Red neuro-difusa para el relleno de datos faltantes en la estación meteorológica Chapingo

Neuro-fuzzy network for missing data population in the meteorological station of Chapingo

Juan Daniel Peña Durán
Centro Universitario UAEM Texcoco
texsmallville@hotmail.com

Irene Aguilar Juárez
Centro Universitario UAEM Texcoco
ireneico@gmail.com

Joel Ayala de la Vega
Centro Universitario UAEM Texcoco
joelayala2001@yahoo.com.mx

Resumen

Esta investigación presenta la aplicación de un modelo de red neurodifusa llamado ANFIS para el problema de estimación de datos faltantes meteorológicos: temperatura, velocidad del viento, humedad relativa y radiación solar en la estimación de la Evapotranspiración de referencia ETo. ANFIS es un método que permite crear la base de reglas de un sistema difuso, utilizando el algoritmo de retro propagación a partir de los datos de un proceso. La estructura de la red neuro-difusa para cada variable meteorológica consiste en dos entradas y una salida. La evaluación del relleno de datos faltantes se realiza mediante la Raíz Cuadrada del Error Cuadrático Medio (RMSE). Los resultados muestran que al usar un mayor número de iteraciones y variación de datos en el entrenamiento puede ayudar a la ANFIS a obtener resultados más precisos.

Palabras clave: ANFIS, datos faltantes, evapotranspiración de referencia.

Abstract

This research presents the implementation of a Neuro-fuzzy network model called ANFIS for the problem of estimation of missing weather data: temperature, speed of the wind, relative humidity, and solar radiation in the estimation of the Evapotranspiration of Reference ETo. ANFIS is a method that allows you to create the basis of rules of a fuzzy system, using the algorithm of retro propagation from a process data. The structure of the neuro-fuzzy network for each meteorological variable consists of two inputs and one output. The evaluation of the filling of missing data is performed by the Mean Squared Error (MSE). The results show that using a larger number of iterations and variance of data in training can help the ANFIS to obtain more precise results.

Key words: ANFIS, missing data, reference evapotranspiration.

Fecha recepción: Agosto 2014 **Fecha aceptación:** Diciembre 2014

Introduction

The national's water Commission (CONAGUA), through the National Weather Service is the official source of weather and climate data in Mexico. Monitoring and registration of such data is done by Automatic Weather Stations (EMAs), which are distributed around all the Mexican Republic. However, meteorological records in many cases information is incomplete due to different factors, among them: flaws in operation and calibration of the instruments, in the maintenance of the station and its instrumentation. For example, the meteorological station of Chapingo, administered by the Mexico Valley Watershed Organisation (OCAVM, located in the municipality of Texcoco, State of Mexico, with geographical coordinates of latitude: 19 ° 50' and longitude West: 98 ° 88') carries out the monitoring of different meteorological variables but occasionally records present lack of data. This situation affects the accuracy of the results when making important calculations as the estimation of the water requirements of crops in areas of irrigation. However, to meet the water requirements of crops it is necessary to calculate the loss of water by evaporation and crop transpiration, which is why the global organization of the

United Nations for food and Agriculture (FAO) in its Guide to the water needs of crops (Rivera, 2008) it introduces the concept of evapotranspiration of cultivation of reference (ETo), which studies the evapotranspiration rate regardless of the type of crop and soil characteristics (Doorenbos & Pruitt, 1997).

Given its definition, the factors affecting ETo are climatic factors can be calculated climate parameters such as temperature, wind speed, relative humidity and solar radiation. This data is provided by the EMA, but as mentioned above suffer from data loss whose effect may be insignificant, but when time increases with data loss monitoring database becomes unreliable. To solve the problem of missing records, the literature suggests the use of different techniques, from traditional, such as linear regression, so-called artificial neural networks; but determining which model is more efficient, some studies found no difference between the results. For their part, they tend to support others with slight superiority to artificial neural networks (Pitarque & Roy, 1998). Thus, to implement a technique to be considered the state of the environment which will develop the study to obtain the desired results.

It is worth mentioning that the Automatic Weather Station Chapingo lacks a history data record one hour period in previous years, the data of the station may have different percentages of missing data; You do not have any other parameter that directly relates to the measured data. Also, you can not extrapolate data from a weather station since they are worse, limiting the options of solution. Therefore, it is considered to use the technique of fuzzy logic because you can work with all complete and incomplete records, and create different scenarios according to the behavior of each meteorological variable during the year and during the day. This allows us to formulate the rules of inference and fuzzy sets for each variable. The objective of this work is to design and implement the ANFIS model for an approximate model of the behavior of meteorological variables from the data recorded or obtained from the Meteorological Station Chapingo, and estimate missing data.

