Daily river level forecast based on the development of an artificial neural network: case study in La Virginia - Risaralda

UNIVERSIDAD DE ANTIOQUIA 1803

Pronóstico del nivel diario de un río basado en el desarrollo de una red neuronal artificial: estudio de caso en La Virginia –Risaralda

Tito Morales–Pinzón[1,2*], Juan David Céspedes–Restrepo[1], Manuel Tiberio Flórez–Calderón[1]

[1]Grupo de investigación Gestión Ambiental Territorial, Facultad de Ciencias Ambientales, Universidad Tecnológica de Pereira. Carrera 27 # 10-02 Barrio Alamos. C. P. 660003. Pereira, Colombia.
[2]Observatorio Ambiental Urbano-Regional de Risaralda, Universidad Tecnológica de Pereira. Carrera 27 # 10-02 Barrio Alamos. C. P. 660003. Pereira, Colombia.

ABSTRACT: The municipality of La Virginia (Risaralda, Colombia) is constantly affected by floods that originate from increased water levels in the Cauca River. Disaster relief agencies do not currently have adequate monitoring systems to identify potential overflow events in time-series observations to prevent flood damage to homes or injury to the general population. In this paper, various simulation models are proposed for the prediction of flooding that contributes as a technical tool to the development and implementation of early warning systems to improve the responsiveness of disaster relief agencies. The models, which are based on artificial neural networks, take hydroclimatological information from different stations along the Cauca River Basin, and the trend indicates the average daily level of the river within the next 48 hours. This methodology can be easily applied to other urban areas exposed to flood risks in developing countries.

RESUMEN: El municipio de La Virginia (Risaralda, Colombia) está constantemente afectado por inundaciones que se originan debido al aumento de los niveles de agua en el río Cauca. Los organismos de socorro y atención de desastres actualmente no tienen sistemas adecuados de vigilancia para identificar eventos potenciales de desbordamiento en el tiempo, y así evitar daños por inundaciones a las viviendas o daños a la población en general. En este trabajo, se proponen diversos modelos de simulación para la predicción de inundaciones que contribuye como una herramienta técnica para el desarrollo e implementación de sistemas de alerta temprana para mejorar la capacidad de respuesta de los organismos de socorro. Los modelos, están basados en redes neuronales artificiales y toman la información hidroclimatológica de diferentes estaciones existentes a lo largo de la cuenca del río Cauca, donde la tendencia indica el nivel medio diario del río durante las siguientes 48 horas. Esta metodología se puede aplicar fácilmente a otras zonas urbanas expuestas a los riesgos de inundación en países en desarrollo.

1. Introduction

La Virginia is a municipality of the Risaralda department in the Andean Region in Colombia. It is located at latitude 4°54'50'' North and longitude 75°54'43'' West [1], in the valley of the Cauca and Risaralda Rivers. This area has important economic agricultural activity, characterised by an emphasis on coffee [2], as part of the economic dynamics of the municipality, which revolves around agriculture and livestock, extraction of drag materials, fishing and tourism [3]. According to projections made by the National Administrative Department of Statistics [4], the municipality has approximately 31,657 inhabitants, located primarily in the urban area (98.3%).

While passing through downtown La Virginia, the Cauca and Risaralda Rivers often produce floods from their convergence. A disaster of this magnitude in this urban area affects approximately 1190 people and 58 households on average per event [This value was calculated based on the 24 events that occurred during the period 1980-2010, according to information compiled by [5]]. In some extreme cases, the floods have caused approximately 20,000 casualties [6]. This condition of coexistence with the disaster, represented by the frequent floods that occur, has

* Corresponding author: Tito Morales Pinzón.
E-mail: tito@utp.edu.co

been present throughout the history of the town. The first reports of flooding date from 1922, just two decades after the first urban development [5].

Flood events in the municipality can occur either suddenly or slowly, but in both cases, the configuration of the phenomena depends on natural and human factors [6]. Natural factors are the distribution of rainfall, morphometry and watershed physiography as well as the Risaralda River impoundment following the increase in the level of the Cauca River. However, despite the importance of these elements, the main factor in defining the threat is of anthropogenic nature and involves the occupation of the flood plains of the Cauca and Risaralda Rivers. The occupancy model that has led to this situation was set during the history of the town and is the result of the weaknesses in the land use planning process, as mentioned in the study regarding housing protection danger areas in the county of La Virginia; "the Municipality of La Virginia is a complex phenomenon, where the lack of coherent planning of the past has allowed location of housing in areas of threat" (quoted by [7]).

Because of the frequency and severity of flooding in the city, strategies for risk management must be incorporated into city planning. This requires the development of technical tools at an early stage, allowing the implementation of early-warning systems (EWS) for flood threats. There are currently tools for prediction and monitoring, as well as different types of mathematical models, such as those based on the Navier-Stokes equations for laminar flows, the Kalman filter, genetic algorithms, and ARIMAs (autoregressive integrated moving average time-series analysis) [8] which usually require great volumes of information of the studied context (hydrological, hydraulic and geomorphological variables).

However, as opposed to the weaknesses of these mathematical models in flood forecasting, such as those discussed by [9], thanks to the massive parallelism of its structure, the Artificial Neural Networks models [onwards ANN] have many advantages including adaptive learning, self-organization, fault tolerance, easy insertion into existing technology, and the possibility of operation in real-time in problems related with filtering applications, control and identification of patterns, among others. This is largely due to their capacity of process information in parallel, allowing the development of calculations at very high speeds and exceeding the real-time constraints of most applications that try to serve to these purposes [10].

In Latin America, a flood warning system has been developed for the Biobío River in Chile, which was constructed from historical data collected from 29 floods [11]. A study conducted by the University of Asunción in Paraguay developed six deterministic and stochastic mathematical models using different methodologies, including ANN, to study the flood phenomenon in the Paraguay River, and they found that the most accurate methodology for describing the river dynamics was ANN [12].

