International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 13 (2018) pp. 11289-11294

Data mining and neural networks for the analysis of the correlation coefficient in C-band meteorological radars

Julián R. Camargo L[1]., Ernesto Gómez Vargas[2] and César A. Perdomo Ch.[3]

[1,2,3] *Engineering Faculty, Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia*

Abstract

This paper shows the analysis of the correlation coefficient measured by a C-band meteorological radar located in an equatorial zone, using data mining and neural networks. For the analysis, the MatLab computer tool and the Neural Network toolbox were used, which allowed us to generate an output set that detected the presence or not of meteorological targets in a specific area of the radar.

Keywords: Meteorological radar, hydrometeor, correlation coefficient, polarimetric variables.

INTRODUCTION

One of the most valuable uses of radar is the possibility of detecting storms and other meteorological phenomena [1]. These devices send and receive signals that provide valuable information about the location, type of hydrometeors and the intensity of rainfall. Doppler radar technology goes beyond the detection of reflectivity allowing obtaining high resolution data and estimated speed data, which is vital for short-term weather forecasting (used in aviation) and weather prediction in severals conditions [2].

When you observe a radar image, you're looking for the distribution of the precipitation and it´s intensity, also the type of hydrometeors that are present. The information of the intensity of the precipitation called radar echoes is represented graphically by a series of colored pixels, each color has an associated intensity scale that represents what is called the reflectivity in dBZ (reflectivity unit), and another scale which represents the corresponding rate of fall, which is an interpretation of the light or heavy form of precipitation [3].

The main difficulty in measurements with radars is related to the diameter of the drops (which determines the type of hydrometeor), it comes to the use of polarimetric radars. These ones have the ability to emit microwaves with double polarization, which incorporates new variables of measurement, in addition to Z (reflectivity), called polarimetric variables, among these can be included among others: the specific phase difference (KDP), the differential reflectivity (ZDR) and the correlation coefficient (RhoHV).

The first of these variables, KDP, gives an estimate of the specific phase difference between the received signals. This is achieved when the drops are large and are deformed generating a difference of optical paths between the radiation with horizontal and vertical polarization. KDP depends on the size of the hydrometeors, their shape and orientation, when the phase difference is close to zero, it means that the meteorological targets are randomly oriented, when the values are positive it indicates that the targets are wider than high, if they are negative, it indicates the opposite [1].

ZDR is defined as the quotient between the horizontal reflectivity Zh and the vertical Zv that the radar receives providing an estimate of the shape of the hydrometeors. This measurement shows that as larger the value of ZDR, as larger and deformed the drops will be and the closer these values are to one, the smaller and more spherical they will be [4].

The correlation coefficient RhoHV is defined as the correlation in horizontal and vertical polarization signals at a given point in space [1]. Equation (1) defines the correlation coefficient.

$$RhoHV = \frac{\langle s_{VV} s_{HH}^{*} \rangle}{\langle |s_{HH}|^2 \rangle^{1/2} \langle |s_{VV}|^2 \rangle^{1/2}} \tag{1}$$

Where:

s = intensity of the signal

s* = the complex conjugate of the intensity of the signal

The value of the correlation coefficient depends on the characteristics of the targets that are being measured. For example, typical rainfall is in the range of 0.97 to 1. Irregularly shaped hydrometeors such as hail, snow or gravel can have correlation coefficients in the range of 0.8 to 0.95 [1]. Also, the correlation coefficient can be used to separate rain targets from non-rain targets, or separate meteorological targets from non-meteorological targets such as birds and insects.

Table 1 shows a summary of the values generated according to the object detected by the weather radar. As can be seen, many values are overlapped, which indicates that this single parameter is not sufficient to classify the type of target in an accurate way

Table 1. Classification of types of targets according to RhoHV [5]

Type of target	Value RhoHV
Cloud	<0.9
Drizzle	<0.9
Rain	>0.95
Snow (just to seasons places)	0.8 a 0.95
Hail	0.9 a 0.95
Insects	~0.8
Birds	~0.9
Earth echoes	<0.8

In Figure 1, an example of the representation of the RhoHV correlation coefficient in a C-band meteorological radar can be seen graphically.

