that is, the domain of  $a$ .

A Decision System (SD) is an  $IS$  augmented with a decision attribute  $d \notin A$ . Formally,  $DS = (U, A \cup \{d\})$ , where  $d \notin A$  is the decision attribute [8].

### 2.3 - Indiscernibility relation

The indiscernibility relation is used as a measure of similarity among objects. Thus, a set of objects with the same attributes are indiscernible if only if their attributes hold the same values from their corresponding domains. This is a equivalence relation that may be used to treat problems as redundancy of attributes or the existence of irrelevant attributes in the data assigned to only one representative of a class. Let  $IS = (U, A)$  be an information system, then with any  $B \subseteq A$  there is associated an equivalence relation [8].

$$I_A(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}, \quad (1)$$

where  $I_A(B)$  is the indiscernibility relation. If  $(x, x') \in I_A(B)$ , then the objects  $x$  e  $x'$  are indiscernible for any attribute in the set  $B$  [8].

### 2.4 - Attribute reduction

The reduction process is identify equivalence classes, i.e. objects that are indiscernible using the available attributes. Savings are to made since only one element of the equivalence class is needed to represent the entire class. The other dimension in reduction is to keep only those attributes that preserve the indiscernibility relation and, consequently, set approximation [8]. Here, the Rosetta package (<http://www.lcb.uu.se/tools/rosetta>) will be used.

The attribute reduction procedure is performed by the discernibility function  $f_A(B)$  derived from the discernibility matrix which is a symmetric matrix constructed by comparing the attribute values that discern the objects. The attribute representing discernible values are inserted into the matrix. Each entry in the matrix consists of a set of attributes that distinguish a pair of objects  $x_i$  and  $x_j$  expressed by:

$$M_{i,j} = \{a \in B \mid a(x_i) \neq a(x_j)\}, \quad (2)$$

where  $1 \leq i, j \leq n$  and  $n = |U/I_A(B)|$ .

The discernibility function  $f_A(B)$  for an information system  $B$  is constructed by concatenating the subsets  $M_{i,j} = \{a^* \mid a \in M_{i,j}\}$  of attributes in each position of the discernibility matrix  $M$ , through a Boolean function of  $m$  variables that correspond to attributes  $a_1, \dots, a_m$ . The function determines the

minimum set of attributes that distinguish any class among the existing classes [9] and it is defined as:

$$f_A(a_1^*, \dots, a_m^*) = \bigwedge \left\{ \bigvee M_{i,j}^* \mid 1 \leq j \leq i \leq n \right\}, \quad (3)$$

where  $M_{i,j} \neq \emptyset$ .

In this paper, the RST was used to extract relevant information from meteorological data to perform climate prediction, from a reduced data set.

### 3. Artificial Neural Network

In this paper, artificial neural networks were used as predictive models of precipitation variable. ANNs are computational techniques that present a mathematical model inspired on the neural structure of biological organisms, acquiring knowledge through experience. The network has parallel and distributed processing units or neurons, which calculates certain mathematical functions, typically nonlinear. These neurons can be divided into one or more hidden layers interconnected by synaptic weights (connections), which store the knowledge represented in the model, and serve to balance the input received by each network neuron [6].

In this article, three types of neural networks (NN) were used: multilayer perceptron (MLP) and the networks recurrent models Elman and Jordan, with supervised training conducted by the error back-propagation algorithm.

The backpropagation learning error consists of two steps through the different layers of the network: forward and backward steps. For the forward step, an activity pattern (input vector) is applied through the nodes of the network, and its effect propagates on the entire the network, layer by layer. Finally, a set of outputs is produced as the actual response of the network. During the backward step, the synaptic weights are all adjusted in accordance with an error correction rule. The response of the NN is subtracted from a desired output to produce an error signal [6].

The recurrent networks are NN with one or more feedback loops. Given a MLP as the basic building block, the application of global feedback can take a variety of forms. We may have feedback from the output neurons to the input layer. Another possibility is the link among the hidden neurons with the input layer. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior. The recurrent networks extend the capability of modeling temporal data, and are successfully used in processing of temporal data, pattern recognition, and time series prediction [2].