Current studies aim to similar conclusions: in Sudan, an ANN model was developed to predict flooding levels in the River Nile. In this study, ANN methodology demonstrated various advantages in relation to other models (especially in the number of variables required to give accurate predictions) [13]. Also, in Taiwan, five models were constructed to forecast the water level of the Tanshui River; according to the results, among all the learning models considered, the artificial neural networks (ANN) yielded more favourable results than other forecast models of water level [14]. There are also many studies of this type in which ANN models demonstrate that they are superior to physical-based hydrodynamic models because they can preserve nonlinear characteristics between input and output variables [15-17].

Other investigations, such as "Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting" in Harvey River, Western Australia and "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control" demonstrate the benefits that learning models based in artificial neural networks have in predicting the dynamics of rivers [18, 19]. The same studies proved also that performance can be improved by creating hybrid models based on ANN models.

These experiences show that ANN models are reliable for predicting floods and identifying threats of this type [12]. This feature is often attributed to the model's ability to combine historical information with algorithms that describe the behaviour of the aspects that determine such phenomena. However, ANN models have also been used in different applications to predict the weather, the temperature, the price of electricity, and landslides, among others [20], thus demonstrating their ability to describe complex problems.

The scope of this research is to do a comparison between different flood forecasting models based on different ANN structures and hydrometeorological information, with the objective of estimating the model performance under different configurations related to limited resources of small municipalities.

2. Methodology
2.1. The Cauca River Basin

The Cauca River Basin is one of the most important basins in Colombia because it contains the sugar industry, part of the coffee industry, the Antioquia mining and agricultural development, and a significant portion of the manufacturing industry in western Colombia.

The primary stream bed of the basin, the Cauca River, originates in the Colombian Massif near Páramo Sotará and crosses the country from south to north, crossing the departments of Cauca, Valle del Cauca and Risaralda, where it meets the urban area of La Virginia [1]. The Cauca River Basin has a total area of 63,300 km^2, equivalent to approximately five per cent of the territory of the country [21] (see Figure 1).
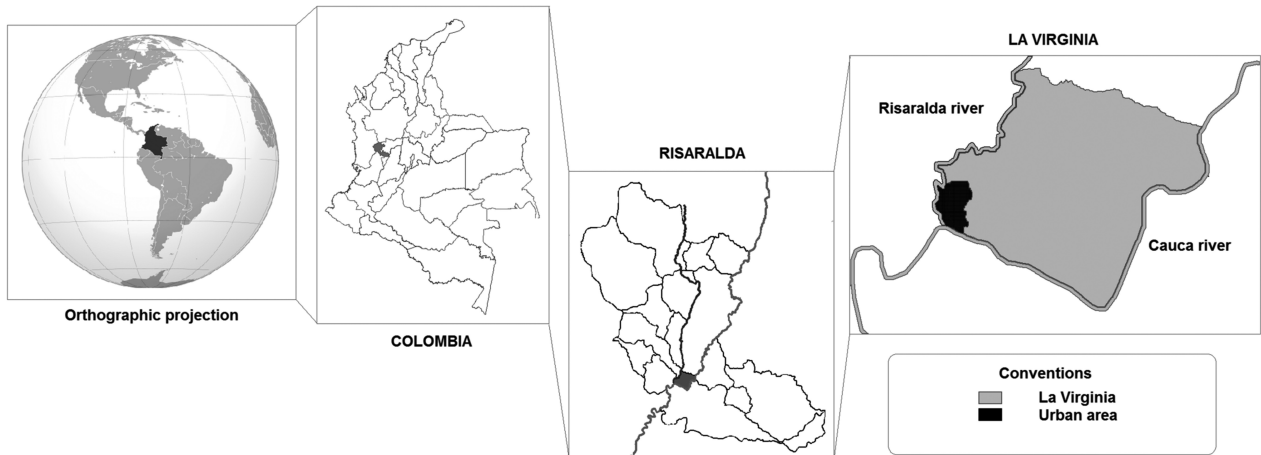
Figure 1 Location of the urban area of La Virginia

From its origin until La Virginia town, the Cauca river crosses Colombia collecting water from other rivers in more than 32 departments. Risaralda and La Vieja rivers are some of the most important tributary sources, with an average flow of 27.2 m³/s and 95.0 m³/s, respectively [22]. La Vieja, the largest tributary river in the middle of the basin, crosses the departments of Valle, Quindío and Risaralda [22].

2.2. Artificial Neural Networks for prediction of average daily levels in the Cauca River

In this study, the typology of ANN chosen to be constructed was the multi-layer perceptron (MLP). This type of ANN is the most widely used in nonlinear problems of time series prediction, for which it has demonstrated high reliability (see Figure 2). As suggested by [10, 15, 16], the MLP is composed of at least three layers, or levels, of neurons. The first level is the input layer; it is responsible for capturing external stimuli or information that feeds the model. The second level, the hidden layer, receives information from the input layer to process. The third layer, the output layer, receives the processed information and delivers the output variables to the network.

Figure 2 Scheme of an Artificial Neuron. Adapted from [23]

In this model, the input variables are the daily average level and total daily precipitation for the Cauca and Risaralda Rivers, which pass through La Virginia municipality. These variables were defined based on the availability and quality of hydrometeorological information in the section studied of the Cauca river basin, and it was assumed that their behaviour represents fundamental aspects related with the floods in La Virginia.

