Neural Network Models for Climate Forecasting based on Reanalysis Data

Juliana Aparecida Anochi, José Demisio Simões da Silva +

National Institute for Space Research - INPE

Computing and Applied Mathematics Laboratory - LAC

{juliana.anochi, demisio}@lac.inpe.br

Abstract – In this work a neural network model for climate forecasting is presented. The model is built by training a neural network with available reanalysis data. In order to assess the model, the development methodology considers the use of data reduction strategies that eliminate data redundancy thus reducing the complexity of the models. The results presented in this paper considered the use of Rough Sets Theory principles in extracting relevant information from the available data to achieve the reduction of redundancy among the variables used for forecasting purposes. The paper presents results of climate prediction made with the use of the neural network based model.

Keywords – Climate Forecasting, Rough Sets Theory, Artificial Neural Networks, Artificial Intelligence.

1. INTRODUCTION

Climate forecasting is the prediction of the state of the atmosphere for a future time in a given location. This is a task that has been pursued by humans, informally and formally, for a long a time and it is accomplished by acquiring quantitative data about the current state of the atmosphere to be analyzed using scientific understanding of the atmospheric processes in order to project how it will be in the future. Usually weather forecasters use computational models of the atmospheric processes to help in their prediction tasks. However, such models demand high computing performance to be executed in a time that is adequate for society to be prepared for possible consequences that bad climate conditions may bring. This problem is even more complicated when dealing with weather forecasting tasks that require higher accuracy once it deals directly with every-day life of the population.

In order to cope with the demand for computing performance, it is necessary to assemble supercomputers or clusters of powerful computers that may work together to produce numerical results that are inputs for forecasting tasks. Nevertheless, it is not only the computing power that counts in a climate forecasting task. Existing knowledge plays an essential role in the whole process that may significantly balance forecasters decisions.

In these regards, it would be very profitable if computational models were constructed based on the existing knowledge. This would permit forecasters to run the climate model for a several different initial conditions thus leading to a variety of different scenarios to be analyzed to support their decision on the prediction of the state of the atmosphere. In order to achieve a computational model that may satisfy the stated conditions, we proposed the use of artificial neural networks to be trained with existing climate data.

However, a drawback for this approach is related to the huge amount of data that may be available for the training processes of the neural network. Thus, in this work we present a method based on Rough Sets Theory that may reduce the number of attributes used in the climate prediction modeling process.

The data reduction 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 rough sets technique used in the data reduction process allows the identification of relevant information of the data for climate prediction. In addition, it is a technique that approaches the problem of dealing with huge amounts of data which is a characteristic of the processes in meteorology. Huge amounts of data in meteorology result from the increase in the number, the variety, and the availability of satellite sensors as well as ground based networks of sensors that have been trying to acquire enough information to better monitor and understand the atmospheric processes. These data need and have been considered in the development of forecasting models which may increase the complexity of the associated computational models.

Nevertheless, due to the complexity of natural phenomena the availability of data is fundamental for a better understanding of the atmospheric phenomena that may significantly impact on the environment and human life on the earth. The better the availability of data, the greater is the chance for humans to map the atmospheric processes and to increase the probability to predict future status for sake of the populations.

In this paper, we consider the use of historic data in the creation of patterns of the atmosphere that may be used to construct an empiric climate forecasting model. The model results from training a supervised neural network to learn the map between the status of the atmospheric processes at a certain time with the future status the variables commonly used in forecasting reports: precipitation and temperature. The results presented in this work are restricted to precipitation.

In order to construct the model the existing data is first preprocessed by rough sets based technique to build the reducts thus identifying the relevant attributes for forecasting purposes that reduce the complexity of the resulting model.

The main advantage of the methodology presented in this paper is the possibility to derive climate forecasting models from simplified data sets gather from observations and to have a means to make prediction of the climate behavior using simplified non-complex computational models based on neural networks. An additional advantage of the approach is the possibility to derive a model without a need for deep understanding of the physics related to atmosphere processes.

The conducted experiments used reanalysis data available at National Center for Environmental Prediction and The National Center for Atmospheric Research (NCEP/NCAR). Two subareas of the South America were selected: Northeast and North Brazil.