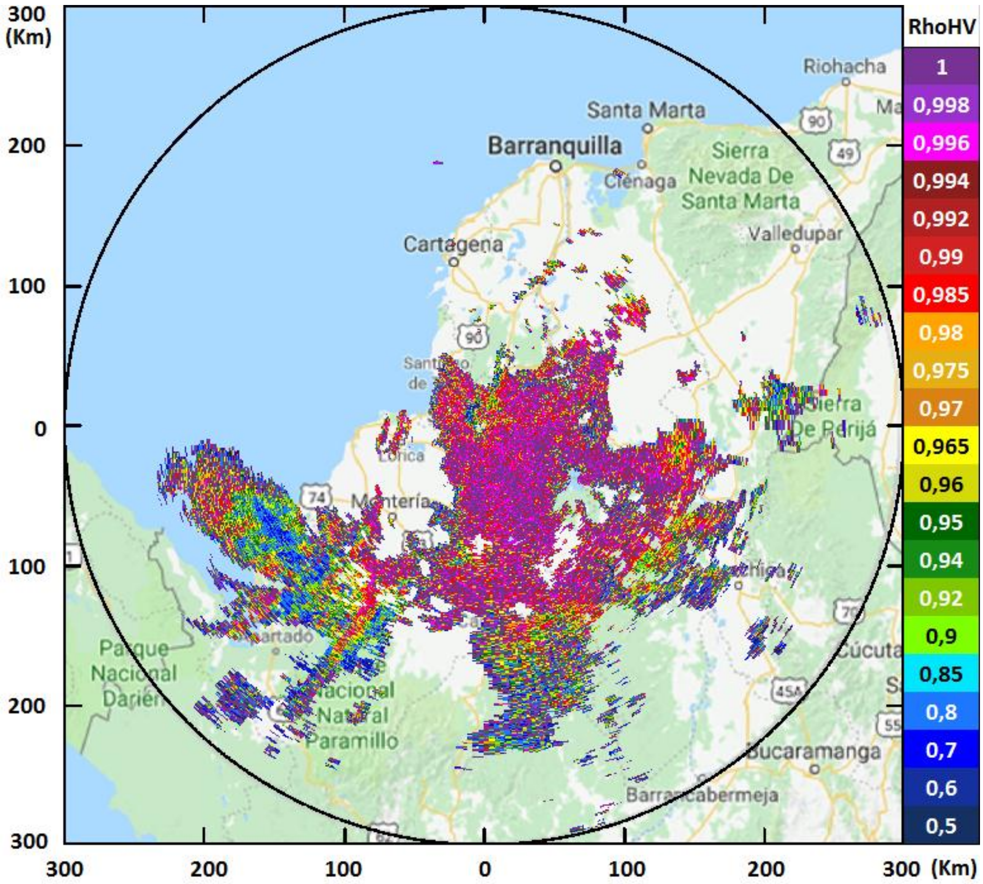

Figure 1. Graphical representation of the correlation coefficient

Table 2. Characteristics of the data measured by the radar [6]

ID	Variable	Units	Fill value	Scale Factor	Add offset
94	Z	dBZ	-128	0.5000	32
96	Z_{DR}	dB	-128	0.0625	0
98	PhiDP	Deg	-128	0.7087	90
99	RhoHV		-128	0.0040	0.5059

CONVERSION OF THE DATA SET

The data generated by the analyzed radar software are stored in proprietary format of the brand of the equipment in a file with extension .NC whose name includes the date and time of the data collection, the file stores 101 variables, which have the format shown in Table 2. Due to the characteristics of the radar beam, there is the possibility of not having the reflection of said beam, this event is represented with a fill value, the number -128.

The information generated by the weather radar is encrypted in a very particular format and it is necessary to use specific software applications to decode it and separate each of the variables generated by the radar.

Commercially, there is a paid software application called Iris that allows the aforementioned decoding, the drawback of the application is its high costs. There is another free software application for Linux called RadX [7] that allows the decoding of data with similar results to the paid application, another useful application for this process is the MatLab computational tool that allows to perform this process directly, in this case, the last one was used to obtain an array of data with each of the variables measured by the radar: Z, ZDR, PHIDP and ROHV.