However, in order to establish some indication about the future variables performance, the selected points to collect information were defined in upstream sections of the Cauca and Risaralda basins before the meeting point of the two rivers in La Virginia. This was done under the assumption that precipitation upstream of the basin (at time t-1) defines the behaviour and the level of the river in the future (time t). To select the optimal points to collect data, a hydrology concept was used, i.e., the concentration time, which is the time it takes the water from a basin to make the journey upstream to its mouth [24]. At this point, it is necessary to explain that the information of La Vieja basin was not used within the defined variables. The assumptions were that given the proximity of its river mouth to the municipality of La Virginia (about 15 km along the drainage) and the hydraulic concentration time of the Cauca river basin, the contribution of La Vieja river in terms of flow would be contained (for the analyzed unit time -one day-) in the measurement of the average level of the Cauca River.
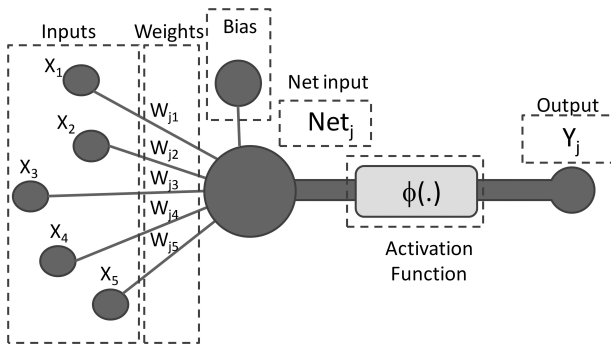
The concentration time was calculated with the equation defined by [25], which also includes variables such as the channel length and the mean slope. The area of the basin is included in the calculation, replacing variables such as runoff and the volume of catchment rainwater.

Next, the hydroclimatological stations were defined due their proximity to the points associated with the concentration times along the watercourse. The average daily water level and total daily precipitation variables were used as the starting point to define the input variables in the model.

The output variables were defined based on their capacity to contribute to the municipal contingency plans with respect to disasters prevention caused by flooding. Thereby, the average daily river level for the Cauca River station at La Virginia point, using time periods of t, t+1, and t+2 (the current day, a day later and two days later, respectively) was considered the best option for the model's output variable (see Table 1).

Table 1 Selected variables for construction of the ANN

Variable	Description	Time	Station from which data were obtained	Unit of measurement
		Input variables		
X1	Average Daily Level, Risaralda River	t-1	La Virgen, water level gauge	cm
X2		t-2		
X3		t-3		
X4	Total Daily Rainfall, Risaralda River	t-1	Taparcal, Rainfall Station	mm/d
X5		t-2		
X6		t-3		
X7	Total Daily Rainfall, Cauca River	t-1	Mediacanoa, Rainfall Station	
X8		t-2		
X9		t-3		
		Output variables		
Y1	Average Daily Level, Cauca River	t	La Virginia, water level gauge	cm
Y2		t+1		
Y3		t+2		

The selected variables were compared using the Pearson correlation coefficient. This criteria was used for deciding the variables that were included in the networks modelled. When implementing an Artificial Neural Network Model, the optimum number of neurons and hidden layers in the network architecture cannot be determined analytically; they are generally found experimentally by trial and error [26]. Thus, various networks were constructed with different configurations and then the number of neurons was gradually increased and decreased to obtain a model with the proper adjustment. The predictive capacity in the different networks was compared by calculating the mean square error.

As recommended by [27], the initial number of neurons in the hidden layer was estimated using the Kolmogorov theorem (For any continuous function, it is possible to create an artificial neural network of forward propagation of three layers that appropriately reproduces the behaviour of the function. This network must have n processing elements in the input layer, m in the output layer and 2n +1 in the hidden layer).

Additionally, in accordance with [10], a sigmoidal function was applied to the neurons of the hidden layer and a linear function to the neurons of the output layer. To avoid noise and to shield the input signal from the model, the original data were normalised.

The available information consisted of 1132 to 1455 patterns, depending on the processes used for training, validation, testing and verification. 23 different networks were tested. Following suggestions from [28], each network was trained using Microsoft Excel spreadsheets and the Premium Solver macro version 12.5, which was designed for Excel 2007 by Frontline Systems. The generalised reduced gradient algorithm was used to train the networks.

Once completed for all test networks, this process helped to estimate the mean square error for the network outputs (Y1, Y2 and Y3) as a performance measure. To check the performance of the selected network, scatter plots and a correlation analysis (r) were used between the predicted and actual values.

2.3. Neural network training times

The training was conducted in two stages, each of which consisted of 500 iterations. This method was employed to continuously confront training error and validation for each model and to avoid over-training the networks.
Office 2007 Excel was used for the training process, with the addition of the Premium Solver software V 12.5, developed by Frontline Systems Company, Inc. This complement allowed the use of a larger number of variables than the standard solver of Excel (this was a fundamental requirement because it was required for training the complex networks).

The training times may vary according to the computing ability (defined by the characteristics of the computer used to perform the calculations). For the training of the networks herein, a Dell Inspiron 580S computer with an Intel ® Core™ i3 processor and six gigabytes of RAM was used. The operating system was the 64-bit version of Windows 7 Home Premium

3. Results and discussion

From the statistical analysis of data, there are significant correlations of input variables related to La Virginia station with the output variables of the network. Significant positive correlations were found for the annual precipitation and the

Table 2 Pearson correlation coefficient for the variables defined

Meteorological station		La Virginia station*		
		t	t+1	t+2
		Rainfall (mm)		
Mediacanoa	t-1	0.17	0.19	0.19
	t-2	0.19	0.19	0.21
	t-3	0.19	0.21	0.21
Taparcal	t-1	0.26	0.27	0.28
	t-2	0.27	0.28	0.29
	t-3	0.28	0.29	0.29
		Water level gauge (cm)		
La Virgen	t-1	0.59	0.58	0.58
	t-2	0.58	0.58	0.57
	t-3	0.58	0.57	0.55
La Virginia	t-1	0.99	0.97	0.95
	t-2	0.97	0.95	0.94
	t-3	0.95	0.94	0.92

* Water level gauge (cm). All Pearson coefficients were highly significant (p<0.01). In bolt are shown the higher Pearson correlations.

water level gauge of different meteorological stations (see Table 2).

Twenty three ANNs were constructed and classified in five groups. These groups were defined based on the input variables used for each one of them. Also, they were classified according to two criteria: the complexity of the networks (implying higher computational requirements) and information constraints.