#### 3.1 - Multilayer Perceptron

The architecture of MLP model is the most used and provides robust solutions in a variety of applications, particularly, areas involving classification, prediction, optimization, pattern recognition, and function approximation. Typically, this architecture consists of a set of units forming an input layer, one or more intermediate (or hidden) layer(s) of computational units, and an output layer. It is a supervised neural network. Figure 1 shows the architecture of a MLP network. [6].

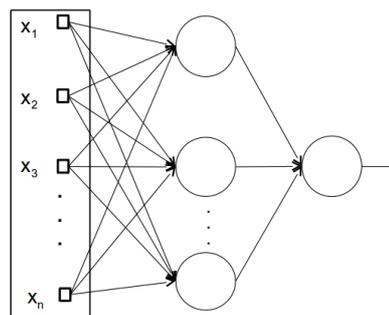


Figure 1 - Multilayer perceptron network with an input layer, a hidden layer, and an output layer.

#### 3.2 - Elman network

The Elman network [4] contains recurrent connections from the hidden neurons to a layer of context units consisting of unit delays. The context unit store the outputs of the hidden neurons for one time step, and then feed them back to the input layer. The hidden neurons thus have some record of their prior activation, which enables the network to perform learning tasks that extend over time. Figure 2 shows the architecture of recurrent Elman network inspired in a multilayer perceptron.

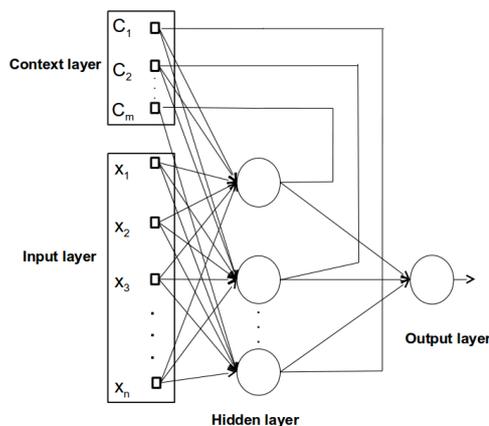


Figure 2 - Elman recurrent network with an input layer, a hidden layer, a context layer, and an output layer.

### 3.3 - Jordan network

The Jordan network [7] first introduced a recurrent network with feedbacks from output units. Jordan network is similar to Elman network. The context units are however fed from the output layer instead of the hidden layer. The context units in a Jordan network are also referred to as the state layer, and have a recurrent connection to themselves with no other nodes on this connection. Figure 3 shows the architecture of Jordan recurrent network.

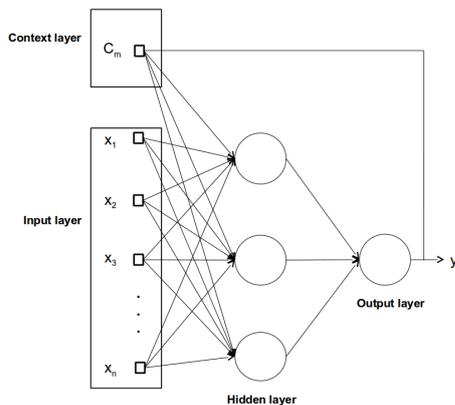


Figure 3 - Jordan recurrent network.

## 4. Case study: climate prediction

Climate prediction is the estimation of the average behaviour of the atmosphere for a future period of time. Thus, the objective of climate prediction is to estimate the statistical properties of the climate in a future period of time. Here, the predictive model developed in this work provides the development of future scenarios that support the studies of impacts and vulnerability and may also enable the preparation of projections of climatic extremes of atmospheric state for seasonal precipitation prediction in Central-West region of Brazil.

The climate in the Central-West is semi-humid tropical, summers are hot and rainy and the winter is cold and dry. This happens between the months of October to March (spring and summer). In the central highlands of parts, the applicant is the tropical climate of altitude.