The remainder of this paper is organized as follows. In Section 2 we briefly introduce the climate forecasting problem. In Section 3 we present an introduction to Rough Sets Theory. In Section 4 we briefly discuss the neural network chosen for the forecasting model. In Section 5 we discuss the conducted experiments and describe the data used. In Section 6 we make some conclusions.

2. CLIMATE PREDICTION

Climate forecasting is the estimation of the average behavior of the atmosphere for a future period of time. For instance, in making a seasonal climate forecast, one may study if the next season (e.g. winter) will colder than the average, or else, if there will be more rain than in the previous season. Thus, the objective of climate forecasting is to estimate the statistical properties of the climate in a future period of time [1].

In Meteorology weather and climate are distinct concepts related to the study of the atmosphere. Weather studies are related to a certain geographic region in a very short time. This is a very complex task due to the inherent dynamics of the physics of the atmosphere that may directly affect human life on the planet. Climate studies search for understanding the average behavior of the atmosphere in a long period of time by integrating the atmospheric conditions to represent an abstract characterization [2].

Climate prediction centers conduct climate forecasting using models that try to describe the behavior of the physical-chemical conditions of the atmosphere. Such models require high performance computational resources to generate possible future status of the atmosphere in high resolution scales. The outputs of such models and systems are used to support the studies of impacts and vulnerabilities, and to permit the development of atmospheric extreme conditions. The models may be local, regional or global in scope.

3. ROUGH SETS

Rough Sets Theory (RST) was proposed in 1982 by Zdzislaw Pawlak as a mathematical theory to treat uncertain and imprecise information, by deriving approximations of a data set [3]. Roguh Sets are 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. The indiscernibility relation allows the derivation of subsets of variables that maintain the existing information by considering the most relevant attributes of the data, thus it reduces possible redundancy inherent to the data.

An important feature of Rough Sets Theory is the possibility to derive decision rules as a result of the quantitative and qualitative processing of the uncertainty in the available data.

In this work, the experiments were conducted using the ROSETTA system (Rough Set Toolkit for Analisys of Data). The ROSETTA system is composed of a set of algorithms for data analysis using Rough Sets. It was developed as a support tool for data mining and knowledge discovery in data bases. The system generates the reducts and rules for analysis [4].

3.1 Information Systems

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 [5]. 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 (DS) is an IS augmented with a decision attribute $d \notin A$. Formally, $DS = (U, A\{d\})$, where $d \notin A$ is the decision attribute [5]. A DS may be deterministic or non-deterministic. A deterministic DS uniquely describes the decisions and actions to be realized when some conditions are satisfied. In the non-deterministic DS the decisions are not determined by the conditions.

Table 1 brings an example of a DS. The decision attribute is Rain that is related to the fact it will rain or not. The set of objects is $U = \{o_1, o_2, o_3, o_4, o_5, o_6\}$, the set of attributes is $A = \{Month, Temperature, Wind\}$, and the decision attribute is $d = \{Rain\}$.

3.2 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.

$$IND_A(B) = \left\{(x, x') \in U^2 | \forall a \in B, a(x) = a(x')\right\}$$ (1)

Table 1: Decision System.

U	Conditional Attributes			Decision Attribute
	$Month$	$Temperature$	$Wind$	$Rain$
o_1	$January$	$High$	$Light$	Yes
o_2	$March$	$Average$	$Moderate$	Yes
o_3	$July$	Low	$Strong$	No
o_4	$September$	$High$	$Light$	No
o_5	$March$	$Average$	$Moderate$	No
o_6	$January$	$High$	$Light$	Yes

Table 2: Indiscernibility Relation.