Then, there are different groups according to criteria of construction and operation. Thus, Group I represents a low complexity of the structure and required information, Group II is medium in both criteria, Group III is medium in complexity of the structure and high in required information, and Group IV and V are high in both criteria. For Group I is much easier to train a simple network because it only requires information of daily average level of Cauca River in La Virginia station (this is located in the same urban area of the municipality).

Similarly, Group II represents models that can be implemented with an average investment of information and computational capability. These models have a simple architecture and they also work with variables related to the total daily precipitation (a much easier variable to measure than variables like the average daily level or flow).

The results showed that the ANN that with the lowest mean square error for the training stage was the NET-IV:05 network corresponding with one of the networks in the group of ANN of major resources required. However, the NET-IV:01 network (one of the simplest networks) showed more skill than NET-IV:05 when processing new information. Thus, we present the NET-IV:01 network as the most accurate because it has better performance than the other networks in the validation and verification process.

Using the mean square error as criteria, models do not behave as expected, because a significant investment in complexity and the high requirement of information does not necessarily generate a proportional effect on the accuracy of predictions. For example, networks belonging to Group V have a higher mean square error when they are compared with simpler networks. This would indicate that small municipalities do not require high investments to develop models with an adequate level of reliability.

Based on the estimates, we verified that the model network NET-IV:01 appropriately fits the behaviour of the output variable Y1 (the level of the Cauca River at time t on the current day). The scatter plot in Figure 3 shows a high level of adjustment and a determination coefficient (r^2) of 0.981. This means that the model has an acceptable error rate when predicting the average daily level of the river using data from previous days.

The constructed network can be used to predict values of the variables Y2 and Y3 (Cauca River level at times t+1 and t+2). The determination coefficients were 0.958 and 0.924 for Y2 and Y3, respectively. The dispersion between the predicted and real data is greater as the forecast horizon increases.

The error analysis shows that the forecast values for Y1, Y2 and Y3 are statistically independent and that they have a constant variance (see Figure 4).

Based on the scatter plot of the error, the model is correctly adjusted to forecast for one day (time t), with an error in many cases of less than 3% (normalised-value, for 75% of the cases), and most are not greater than 5% (90% of the cases). The variable Y2 has a similar adjustment level, i.e., a strong trend toward the line that indicates the error is zero (0), and reason to suggest that for 75% of the cases, the error will be less than 4 % and generally (90% of the data), the error less than 6%.
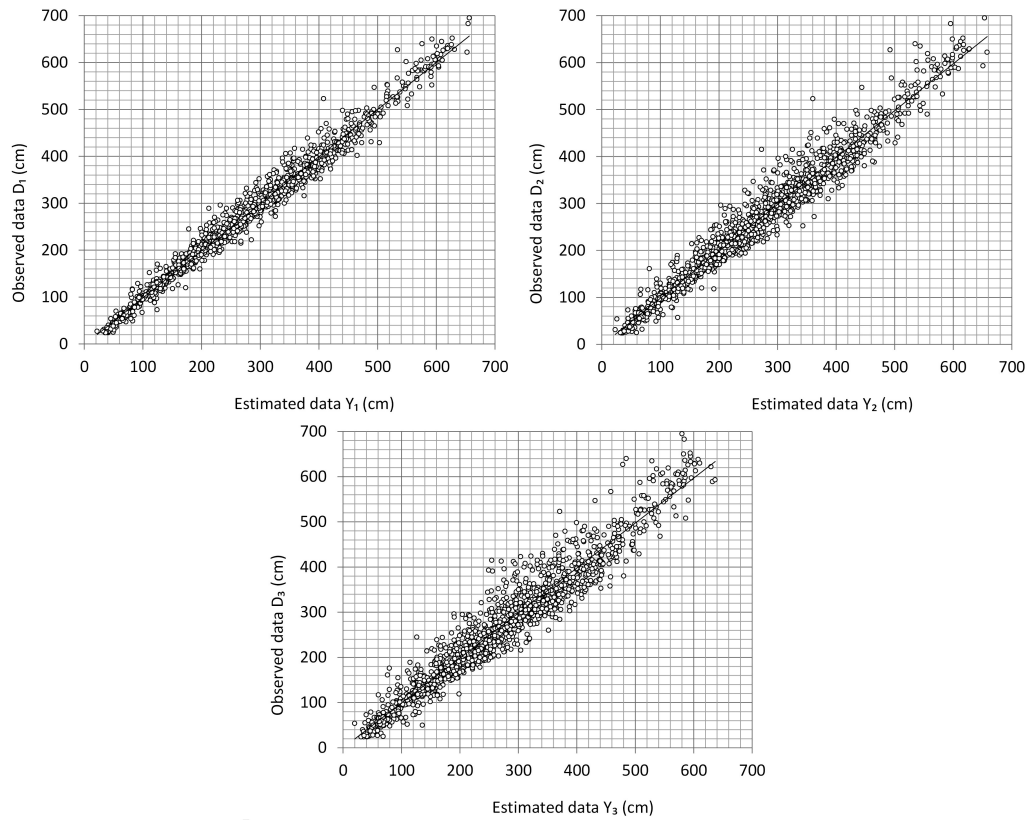
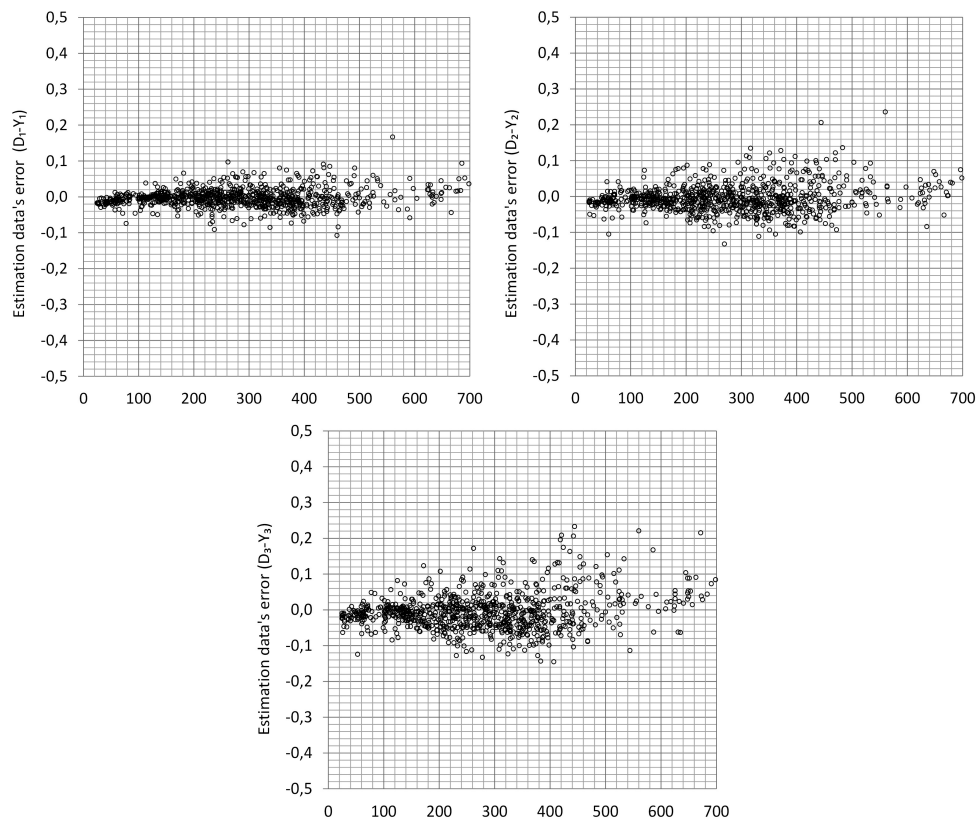**Figure 3** Observed data vs. Estimated data for each ANN output