### 4.1 - Meteorological data

The study area was selected to test the adequacy of the proposed methodology to derive forecasting models from data: Central-West Brazil (Lat 25°S, 7.5°S) to (Lon 62°W, 48°W).

The data consists of monthly means from January 1980 to December 2009. The data was downloaded

from the reanalysis data repository from National Center for Environmental Prediction & National Center for Atmospheric Research (NCEP/NCAR). The global data reanalysis grid uses a set with horizontal resolution of  $2.5^\circ \times 2.5^\circ$  (latitude  $\times$  longitude). The available variables are presented in Table 1.

The training data subset was formed with data from January 1980 up to December 2006. This subset was used to derive a NN model. The period from January 2007 to December 2009 was used for the generalization test subset.

Table 1 - Variables NCEP/NCAR.

Variable	Description	Level
u300	Zonal Wind [m/s]	300 hPa
u500	Zonal Wind [m/s]	500 hPa
u850	Zonal Wind [m/s]	850 hPa
v300	Meridional Wind [m/s]	300 hPa
v500	Meridional Wind [m/s]	500 hPa
v850	Meridional Wind [m/s]	850 hPa
air	Air temperature [ $^\circ$ C]	850 hPa
spres	Surface pressure	surface
shum	Specific humidity [g/kg]	500 hPa
prec	Precipitation [mm/day]	surface

For the visualization of the results we used the GrADS software (Grid Analysis and Display System). This software is available for free download at (<http://www.iges.org/grads/>) [3].

## 5. Experiments and results

The models of MLP, Elman and Jordan networks were treated in two different ways: firstly were trained using all available variables in the database. The second way the networks were trained with reduced number of variables indicated by RST.

In the dimensionality reduction process the relevant attributes are those that mostly occur in the data, in terms of the indiscernibility relation. For the attribute reduction process the training data set previously mentioned were first discretized, then submitted to the reduction algorithm for selection of the relevant attributes chosen as those with a presence greater than 70% in the discernibility function. The data reduction process was used Rosetta software.

The variables that were reduced by the use of RST are show in Table 2. It is observed that of 6 variables out of 10, were considered relevant for Central-West region. These variables were used as initial conditions for the climatic prediction process, and the variable precipitation is the target.

Table 2 - Variables reduced by RST.

Variables	% ( $f_A$ )
airt	72
u300	74
u500	79
v300	73
v500	75
spres	78

In order to evaluate the performance of the three models, the mean square error (MSE) is used:

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_k - d_k)^2, \quad (4)$$

where  $N$  is the number of patterns in the data set,  $y_k$  is the true observational value, and  $d_k$  is the estimation computed by the neural model.

Table 3 shows the mean square error obtained for the three models MLP, Elman, and Jordan, that were trained with all available variables in database for summer season 2008. Indeed, the recurrent Elman network had the lowest squared than the other network models, which shows that has better performance.

Table 3 - Mean square error.	
Model network	MSE
MLP	0.115
Elman	0.002
Jordan	0.061

Table 4 shows the mean square error obtained for the three networks MLP, Elman, and Jordan, that were trained with a reduced set of variables by RST. From the errors showed in Table 4, clearly the best result was obtained by Elman network with the lowest squared error.

Table 4 - Mean square error: NN with RST.

Model network	MSE
MLP (RST)	0.066
Elman (RST)	0.009
Jordan (RST)	0.137

The observed precipitation for Central-West region of Brazil in Summer 2014 season is shows in Figure 4. The results for the climate prediction for the Central-West region of Brazil are show in Figure 5; Figures 5a, 5b, and 5c show the results of seasonal climate precipitation prediction produced by the MLP, Elman and Jordan networks that were trained with all available variables in database, respectively. Figures 5d, 5e, and 5f show the results obtained by the

MLP, Elman and Jordan networks trained with a reduced set, respectively.

From the results, the Elman network trained with the set reduced by RST present the better performance than the models trained with all available variables. Indeed, both models NN-Elman 5b and NN-Elman(RST) 5e show a high fidelity with real dynamics.