$U/IND_A(Month)$	$\{o_1, o_6\}, \{o_2, o_5\}, \{o_3\}, \{o_4\}$
$U/IND_A(Temperature)$	$\{o_1, o_4, o_6\}, \{o_2, o_5\} \{o_3\}$
$U/IND_A(Wind)$	$\{o_1, o_4, o_6\}, \{o_2, o_5\}, \{o_3\}$
$U/IND_A(Month, Temperature)$	$\{o_1, o_6\}, \{o_2, o_5\}, \{o_3\}, \{o_4\}$
$U/IND_A(Month, Wind)$	$\{o_1, o_6\}, \{o_2, o_5\}, \{o_3\}, \{o_4\}$
$U/IND_A(Temperature, Wind)$	$\{o_1, o_4, o_6\}, \{o_2, o_5\}, \{o_3\}$

For a given information system $IS = (U, A)$, for any set $B \subseteq A$ there exists an equivalence relation $IND_A(B)$, as defined in 1, that establish two objects x and x' in U are indiscernible if for the objects x and x' the attribute a in the set B, hold the same value.

Table 2 presents an indiscernibility relation $U/IND_A(B)$ for a subset $B \subseteq A$ in the example presented in Table 1. For subset $B = \{Temperature\}$, the sets of objects $\{o_1, o_4, o_6\}$ and $\{o_2, o_5\}$ will be belong to the same equivalence class because they are indiscernible. Then, if a number of objects are grouped together they cannot be distinguished one from another. Each group of objects is a class. Table 3 presents the classes resulting for the subset $B = \{Month, Temperature, Wind\}$. Class $C1$ includes objects $\{o_1, o_6\}$; class $C2$ includes objects $\{o_2, o_5\}$; class $C3$ includes object $\{o_3\}$; and class $C4$ includes object $\{o_4\}$.

Table 3: Classes for U/IND_A $B = \{Month, Temperature, Wind\}$

U	Conditional Attributes		
	$Month$	$Temperature$	$Wind$
$C1$	$January$	$High$	$Light$
$C2$	$March$	$Average$	$Moderate$
$C3$	$July$	Low	$Strong$
$C4$	$September$	$High$	$Light$

3.3 Attribute Reduction

The reduction process is achieved by deriving subsets Reducts (RED) of attributes that maitain the knowledge in the database. Indiscernible objects are clustered together in distinct classes based on the discernibility matrix. The reducts hold those attributes that preserve the indiscernibility relation taken as the conditional attribute that maintain the same partition based on the classes identified by the decision attribute. The remaining attributes are considered to be redundant if they do not alter the classification if they are removed from the set [5].

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 as it is stated in equation 2:

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

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

Considering the attributes in Table 3 (Month = Mo, Temperature = Te and Wind = Wi) we may derive the discernibility matrix $M_i, j(B)$ as in Table 4. From Table 3 it is observed that classes $C1$ and $C2$ hold different values for the $\{Mo, Te, Wi\}$. Then,

they are inserted in position (1,2) of the matrix. The same happens for classes $C1$ and $C3$. The three attributes are inserted in position (1,3) of the matrix. Position (1,4) of the matrix holds only one attribute because classes $C1$ and $C4$ have only one attribute that hold a different value [6].

The discernibility function $f_A(B)$ for subset B is constructed by concatenating the subsets $M_{i,j} = \{a^*|a \in M_{i,j}\}$ of attributes in each position (cell) of the discernibility matrix M, through a Boolean function of m variables that correspond to attributes $a_1,...,a_m$. The function determines the minimum set of attributes that distinguish any class among the existing classes [3] and it is defined as:

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

Table 4: Discernibility Matrix

	$C1$	$C2$	$C3$	$C4$
$C1$	$\oslash$	Mo,Te,Wi	Mo,Te,Wi	Mo
$C2$	Mo,Te,Wi	$\oslash$	Mo,Te,Wi	Mo,Te,Wi
$C3$	Mo,Te,Wi	Mo,Te,Wi	$\oslash$	Mo,Te,Wi
$C4$	Mo	Mo,Te,Wi	Mo,Te,Wi	$\oslash$

4. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) have emerged as excellent tools for deriving data oriented models because of their inherent characteristic of plasticity that permits the adaptation of the learning task when data is provided. In this regards, the use of an ANN is adequate to derive the forecasting model proposed in this paper. In addition to plasticity, ANNs also present generalization and fault tolerance characteristics that are fundamental for systems that depend on observational data that may be incomplete and slightly different from the data used to derive the models [7].