Figure 4 Error dispersal for output variables

A similar behaviour is observed for the variable Y3; the error dispersion plot shows a behaviour where, although the values are grouped around the straight line, there is a greater range of errors than for the first two variables, Y1 and Y2 (see Figure 4).

The validation and verification simulation sets had a standard deviation of 18.04 cm error in the model estimates (Error = Real Data - Estimate) for variable Y1. A statistical analysis based on the calculation of quartiles and percentiles suggests that, based on the error distribution obtained, for 50% of the estimates made with the ANN, the error will be less than 9.47 cm, seventy-five per cent of the estimates will have an error of less than 17.33 cm, and ninety per cent of the estimates will have an error of less than 31.19 cm.

Comparing the output variable Y1 (Cauca River level at time t) with the output variable Y2 (Cauca River level in period t +1), the change in the degree of fit of the model is low, presenting a variation in the mean error of 6.63 cm. However, the variation in the error distribution for variable Y2 is larger than that of Y1, presenting a difference of 12.40 cm at the 90th percentile with respect to the output variable Y1. According to the error distribution of the simulations of the output variable Y2, 50% of the estimates made within two days are accurate within 14.53 cm (the median value represents the distribution divided in half according to the frequency of the observed data), seventy-five per cent within 27.23 cm, and in ninety per cent of the estimates, the error will be less than 43.59 cm.

In contrast, the behaviour of the variable Y3 behaves differently than output variables Y1 and Y2 in regard to the distribution of the data. For Y3, the average error is 13.17 cm greater than that found for Y1. According to the distribution of the simulated data, fifty per cent of the estimates of Y3 will have errors less than 19.70 cm, seventy-five per cent will have errors less than 36.05 cm, and ninety per cent will have error values of less than 57.86 cm.

In addition to resulting in a higher error than the other output variables, the Y3 variable has a behaviour that matches the behaviour of the problem addressed. Thus, Y3 can be used to track the general trend of the Cauca River level in specific conditions.

3.1. Neural network training time

ANN models with a less training time and a simpler structure represent a lower investment of development and operation, this is a useful characteristic for small municipalities. The increase of neurons in the hidden layer of the model increases the complexity, which is reflected in the need to invest more time in the training process of the network. Thus, the net that required less training time was network NET I:04, having only 5 neurons in its hidden layer, as opposed to network NET V:02, with 25 neurons in its hidden layer and whose total training time is nearly three hours.

Additionally, there is usually a higher requirement on the computation time in those networks that have three output variables, as observed in networks NET I:01/NET I:05, NET III:01/NET III:04, NET III:02/NET III:05, NET III:03/NET III:06, and NET IV:01/NET IV:05, which share the same architecture in terms of the number of neurons in the input and hidden layers but that differ in the number of neurons in the output layer.

With respect to the current problem, an ANN with a large number of variables and that requires large training or retraining times, such as networks having a number equal to or greater than 19 neurons in the hidden layer (see Figure 5), are highly inefficient if the level of adjustment that they require can be reached or exceeded by less complex networks. However, an increase in complexity will not occur in all cases from a large number of neurons in the hidden layer and in the output; therefore, the long computation time for training necessarily implies a low level of the ANN's efficiency. This is evidenced in network group NET II, which contains the ANNs with high mean square error values and low training times (see Figure 5). These models have a low demand on computational time and a low level adjustment, a situation that makes them less suitable in relation to the current problem in comparison with the other ANNs evaluated.

Results obtained from the measurement of training times vs. the number of neurons in the hidden layer and the level of adjustment suggest that, for this problem, the NET IV:01
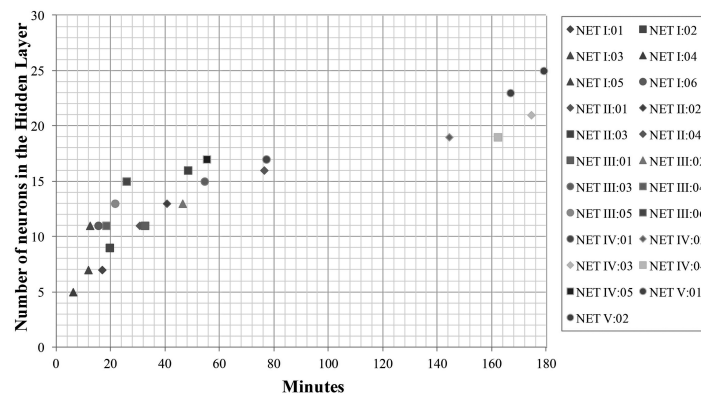

Figure 5 **Number of neurons in the hidden layer vs. Training time**

and NET IV:05 networks presented high levels of efficiency because they showed a good level of adjustment, are less vulnerable to failures by partial damage to the network and have training times that are close to the mean of the other networks (1 h 02 ´28´´).