The decoded data has a format that is not suitable for analysis, so it became necessary to transform it to standard units for each variable. In equation (2) the general mathematical function used to convert to the appropriate format each of the variables in Table 2 is shown.

$$ConvData = (Data * ScaleFactor) + Offset \quad (2)$$

Where Data corresponds to the value to converted, FactorScale is the multiplication factor of the data, Offset the value added to the product between the Data and the FactorScale and

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 13 (2018) pp. 11289-11294
© Research India Publications. http://www.ripublication.com

DatoConv corresponds to the result of the conversion. When a fill value appears, it does not change to remain at -128.

Using the MatLab computational tool as indicated in previous paragraphs, the netcdf.open () function [8] that opens the file with the .NC extension as a read-only file and stores the file is used to start the data conversion process information read in a single variable (in the algorithm developed it is called x). With this data, information is obtained on the number of variables in the file, global attributes and the ID of each of them with the function netcdf.inq() [9].

The next step is separate each of the variables of interest so they can be analyzed individually. For this, the function netcdf.inqVar () [10] is used, which requires fot the input parameters the information stored in a variable (for this case the variable x) and the number of the variable to be separated (corresponds to the ID shown in the Table 2).

The information of the read variable is stored in a matrix by means of the function netcdf.getVar () [11], to complete the process, the adjustment of the data is made using the information shown in equation (2). In Figure 2 you can see the code of the conversion algorithm designed for the conversion of the RhoHV parameter (correlation coefficient).

```
% Correción RhoHV
[varname, xtype, varDimIDs, varAtts] = netcdf.inqVar(x,99);
data5 = netcdf.getVar(x,varid);
ROHV=data5;
for ff=1:F
  if(ROHV(ff,1)==-128)
    ROHV(ff,1)=ROHV(ff,1);
    else
    ROHV(ff,1)=((ROHV(ff,1))*0.00395257)+0.505929;
  end
end
```

Figure 2. MatLab code for the conversion of data for the parameter RhoHV

With this treatment, data matrices of 360x664 (360° x 664cells) were finally obtained for each of the variables analyzed.

ANALYSIS OF THE CORRELATION COEFFICIENT WITH NEURAL NETWORKS

A neural network is defined with an input matrix of 360x664 which, as was indicated above, corresponds to the size of the data measured by the C-band weather radar, specifically for the variable RhoHV. With an output set that validates with a value 0 or 1 the possibility of meteorological targets in a specific area of the radar.

Training with 10 neurons

Figure 3 shows the representation of the neural network making use of the Neural Network tool of Matlab, with the aforementioned parameters, trained with 10 neurons and the time taken for this process.

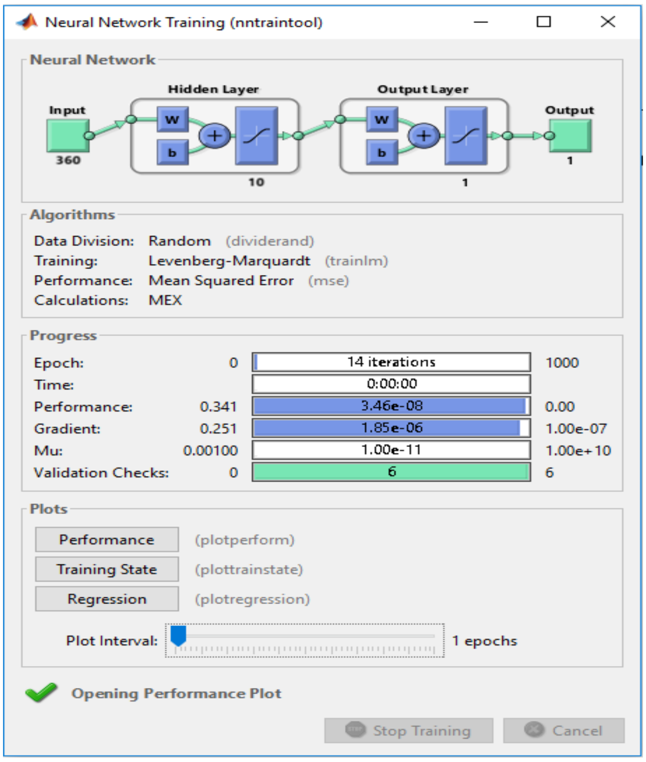

Figure 3. System training with 10 neural networks

As a result of training you can see the margin of error in Figure 4.