The problem approached in this paper is the derivation of an ANNs based model for climate forecasting using reanalysis data. For the problem of building prediction models, as proposed in this paper, are used Multilayer Perceptron networks and models of recurrent Elman and Jordan, with supervised training conducted by the error backpropagation algorithm, as shown in Figure 1.
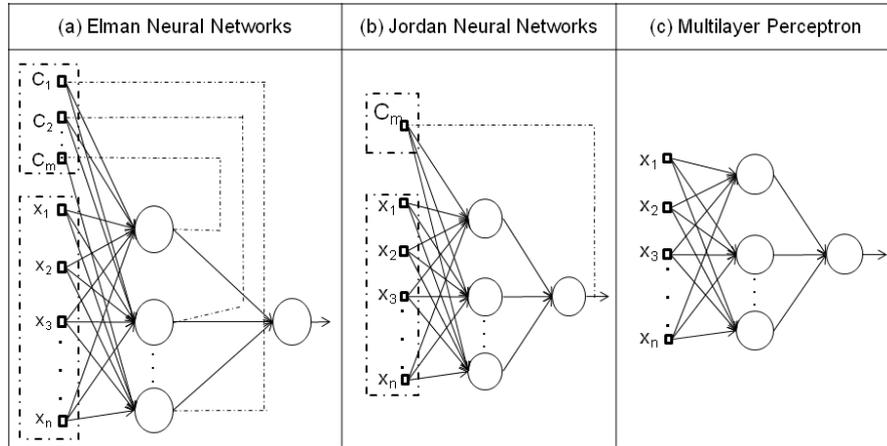

Figure 1: (a) Elman Recurrent Network; (b) Jordan Recurrent Network; (c) Multilayer Perceptron Network.

5. EXPERIMENTS AND RESULTS

Two study areas were selected to test the adequacy of the proposed methodology to derive forecasting models from data: Northeast Brazil (Lat 0°, 17.5°S) to (Lon 47°W, 35°W) and North Brazil (Lat 15°S, 10°N) to (Lon 75°W, 45°W), as it is stated in Figure 2.

Six models of neural networks were constructed following the same methodology presented in [8]: three models are generated from the training of MLP networks, Elman and Jordan, using all available variables in the database, and the other three models (MLP, Elman and Jordan), containing a reduced number of variables indicated by the attribute reduction process. The experiments conducted aimed to verify the adequacy of the reduced training set in the climate forecasting task.

The seasonal climate prediction model developed in this paper has the objective to develop estimates of the statistical properties of some future climate state, which can be for a month or a season. For example, assuming the initial conditions as model

receives data regarding the autumn season, as output will have a forecast for the season ahead, which in this case is the estimate of precipitation for the winter season.
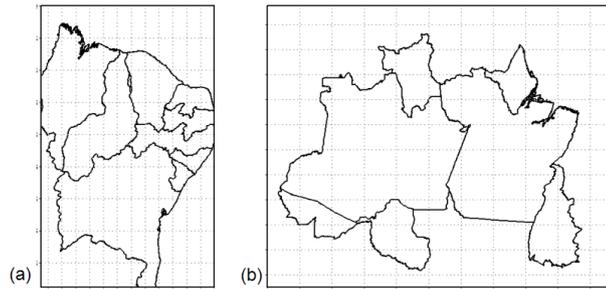

Figure 2: Study areas: (a) Northeast Brazil; (b) North Brazil

In the dimensionality reduction process the relevant attributes are those that mostly occur in the data, in terms of the indiscernibility relation. The data was downloaded from the reanalysis data repository from NCEP/NCAR (National Centers for Environmental Prediction / The National Center for Atmospheric Research) [http://www.ncep.noaa.gov/]. The data consists of monthly means from January 2000 to December 2009 in South America within the geographic coordinates (10°N; 35°S) latitudes and (80°W; 30°W) longitude with a spatial resolution of 2.5°. The available variables are presented in Table 5.