4. Conclusions

According to the methods explored in this study, it is possible to implement an artificial neural network multilayer perceptron model to predict the level in the Cauca River, with appropriate model adjustments and thereby estimate the flood's risk in La Virginia. This approach could eventually allow the development of early warning systems for disaster prevention and response in cities that are located beside the river valley in Colombia.

Because models constructed using the ANN methodology do not require a priori conceptualisation of the interrelationships that develop within the studied system, they can be built and deployed quickly, efficiently and with very low development and operation costs. In the case study described, the proposed Artificial Neural Network only requires Microsoft Office Excel as platform for development/operation and information of only two meteorological stations.

The ability to adapt and learn from ANN models through historical data allows the models continuously adjust to changes that arise in the immediate context of the studied system. This feature allows flooding problems to be addressed even in the face of high variability phenomena, such as climate change.

For small municipalities having limited resources or information, it is possible to build an ANN with an appropriate level of adjustment and having a good performance of prediction of risk flood (models with low mean square error values). These models would help to small municipalities to management processes focused on early warnings.

Appendix

Appendix A-1

The concentration time given by Bransby-Williams is [25] (Eq. 1):

$$T_c = 14.6 * L * A^{-0.1} * S^{-0.2}$$

(1)

where:

T_c is the concentration time for the basin (hours),
L is the length of the longest watercourse (km),
A is the watershed area (km^2), and
S is the average slope (m/m).

Using Eq. (1), the length of the longest watercourse for 24 and 48 hours for each of the streams involved was estimated.

Appendix A-2

The predictive capacity is calculated using a modified estimator for the mean square error (Eq. 2):

$$MSE = \frac{1}{3n} \sum_{Y=1}^{3} \sum_{k=1}^{n} \left[(Xp_i)_{Yk} - (Xp_o)_{Yk} \right]^2$$

(2)

where:

Xp_i is the normalised output,
k is the sample size, and
Y is the number of prognostic variables.

Appendix B

In order to correct the different scales used in the ANN, the original input data were normalised using next expression (3):

$$Xp_i = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

(3)

where:

Xp_i is the normalised input,
X_i is the independent input data,
X_{min} is the minimum input data, and
X_{max} is maximum input data.

Appendix C-1

Summary of characteristics represented by the ANN modelled

ANN Group	Input Variables	Arquitectures	Output Variables
I.	Variable: Average Daily Level, Cauca River. Times t-1, t-2, t-3 Station: La Virginia, water level gauge	3-7-3; 3-9-3;3-11-3; 3-5-2; 3-7-2; 3-11-2	Variable: Average Daily Level, Cauca River. Times t, t+1, t+2 Station: La Virginia, water level gauge

II.	**Variable:** Total Daily Rainfall, Cauca River. Times t-1, t-2, t-3 **Station:** Mediacanoa Station **Variable:** Total Daily Rainfall, Risaralda River. Times t-1, t-2, t-3 **Station:** Taparcal Station	6-11-3; 6-13-3; 6-16-3	**Variable:** Average Daily Level, Cauca River. Times t, t+1, t+2 **Station:** La Virginia, water level gauge
III.	**Variable:** Average Daily Level, Cauca River. Times t-1, t-2, t-3 **Station:** La Virginia, water level gauge **Variable:** Average Daily Level, Risaralda River. Times t-1, t-2, t-3 **Station:** La Virgen––, water level gauge	6-11-3; 6-13-3; 6-15-3	**Variable:** Average Daily Level, Cauca River. Times t, t+1, t+2 **Station:** La Virginia, water level gauge
IV.	**Variable:** Average Daily Level, Cauca River. Times t-1, t-2, t-3 **Station:** La Virginia, water level gauge **Variable:** Total Daily Rainfall, Cauca River. Times t-1, t-2, t-3 **Station:** Mediacanoa Station **Variable:** Total Daily Rainfall, Risaralda River. Times t-1, t-2, t-3 **Station:** Taparcal Station	9-17-3; 9-19-3; 9-21-3; 9-17-2	**Variable:** Average Daily Level, Cauca River. Times t, t+1, t+2 **Station:** La Virginia, water level gauge
V.	**Variable:** Average Daily Level, Cauca River. Times t-1, t-2, t-3 **Station:** La Virginia, water level gauge **Variable:** Average Daily Level, Risaralda River. Times t-1, t-2, t-3 **Station:** La Virgen, water level gauge **Variable:** Total Daily Rainfall, Cauca River. Times t-1, t-2, t-3 **Station:** Mediacanoa Station **Variable:** Total Daily Rainfall, Risaralda River. Times t-1, t-2, t-3 **Station:** Taparcal Station	12-23-3; 12-25-3	**Variable:** Average Daily Level, Cauca River. Times t, t+1, t+2 **Station:** La Virginia, water level gauge

For each model of ANN, the Premium Solver was configured with the following parameters during the training: Iterations: 500 Accuracy: 1e-005	Convergence: 0.0001 Estimate: Quadratic Derived: Progressive Search: Conjugate Gradient