The paper is organized as follows: section history where related to the recovery of missing data are revised jobs; neurofuzzy Focus section, this section elements and relationships that make up the structure of ANFIS network are shown; the proposed through the ANFIS Model section presents the modeling of fuzzy sets, the range of values and functions of our neuro network; Experiment and results section shows the implementation of the neuro network in the calculation of evapotranspiration and estimated RSME; Finally, the Conclusions and Future Work section shows an overview of the benefits of ANFIS model in satisfactory estimate of missing data.

Background

In the literature there have been different approaches to the problem of missing data filling depending on the variable to be completed in records; research (Campos Quispe, & Tatiana, 2012) management methodology for work on meteorological precipitation data and the application of the method of filling the inverse equation of the Euclidean distance to the correlation matrix is proposed if It is that spaces are presented. The results were satisfactory with the equation of the inverse of the Euclidean distance with the correlation matrix, allowing a variation in both the mean and variations that occur during the time period handled. However, the study concludes that during the selection of stations should be careful selecting stations with correlation coefficients greater than 0.75 percent and less than 20% empty, which ensures reliable data filled since they are percentages of variation ranging from 0 to 15%. In this writing (Ferreira, 2003) various methods of analysis and imputation of incomplete data, from the standpoint of its application to complete the missing values in series of wind speed were studied; his results are interesting for wind speed. In the work described (R. Alfaro, Alfaro, & Pacheco, 2000) different methods for filling gaps in the series of annual precipitation applied to records of meteorological stations located in different regions of Costa Rica is. Filling methods used to reproduce a series of data they estimated using at least one station near the station study are: simple regression, the reason why the set, multiple regression and normal reason. The results of the study show absolute maximum differences between actual and estimated values of about 30%, suggesting the use of more complex from those presented in this study if you want to make more precise estimates of annual precipitation data methods. Reference (Valesani & Quintana, 2009) concerns the application of Artificial Neural Network (ANN) as imputation method simulating absence of data rates using the MCR (Missing Completely at Random) technique. Its efficiency was then evaluated in different situations in order to assess the performance with different parameters such as MAE (mean absolute error), RMSE (Mean Square Error), and regression in order to determine whether the RNA are suitable for data imputation in this particular case. The results shown in research are satisfactory and deemed acceptable in statistical terms. Similarly, the proposal (Solana & Boat, 1998) highlights the Artificial Neural Networks mentioning technical concepts as well as specific applications in the field of information retrieval. The proposal (Cruz, 2012) a methodology for filling of missing meteorological data developed in the Matlab programming environment using the interpolation technique Piecewice Cubic Hermite Interpolating Polynomial (PCHIP) featuring software is established. To evaluate the performance of the interpolation root mean square error (RMSE) is used. Note that the results on the filling of missing data are not presented in this work. To solve the problem of missing meteorological data have been used some traditional techniques such as regression, the homogeneity of nearby stations and use the latest techniques Artificial Intelligence, based on fuzzy logic (inductive learning), genetic algorithms and neural networks (Saba & Ortega, 2008) (Alfaro & Soley, 2009), (Chen, 1995); however, the problem arises when we found conflicting results when determining which models are more efficient in solving the problem of lack of data.

This research has considered using as a technique to fill in missing data in the Automatic Weather Station Chapingo to neuro-fuzzy networks. Thus the study of artificial intelligence in complex nonlinear systems is reinforced.