Table 5: Available Variables

	$Variables$	$Descriptions$
1	$airt$	$Air\ temperature$
2	$u300$	$Zonal\ Wind Components\ 300hPa$
3	$u500$	$Zonal\ Wind\ Components\ 500hPa$
4	$u850$	$Zonal\ Wind\ Components\ 850hPa$
5	$v300$	$Meridional\ Wind\ Components\ 300hPa$
6	$v500$	$Meridional\ Wind\ Components\ 500hPa$
7	$v850$	$Meridional\ Wind\ Components\ 850hPa$
8	$spres$	$Surface\ pressure$
9	$shum$	$Specific\ humidity$
10	$prec$	$Precipitation$

The training data set was formed with data from January 2000 to December 2006. This set was used to derive a neural network based model using all the variables and to derive the reducts with the rough set based process. The remaining three years (2007-2009) were used to derive the generalization data set. A testing set corresponding to 25% of the training set was randomly created to be used as an anticipated stopping criterion. In order to evaluate the performance of the forecasting models, we used the mean square error presented as in equation 4.

$$E = \frac{1}{N} \sum_{k=1}^{N} \left(y_k - \hat{y_k} \right)^2 \tag{4}$$

where N is the number of grid points, y_k is the true observational value and $\hat{y_k}$ is the estimation given by the model.

For the visualization of the results we used the GrADS software (Grid Analysis and Display System). This software is used for visualization and data analysis grid points, which provides an integrated environment for access, manipulation and display of data. Software GrADS is a available for free download at [9].

5.1 Resulting Reducts

For the attribute reduction process the training data set previously mentioned was first discretized and then submitted to the reduction algorithm for selection of the relevant attributes chosen as those with a presence greater than 70% (this number was chosen in ad hoc way) in the discernibility function. The resulting reducts for both study areas are presented in Table 6.

Table 6 shows that only 5 variables out of 10, were considered for Northeast and only 7 variables out of 10 were considered for North. It is to be noticed that the variables that form the reducts have a presence greater than 70% in the dicernibility function. These reducts were then used to derive the second forecasting models for both study areas.

5

5.2 Forecasting Results

The climate forecasting problem as presented in this paper is an inverse approach that uses supervised learning to train the neural networks. In order to tune the neural networks we used 10 hidden neurons with sigmoidal activation functions.

The number of neurons was chosen by testing different network architectures in several training generalization processes. Each training phase was set to 10.000 training epochs but all the neural networks stopped their training by the anticipated stopping criterion.

The generalization (or validation phase) of the neural network based models for both study areas was then started. Visual qualitative analysis are possible through Figures 3 to 6. They show the generalization results for the four seasons of the year 2008 produced by both models for each study area.

Table 6: Resultting Reducts

Reduct for Northeast		Reduct for North	
$Variable$	$Presence\%$	$Variable$	$Presence\%$
$airt$	75%	$airt$	74%
$u850$	82%	$u300$	81%
$u500$	71%	$u500$	81%
$v500$	73%	$v300$	76%
$v300$	85%	$v500$	78%
		$spres$	77%
		$shum$	88%

Table 7 presents the forecasting performances of the three neural networks models trained with all the available variables for both areas for the four seasons in the year 2008.

Table 7: Mean square error in the generalization phase

Study areas	Northeast		North	
Models ANNs	$Winter$	$Summer$	$Autumn$	$Spring$
Jordan Network	0.0081	0.0114	0.0808	0.0233
Elman Network	0.0022	0.0037	0.0281	0.0055
Multilayer Perceptron	0.0388	0.0020	0.0668	0.0557

Table 8 presents the forecasting performances of the neural networks models trained with the reducts for both areas for the four seasons in the year 2008.

Table 8: Mean square error in the generalization phase using RST

Study areas	Northeast with the reducts		North with the reducts	
Models ANNs	$Winter$	$Summer$	$Autumn$	$Spring$
Jordan Network	0.0484	0.0071	0.0011	0.2689
Elman Network	0.0003	0.0004	0.0094	0.0274
Multilayer Perceptron	0.0006	0.0037	0.1499	0.1299

Figure 3-a corresponds to the observed precipitation in Northeast in Winter 2008; Figure 3-b corresponds to the estimation produced by the model Jordan trained with all the available variables; Figure 3-c corresponds to the estimation produced by the model Elman using all the available variables in the database; Figure 3-d corresponds to the estimation produced by the model Multilayer Perceptron trained with all the variables; Figure 3-e corresponds to the estimation produced by the model Jordan trained with the data processed by the RST; Figure 3-f corresponds to the estimation produced by the model Elman trained with the reducts; and Figure 3-g correspond to the estimation produced by the model Mutilayer Perceptron trained with the reducts. The same order is followed in the remaining figures (figures 4, 5 and 6).