Appendix C-2

Mean square error for each ANN

ANN Code	Arq	Training Y1	Training Y2	Training Y3	Validation Y1	Validation Y2	Validation Y3	Test Y1	Test Y2	Test Y3
NET I:01	3-7-3	0.00079	0.00165	0.00253	0.00076	0.00166	0.00275	0.00090	0.00204	0.00295
NET I:02	3-9-3	0.00084	0.00171	0.00259	0.00078	0.00173	0.00279	0.00097	0.00210	0.00300
NET I:03	3-11-3	0.00082	0.00166	0.00255	0.00081	0.00169	0.00275	0.00093	0.00204	0.00294
NET I:04	3-5-2	0.00079	0.00165	---	0.00075	0.00163	---	0.00090	0.00202	---
NET I:05	3-7-2	0.00080	0.00166	---	0.00076	0.00163	---	0.00090	0.00203	---
NET I:06	3-11-2	0.00079	0.00165	---	0.00076	0.00163	---	0.00090	0.00203	---
NET II:01	6-11-3	0.03062	0.03033	0.03069	0.03742	0.03628	0.03663	0.03229	0.03189	0.03180
NET II:02	6-13-3	0.02952	0.02955	0.02948	0.03637	0.03569	0.03564	0.03353	0.03290	0.03206
NET II:03	6-16-3	0.03229	0.03134	0.03103	0.03398	0.03471	0.03393	0.03234	0.03135	0.03181
NET II:04	6-16-3*	0.03114	0.03147	0.03192	0.03664	0.03682	0.03673	0.02990	0.02994	0.02954
NET III:01	6-11-3	0.00095	0.00198	0.00290	0.00091	0.00226	0.00334	0.00113	0.00199	0.00339
NET III:02	6-13-3	0.00091	0.00191	0.00282	0.00088	0.00223	0.00328	0.00110	0.00199	0.00336
NET III:03	6-15-3	0.00093	0.00192	0.00284	0.00093	0.00229	0.00336	0.00109	0.00197	0.00333
NET III:04	6-11-2	0.00100	0.00201	---	0.00098	0.00231	---	0.00114	0.00199	---
NET III:05	6-13-2	0.00090	0.00189	---	0.00091	0.00229	---	0.00111	0.00201	---
NET III:06	6-15-2	0.00092	0.00191	---	0.00091	0.00227	---	0.00110	0.00201	---
NET IV:01	9-17-3	0.00075	0.00177	0.00274	0.00067	0.00150	0.00249	0.00071	0.00155	0.00287
NET IV:02	9-19-3	0.00124	0.00246	0.00329	0.00120	0.00208	0.00286	0.00125	0.00216	0.00342
NET IV:03	9-21-3	0.00117	0.00216	0.00303	0.00120	0.00192	0.00275	0.00138	0.00197	0.00307
NET IV:04	9-19-3*	0.00104	0.00215	0.00326	0.00102	0.00180	0.00290	0.00089	0.00234	0.00364
NET IV:05	9-17-2	0.00066	0.00158	---	0.00070	0.00154	---	0.00070	0.00168	---
NET V:01	12-23-3	0.00083	0.00201	0.00288	0.00101	0.00213	0.00321	0.00145	0.00267	0.00359
NET V:02	12-25-3	0.00074	0.00129	0.00205	0.00096	0.00195	0.00281	0.00105	0.00224	0.00363

* The input variables for these ANN have been modified as follows: data for inputs from Mediacanoa and Taparcal stations was changed according to a correlation analysis. In this way, the employees' times were t-3, t-4, and t-5 instead of t-1, t-2 and t-3.

Appendix C-3

Training times for different ANNs

ANN Code	Architecture	Training phase time	
		1	2
NET I:01	3-7-3	1'58''	14'54''
NET I:02	3-9-3	1'46''	17'52''
NET I:03	3-11-3	1'58''	10'27''
NET I:03	3-11-3	1'58''	10'27''
NET I:04	3-5-2	1'21''	4'51''
NET I:05	3-7-2	1'35''	10'12''
NET I:06	3-11-2	1'47''	13'41''
NET II:01	6-11-3	2'57''	27'42
NET II:02	6-13-3	3'02''	37'35''
NET II:03	6-16-3	2'59''	45'23''
NET II:04	6-16-3*	3'34''	1 h 12'47''
NET III:01	6-11-3	2'05''	30'29''
NET III:02	6-13-3	1'32''	44'52''
NET III:03	6-15-3	1'53''	52'36''
NET III:04	6-11-2	1'59''	16'25''
NET III:05	6-13-2	1'23''	20'13''
NET III:06	6-15-2	1'53''	23'56''
NET IV:01	9-17-3	4'30''	1 h 12'38''
NET IV:02	9-19-3	5'23''	2 h 18'54''
NET IV:03	9-21-3	6'41''	2 h 47'44''
NET IV:04	9-19-3*	5'52''	2 h 36'15''
NET IV:05	9-17-2	4'05''	51'12''
NET V:01	12-23-3	3'25''	2 h 43'22''
NET V:02	12-25-3	3'41''	2 h 55'19''

* The input variables for these ANN have been modified as follows: data for inputs from Mediacanoa and Taparcal stations was changed according to a correlation analysis that was performed. In this way, the employees' times were t-3, t-4, t-5 instead of t-1, t-2 and t-3

6. References

1. J. Pulecio. *Mitigación de inundaciones ocasionadas por los ríos Risaralda y Cauca en el Municipio de La Virginia, Departamento de Risaralda.* Informe final de Consultoría a la Corporación Autónoma Regional de Risaralda (CARDER). Pereira, Colombia. 2010. pp. 1-24.
2. H. Vásquez, M. Holguín, G. Osorio. *Base Ambiental con Énfasis en Riesgos*. Municipio de La Virginia. Documento técnico de la Corporación Autónoma Regional del Risaralda -CARDER-. Pereira, Colombia. 2002. pp. 1-104.
3. Alcaldía de La Virginia - Risaralda. *Nuestro Municipio: Geografía.* Available on: http://www.lavirginia-risaralda.gov.co/informacion_general.shtml. Accessed: March 22, 2011.
4. Departamento Administrativo Nacional de Estadística (DANE). *Boletín Censo General 2005. Perfil La Virginia, Risaralda*. DANE. Bogotá, Colombia. 2010. pp. 1-6.
5. Dirección de Gestión de Riesgos (DGR), Corporación OSSO. *Disaster Information System – DESINVENTAR.* Available on: http://www.desinventar.net/DesInventar/main.jsp?countrycode=col. Accessed: May 15, 2011.
6. Comité Local para la Prevención y Atención de Desastres del Municipio de La Virginia (CLOPAD). *Plan Local de Emergencia y Contingencia Municipio de La Virginia.* CLOPAD. La Virginia, Colombia. 2009. pp. 1-151.