Focus neurofuzzy

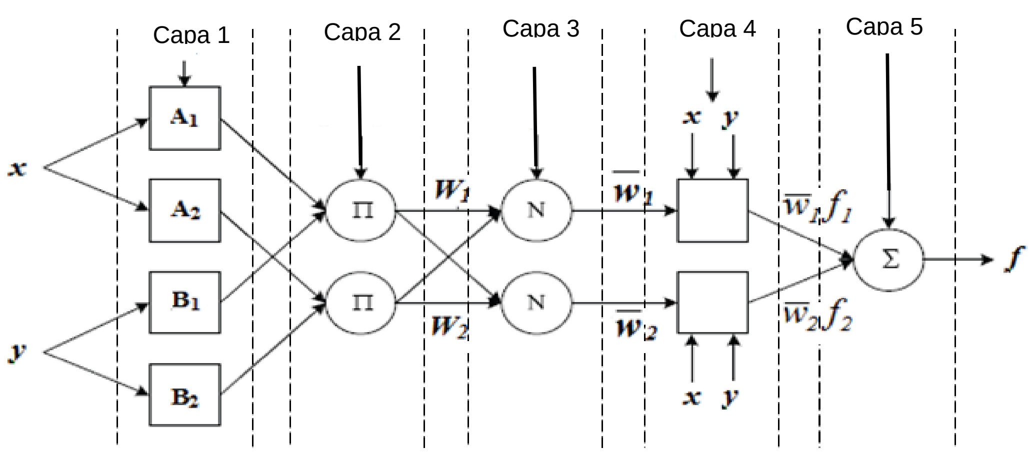
The neurodifusas networks are systems that take advantage of features of neural networks as the ability to learn or adjust itself and generalize, added to the characteristics of fuzzy logic, which works with logical reasoning based on membership functions that let you work with linguistic variables, very natural for humans. The fuzzy inference systems can represent knowledge based on if-then rules, but lacks the ability to adapt when external conditions change. For this reason, the concept of recognition of neural networks are included.

Neurofuzzy the basis of systems found in the fuzzy neurons, based emulate biological neuron morphology, followed by a learning system feature more diffuse. We can classify neurons diffuse into two classes: first, diffuse characteristic is the description of the synaptic weights, and second, the signals transmitted are diffuse with the synaptic weights.

We must take into consideration very important actions for the development of neural diffusion systems: a) definition of the input and output values, b) definition of fuzzy sets that are required to use, c) definition of fuzzy rules, d) structuring the neural network and e) modeling of synaptic connections which can incorporate fuzzy interpretation.

In recent decades the neural diffusion systems have been positioned in important applications in different areas such as: control (in most systems); quantitative analysis (operations, data management); inference (expert systems for diagnosis, planning and forecasting, natural language processing, robotics, software engineering) and information retrieval (database), among other applications (Lin, Lee, & CS, 1996).

Therefore, for this study it was decided to model the dynamics of the process through the first neuro system known and most established, ANFIS; besides being one of the pioneering works, it is one of the simplest computationally (Jang., 1993). ANFIS implements the Takagi-Sugeno model for the structure of if-then rules of the fuzzy system. A ANFIS model is composed of five layers in which all nodes of the same layer have a similar function. The first layer is used for inputs. The last layer for output and has 3 hidden intermediate layers. The number of hidden layers remains constant in all types of ANFIS to implement, regardless of the entries that have the system and has only one way out. In Figure 1 the five layers of ANFIS network nodes and the relationship between input linguistic variables, nodes rules, standardized rules and parameters are displayed, then the function of the network layer by layer is explained.



ANFIS Figure 1. System Architecture [16].

The behavior of each layer ANFIS described below:

Layer 1: Each node i of this layer is adaptable, ie has adjustable parameters and described by equation 1.$0_i^1 = \mu A_i(x)$

(1)

Where x is the input node i, Ai is a linguistic variable associated with the function of this node. In other words, 0i1 is the membership function of Ai and specifies the degree of membership of x with respect to Ai.

Layer 2: Each node in this layer is labeled with Π (Figure 1). This layer input signals are multiplied and the product at the outlet, that is, when multiple signals are input to this node resulting product sent to each instance i is sent. For a moment: $0_i^2 = \mu A_i(x)\,\mu B_i(y),\ i=1,2$

(2)

Each output node represents the degree of activation of a rule. They also represent the T-rule or T-conorm to model the logical AND and OR operations. They are often referred to as nodes of rules.

Layer 3: Each node labeled N (Figure 1), indicating the normalization of the activation degree. The ith node calculates the normal activation rules with the sum of all activated rules according to equation 3. The outputs of this layer may be called activation of standardized rules.$0_i^3 = w1 = w_i / w_i + w_2$ (3)

Layer 4: Each node i in this layer is square and has a function of node:

$$0_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i),\ i=1,2 \quad (4)$$

Where $\bar{w}_i$ It is the output layer 3 and the parameter set {pi, qi, ri} They are referred to parameters accordingly.

Layer 5: Circular presents a single node labeled ∑, here the output is calculated from the input signals (Equation 5):$0_i^5 = f = \sum w_i f_i = \sum_i w_i f \,/\, \sum_i w_i = w_i f_i + w_2 f_2$

(5)

The training process is performed with two sets of parameters: the antecedent (constant characterizing the membership functions) and the result (coefficients of linear functions of the consequent of the rules). Only links between nodes indicate the direction in which signals flow, do not have associated weights (Villada F & Garcia).