From the figures we may visually devise that the networks recurrent models trained with the reducts did approximate better to the observations than the models trained with all the available variables.
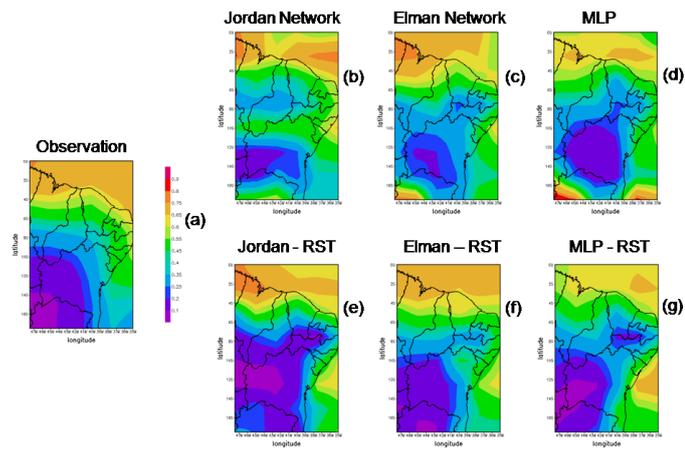
Figure 3: Precipitation in Winter 2008 in the study area of Northeast; (a) actual precipitation; (b) result of Jordan model using all data; (c) result of Jordan model using RST; (d) result of Elman model with all data; (e) result of Elman model with RST; (f) result of the MLP model using all data; (g) of the model result MLP using RST.
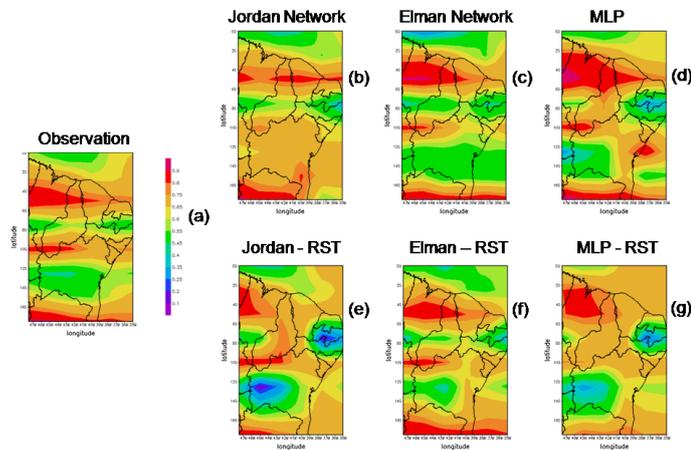

Figure 4: Precipitation in Summer 2008 in the study area of Northeast; (a) actual precipitation; (b) result of Jordan model using all data; (c) result of Jordan model using RST; (d) result of Elman model with all data; (e) result of Elman model with RST; (f) result of the MLP model using all data; (g) of the model result MLP using RST.
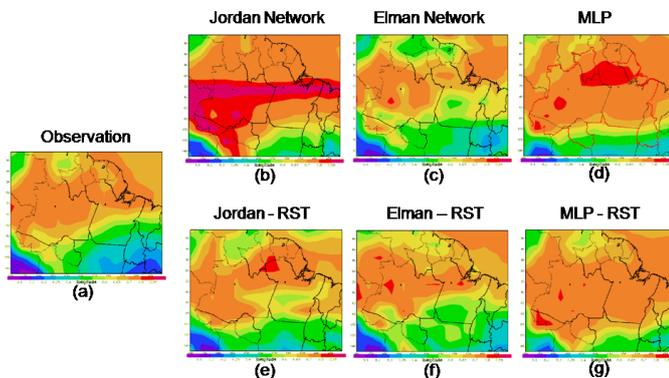