7. Corporación Autónoma Regional del Risaralda (CARDER). Diagnóstico de Riesgos Ambientales, Municipio de La Virginia. CARDER. Pereira, Colombia. 2002. pp. 1-50.

8. Consorcio Evaluación de Riesgos Naturales (ERN). *Metodología de evaluación probabilista de riesgos naturales*. Central American Probabilistic Risk Assessment (CAPRA). Available on: http://www.ecapra.org/es/metodolog%C3%ADa-de-evaluaci%C3%B3n-probabilista-de-riesgos-naturales. Accessed: February 26, 2011.

9. D. Matich. Redes Neuronales: *Conceptos Básicos y Aplicaciones.* Informe. Universidad Tecnológica Nacional. Rosario, Argentina. 2001. Available on: http://www.scribd.com/doc/48697993/Virus-Hack-Redes-Neuronales-Conceptos-Basicos-y-Aplicaciones. Accessed: March 3, 2011.

10. E. Caicedo, J. López. *Una aproximación práctica a las redes Neuronales Artificiales.* 1st ed. Ed. Programa Editorial Universidad del Valle. Cali, Colombia. 2009. pp. 1-217.

11. R. González, X. Vargas, A. López, L. Belmar. *Sistema de alerta de crecidas en el río Biobío.* Uso recursivo de un modelo de Redes Neuronales. Proceedings of the III Encuentro de Aguas. Santiago, Chile. 2002. Available on: http://www.bvsde.paho.org/bvsacd/encuen/gonza.pdf. Accessed: March 14, 2011.

12. C. Guerreño, O. Ferreira, L. Chamorro, R. Almirón. "Análisis de Modelos Matemáticos Aplicados al Estudio del Río Paraguay". *Revista de Ciencia y Tecnología*. Vol. 1. 2001. pp. 111-118.

13. S. Hag. "Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan". *Alexandria Engineering Journal*. Vol. 53. 2014. pp. 655-662.

14. W. Chih. "Comparing lazy and eager learning models for water level forecasting in river-reservoir basins of inundation regions". *Environmental Modelling & Software*. Vol. 63. 2015. pp. 137-155.

15. B. Unal, M. Mamak, G. Seckin, M. Cobaner. "Comparison of an ANN approach with 1-D and 2-D methods for estimating discharge capacity of straight compound channels". Advances in Engineering Software. Vol. 41. 2010. pp. 120-129.

16. C. Wei, L. Wen, H. Ming. "Comparison of ANN approach with 2D and 3D hydrodynamic models for simulating estuary water stage". *Advances in Engineering Software*. Vol. 45. 2012. pp. 69-79.

17. B. Balouchi, M. Nikoo, J. Adamowski. "Development of expert systems for the prediction of scour depth under live-bed conditions at river confluences: *Application of ANNs and the M5P model tree". Applied Soft Computing Journal*. Vol. 34. 2015. pp. 51-59.

18. H. Badrzadeh, R. Sarukkalige, A. Jayawardena. "Impact of multi-resolution analysis of artificial intelligence models inputs on multi-step ahead river flow forecasting". *Journal of Hydrology.* Vol. 507. 2013. pp. 75-85.

19. F. Chang, P. Chen, Y. Lu, E. Huang, K. Chang. "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control". *Journal of Hydrology.* Vol. 517. 2014. pp. 836-846.

20. G. Ovando, M. Bocco, S. Sayago. "Redes Neuronales para Modelar la Predicción de Heladas". *Revista Agricultura Técnica*. Vol. 65. 2005. pp. 65-73.

21. Universidad del Valle, Corporación Autónoma Regional del Valle del Cauca (CVC). Volumen I: Caracterización del Río Cauca Tramo Salvajina – La Virginia. Available on: http://www.cvc.gov.co/cvc/Mosaic/Publicacionesf1.html#volumen1. Accessed: April 12, 2011.

22. Universidad del Valle, Corporación Autónoma Regional del Valle del Cauca (CVC). *Volumen IV y V: Caracterización de Ríos Tributarios del Rio Cauca Tramo Salvajina – La Virginia*. Available on: http://www.cvc.gov.co/cvc/Mosaic/Publicacionesf1.html#volumen4. Accessed: April 3, 2011.

23. M. Gestal. *Introducción a las Redes de Neuronas Artificiales*. Universidade da Coruña. http://sabia.tic.udc.es/mgestal/cv/RNAtutorial/TutorialRNA.pdf. Accessed: March 20, 2011.

24. F. Sánchez. *Medidas Puntuales de Permeabilidad*. Universidad de Salamanca. Available on: http://hidrologia.usal.es/temas/Slug_tests.pdf. Accessed: March 20, 2011.

25. M. Wanielista, R. Kersten, R. Eaglin. *Hydrology: water quantity and quality control*. 2*nd ed*. Ed. John Wiley & Sons. New York, USA. 1997. pp. 1-592.

26. M. Bruen, J. Yang. "Functional networks in real-time flood forecasting -a novel application". *Advances in Water Resources*. Vol. 28. 2005. pp. 899-909.

27. L. Feng, J. Lu. "The practical research on flood forecasting based on artificial neural networks". *Expert Systems with Applications*. Vol. 37. 2010. pp. 2974-2977.

28. J. García, A. López, J. Romero, A. García, C. Camacho, J. Cantero, et. al. "Hojas de cálculo para la simulación de redes de neuronas artificiales (RNA)". *Questiió*. Vol. 26. 2002. pp. 289-305.