Figures 6a, 6b, 6c show the error maps for three models of neural networks, MLP, Elman, Jordan, that were trained with all available variables in database, respectively. Figures 6d, 6e, and 6f show the error maps obtained by the MLP, Elman and Jordan networks trained with a reduced set, respectively.

The error is computed by:

$$\text{error} = |y_k - d_k^m|, \quad (5)$$

where  $y_k$  is the measured precipitation, and  $d_k^m$  is the precipitation calculated by ANN.

Figures 5b, 5c, 5e, and 5f show that all the recurrent networks are reasonably good for predictions in Summer 2008. The MLP network has the worst performance and cannot predict the trend at all (see Figures 5a and 5d).

However, clearly the best prediction was obtained by the Elman trained with the reduced set (see Figure 6e), by using the RST.

## 6. Conclusions

In this article, an approach of data reduction was used to investigate the climate condition in order to predict the precipitation variable, in Central-West region of Brazil. This approach allows the derivation of smaller data sets derived from the resulting reducts for the training phase of the neural network without losing data expressiveness for forecasting purposes. Thus, the RST technique used in the data reduction process allows the identification of relevant information of the data for climate prediction.

The results obtained for the seasonal climate precipitation prediction using the dimensionality reduction of meteorological data clearly showed the best results in relation to networks trained with all available variables in the database. The results have shown that Elman network is significantly better than the other two network models (see Figures 6b, and 6e).

It is seen that Elman network is capable of yielding good results and can be considered as an alternative to traditional meteorological approaches for climate precipitation prediction.