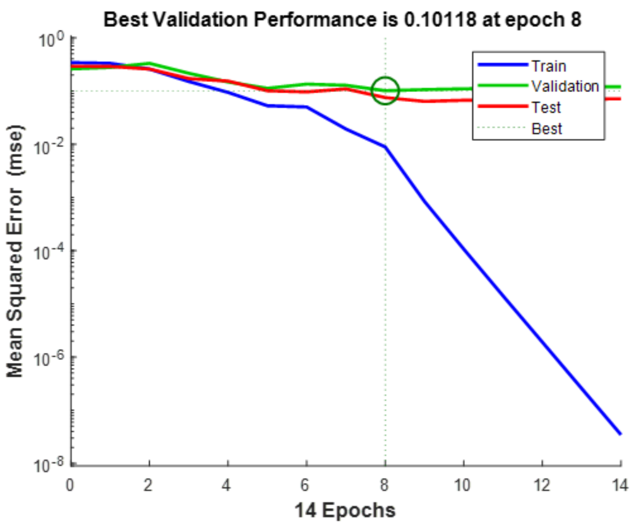

Figure 4. Error generated by training with 10 neurons

Finally, in Figure 5 it can be seen that training and simulation for this particular case present very good results, giving a very high confidence margin of 98% and 84% respectively. Although, the results of the validation reflect a somewhat lower degree of confidence of around 75%.

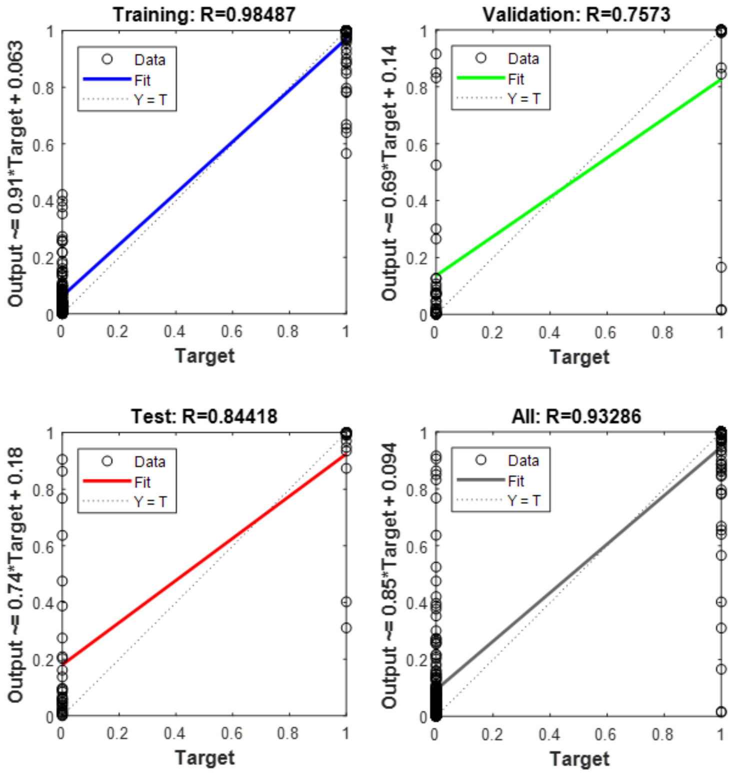

Figure 5. Result of training with 10 neurons

Training with 20 Neurons

Since the expected results were not the desired ones, the neural network is modified by training it with the same data and with 20 neurons, in Figure 6 the modified network and the time taken for the training process can be appreciated.