Figure 5: Precipitation in Autumn 2008 in the study area of North; (a) actual precipitation; (b) result of Jordan model using all data; (c) result of Jordan model using RST; (d) result of Elman model with all data; (e) result of Elman model with RST; (f) result of the MLP model using all data; (g) of the model result MLP using RST.
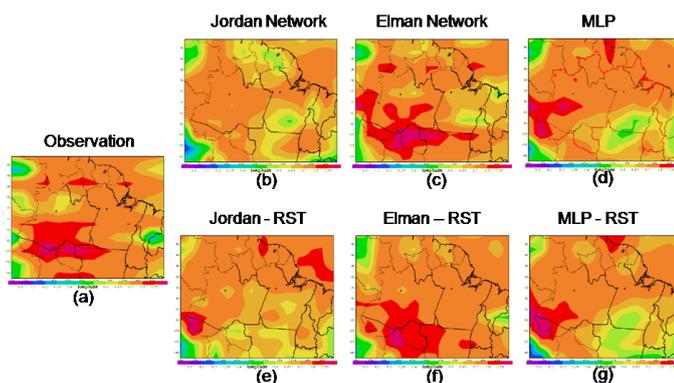
Figure 6: Precipitation in Spring 2008 in the study area of North; (a) actual precipitation; (b) result of Jordan model using all data; (c) result of Jordan model using RST; (d) result of Elman model with all data; (e) result of Elman model with RST; (f) result of the MLP model using all data; (g) of the model result MLP using RST.

6. CONCLUSIONS

This paper presented results of applying supervised neural networks to derive computational models for climate forecasting. The methodology was applied to two different areas located in South America in the Brazilian territory. The mean square errors obtained show the models have approximated the observations in terms of generalization.

In addition, a significant performance was obtained when the training data was preprocessed to eliminate the existing redundancy in the data. The results show that it is worth to search for hidden information in the data in a way that less computational efforts are necessary to construct efficient forecasting models.

The main advantage in using the proposed methodology for deriving models the prediction using neural networks, is the possibility that the technique of the data mining offers to reduce the complexity of the problem and in fact can be used forecasting methods based on models derived directly from data, without consideration of the physics of the problem, allowing the capture of application knowledge. From the conducted experiments in this paper we may visually devise that the results with the recurrent neural networks Elman and Jordan shown that are better than the model multilayer perceptron networks. Therefore the proposed neural networks models shown that are feasible and effective in the climate forecasting field.

However, in this paper we experimentally show the adequacy of a preprocessing phase to eliminate existing redundancy, in a search for accuracy increases. Future work will focus on new neural network architectures and the study of temperature forecasting under the same methodology.

Dedication

I dedicate this work to the lecturer and friend José Demisio Simões da Silva in memoria. I was honored to have been student of one of the greatest Brazilian scientific researchers in Artificial Intelligence.

REFERENCES

[1] I. A. Cavalcanti. "Previsão climática no CPTEC-INPE", April 2009.

[2] R. Vianello and A. Alvez. *Meteorologia Básica e Aplicações*. UFV, Viçosa, 2000.

[3] Z. Pawlak. "Rough sets". In *Proceedings of the International Journal of Computer and Information Sciences*, volume 1, pp. 341–356, Poland, 1992.

[4] A. Ohrn. "Discernibility and Rough Sets in Medicine: Tools and Applications". Ph.D. thesis, Ecole polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suíça, 2001.

[5] J. Komorowski. and A. Øhrn. "Modelling prognostic power of cardiac tests using rough sets". *Artificial Intelligence in Medicine*, vol. 15, pp. 167–191, 1999.

[6] J. Komorowski, Z. Pawlak, L. Polkowski and A. Skowron. *Rough Set: a tutorial*. Department of Computer and Information Science Norwegian University os Science and Technology (NTNU), Polond, 1998.

[7] S. Haykin. *Redes Neurais Princípios e Práticas*. Bookman, Porto Alegre, second edition, 2001.

[8] J. Anochi. "Modelos baseados em redes neurais para o estudo de padrões climáticos sazonais a partir de dados tratados com a teoria dos conjuntos aproximativos". Dissertation, Intituto Nacional de Pesquisas Espaciais, 2010.

[9] B. Doty. "Grid Analysis and Display System (GrADS)", February 2009.