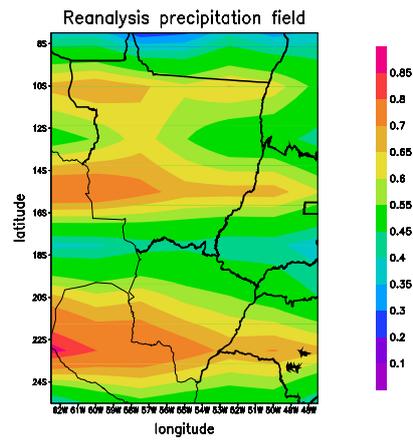


Figure 4: Reanalysis precipitation NCEP/NCAR: Central-West region of Brazil

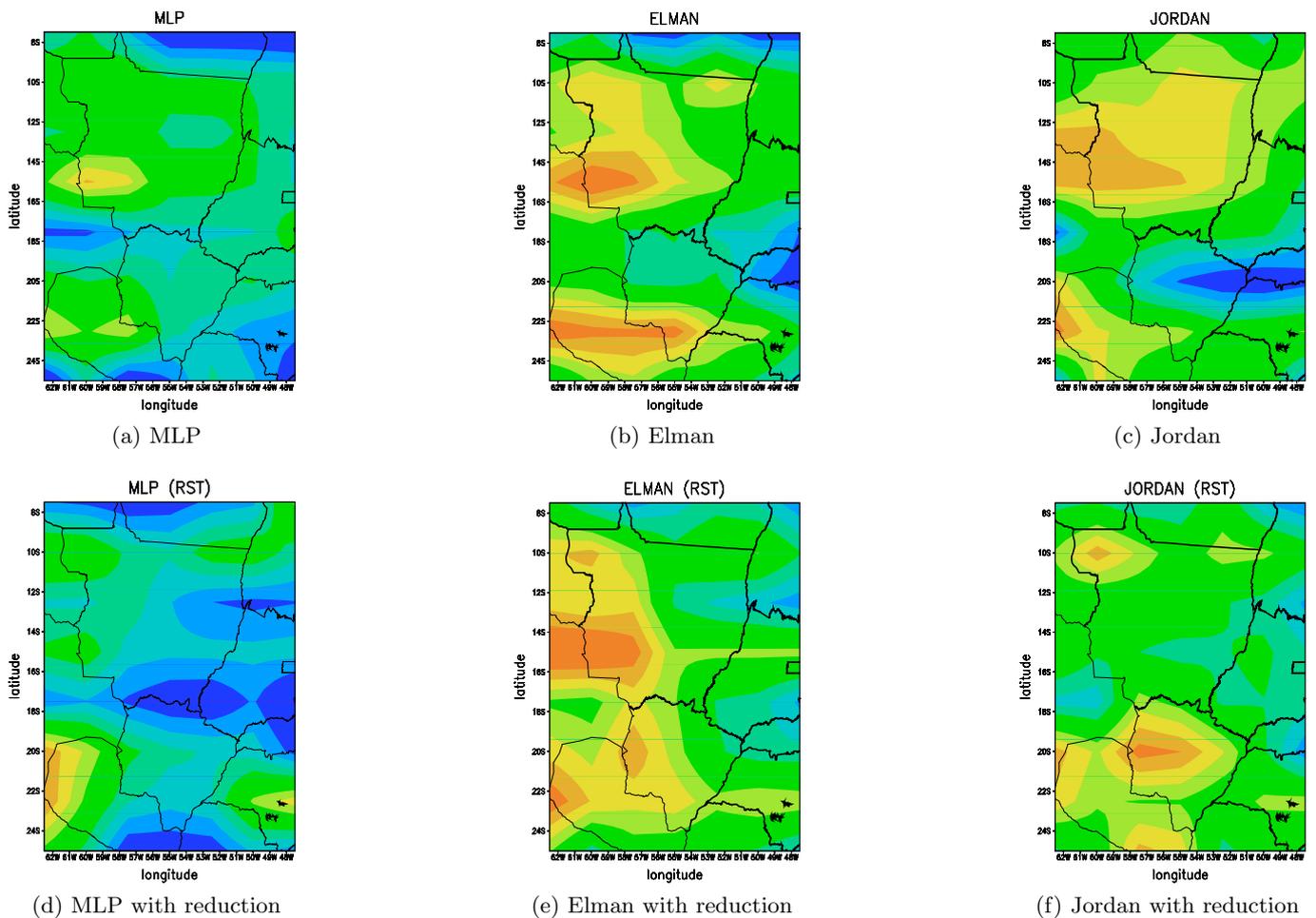


Figure 5: Results for the climate prediction in a Central-West region of Brazil for Summer 2008.