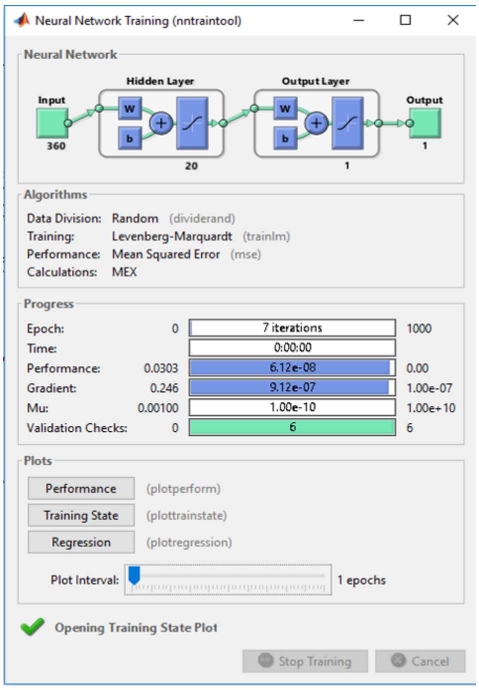

Figure 6. Training of the neuronal network with 20 neurons

The error generated decreases as the number of neurons increases, as can be seen in Figure 7.

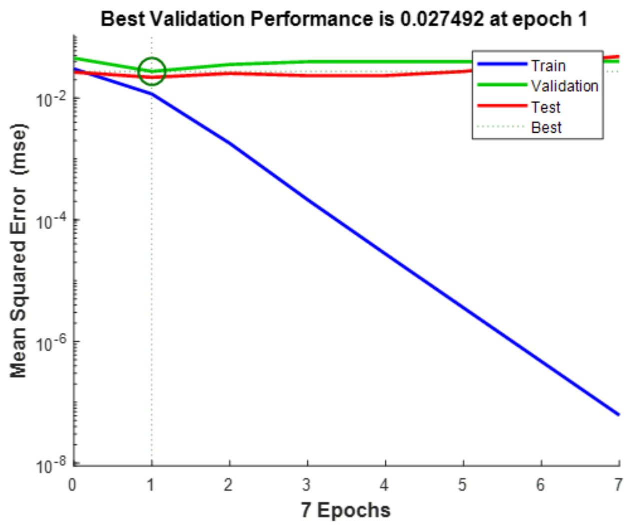

Figure 7. Error generated by the training with 20 neurons

The training with 20 neurons presents better results than the case with 10, giving a very high margin of confidence and showing results of the simulation with a confidence level of around 95%. For training and validation also had better results, especially in the validation that went from 75% for training with 10 neurons to 94% with 20 neurons (Figure 8).

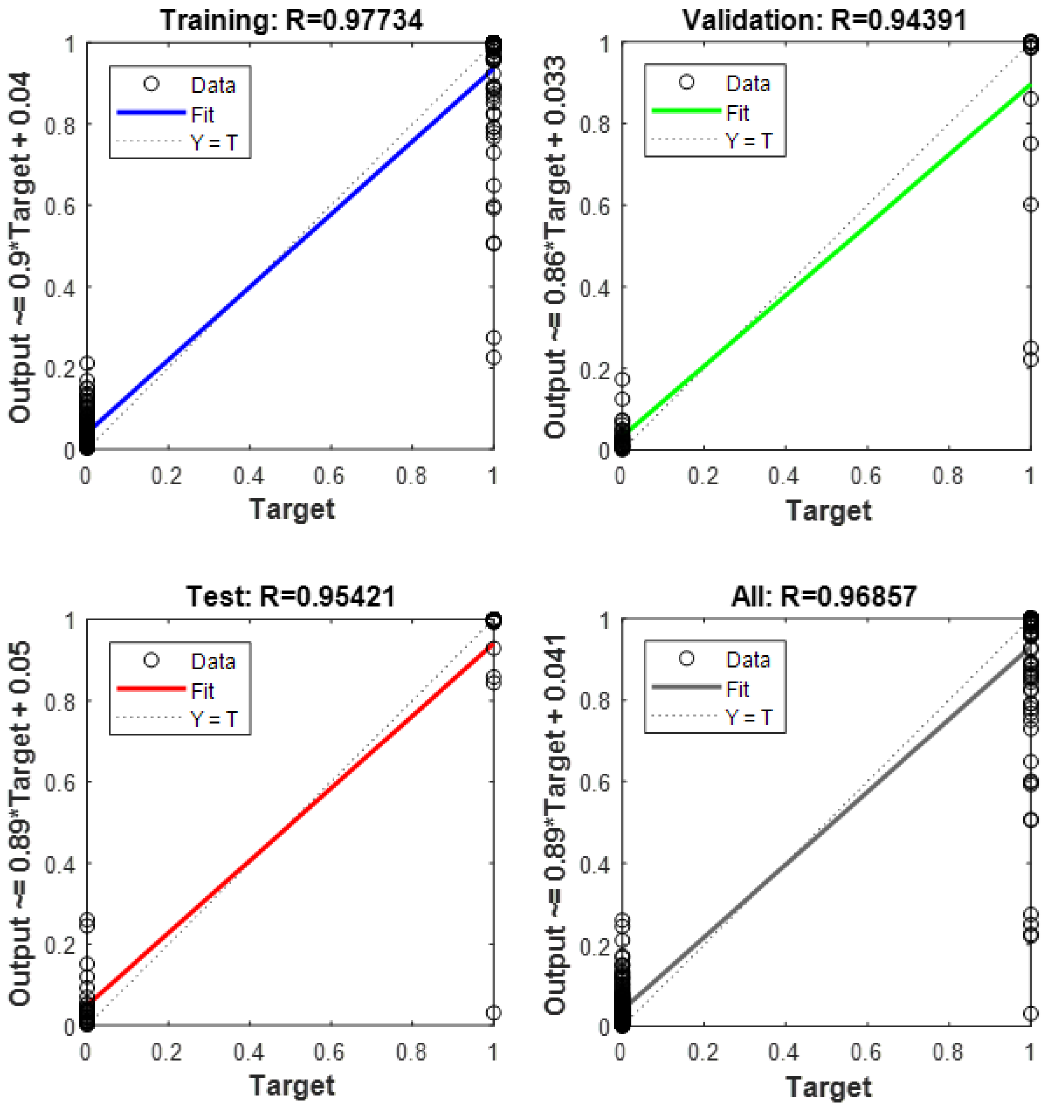

Figure 8. Results of the training with 20 neurons

CONCLUSION

Neural networks are an efficient tool to perform data mining with low computational costs, sin this case to make any type of analysis, the most important process and normally the one who take more time is the preprocessing and the adjustment of the data, for this there are various tasks and methods. In the case of the work done with the data measured by the meteorological radar, it was necessary to carry out an initial decoding process, then transform them into an appropriate format, clean them by eliminating erroneous information and normalize them to the scale used according to the standard.

REFERENCES

[1] R.E. Rinehart. El radar para los Meteorólogos. Traducido por O. L. Rodríguez. Rinehart Publications. 2007.

[2] Radar Meteorology. July 1997. [On line]. Available: WW2010 University of Illinois: http://ww2010.atmos.uiuc.edu/(Gh)/guides/rs/rad/home.rxml

[3] About Radar: Enviroment Canadá, August 2014. [On line]. Available: http://www.ec.gc.ca/meteo-weather/default.asp?lang=En&n=2B931828-1D.

[4] A. Ryzhkov, T. Schuur, D. Burgess, P. Heinselman, S. Giangrande y D. Zrnic, "The joint polarization expert. Polarimetric Rainfall Measurements and Hydrometeor Classification" *American Meteorological Society*. pp809-824, June 2005

[5] R. J. Doviak and D. S. Zrnic. Doppler Radar and Weather Observations, Second Edition. San Diego, Academic Press, Inc., 1993.

[6] Vaisala. 2018. Programming Guide. IRIS Radar. [On line]. Available: ftp://ftp.sigmet.com/outgoing/manuals/IRIS_Programmers_Manual.pdf

[7] Earl Observing Laboratory. 2016. [On line]. Available: https://www.eol.ucar.edu/content/running-radx-applications.

[8] Mathworks. 2018. [On line]. Available: https://la.mathworks.com/help/matlab/ref/netcdf.open.html

[9] Mathworks. 2018. [On line]. Available: https://la.mathworks.com/help/matlab/ref/netcdf.inq.html

[10] Mathworks. 2018. [On line]. Available: https://la.mathworks.com/help/matlab/ref/netcdf.inqvar.html

[11] Mathworks. 2018. [On line]. Available: https://la.mathworks.com/help/matlab/ref/netcdf.getvar.html