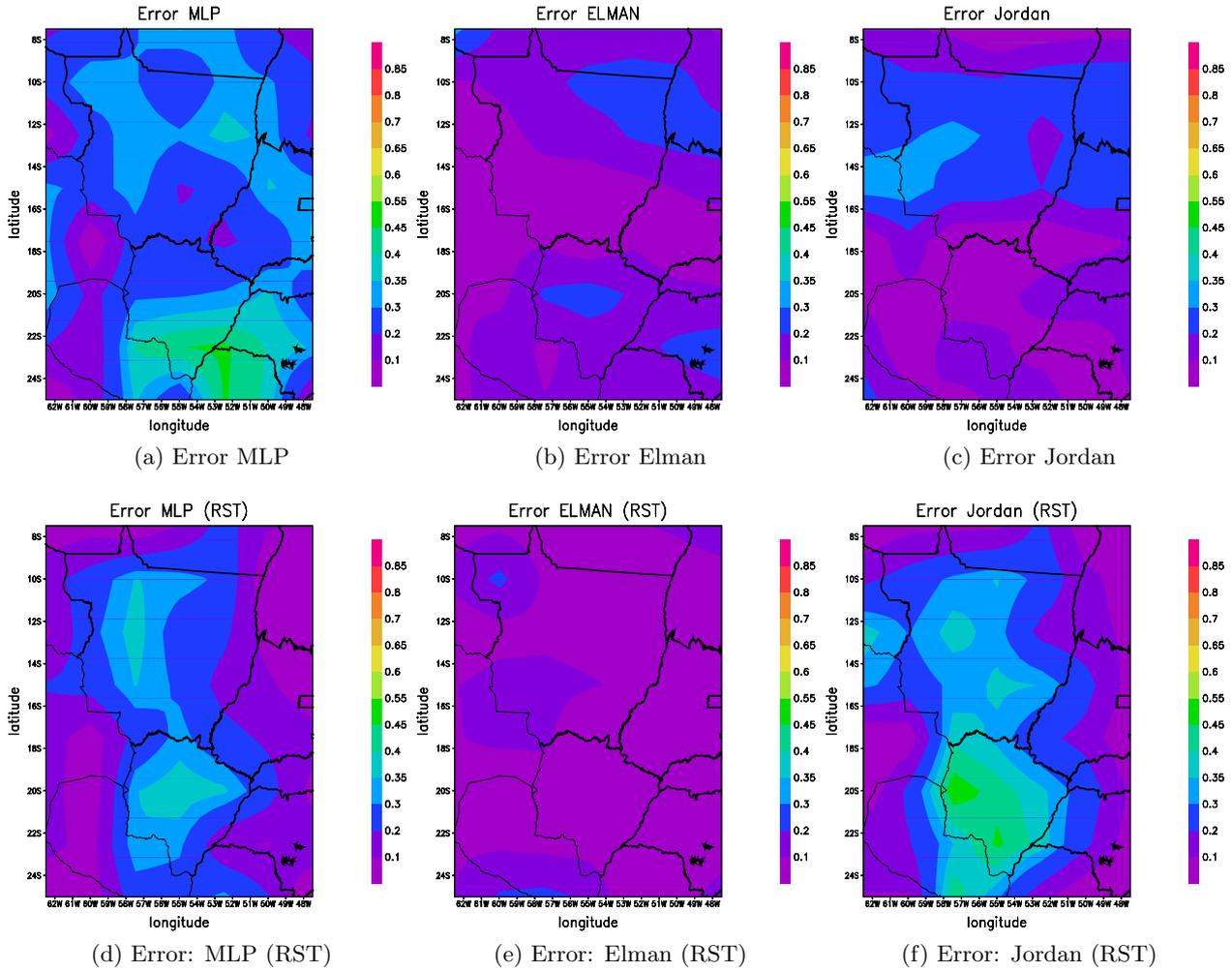


Figure 6: Error maps for the results of climate prediction.

**Acknowledgements:** The authors thank CNPq and CAPES for the physical and financial support provided to this research.

## References

- [1] Aprianti, W., Mukhlash I., *The application of rough set and fuzzy rough set based algorithm to classify incomplete meteorological data*, International Conference on Data and Software Engineering, 10: 1-6, 2014.
- [2] Braga, A., Carvalho, A., Ludemir, T. *Redes neurais artificiais teoria e aplicaes*, 1st. Edition, JC, 2000.
- [3] Doty, B. *Using the grid analysis and display system (GrADS)*, Maryland: Center for Ocean Land Atmosphere Studies (COLA), College Park, MD, University of Maryland.
- [4] Elman, J.L. *Finding structure in time*, Cognitive Science 14:179-211, 1990.
- [5] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. *From data mining to knowledge discovery in databases*, AI magazine, 3:37-54, 1996.
- [6] Haykin, S. *Neural networks principles and practices*, 1st. Edition, Prentice Hall, 1998.
- [7] Jordan, M.I. *Serial order: a parallel distributed processing approach*, Technical report, Institute for Cognitive Science Report 8604, UC San Diego, 1986.
- [8] Komorowski, Z., Pawlak, Z., Polkowski, L., Skowron, A. *Rough sets: a tutorial*, Rough fuzzy hybridization: A new trend in decision-making, 3-98, 1999.
- [9] Pawlak, Z. *Rough sets*, International Journal of Computer and Information Sciences, 1:341-356, 1982.
- [10] Pessoa, A.S.A., Stephany, S. *An innovative approach for attribute reduction in rough set theory*, Intelligent Information Management, 5:223-239, 2014.
- [11] Peters, J.F., Suraj, Z., Shan, S., Ramanna, S., Pedrycz, W., Pizzi, N. *Classification of meteorological volumetric radar data using rough set methods*, Pattern Recognition Letters: 6:911-920, 2003.