Proposed model

To implement the ANFIS actual climatic data sets were used, which we get from the weather station Chapingo. The meteorological variables were recorded every 10 minutes during daily periods September 2013 to the current day. Modeled variables are wind speed (VELS), temperature (TEMP), relative humidity (RH) and solar radiation (RAD-SOL). In the neuro model variables are assigned H (Time) and Station (EA) as input variables and meteorological variables wind speed, temperature, relative humidity and solar radiation as output variables. Time variables and are dependent Station of the Year, while wind speed, temperature, relative humidity and solar radiation are dependent on the first two. As linguistic values of H (Hour) are proposed: Early morning (M1), Tomorrow (M2), Noon (M3), Sunset (A1), Dusk (A2) and Night (N). Fuzzy set for time of day shown in section A of Figure 2.

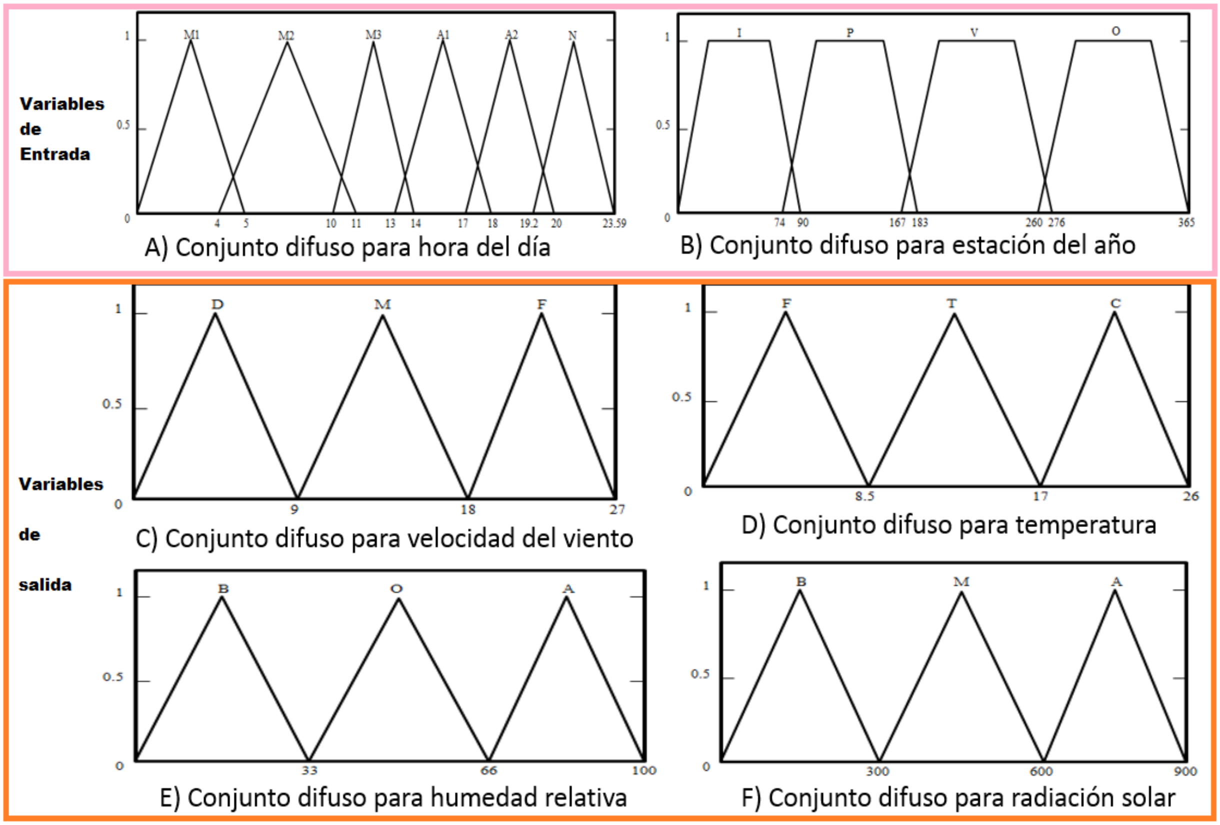


Figure 2: Fuzzy sets for: a) time of day, b) season, c) wind speed, d) temperature, e) and f relative humidity) solar radiation

Not to incur a minimum poorly defined states crosslinking was used, considering the behavior of the meteorological variables during the day, so that each element remains represented in at least two membership functions. The triangular function to specify the height of each linguistic label is used.

Because of the difficulty of establishing a uniform scale in the length of the months of the year, it was decided to distribute the days of the year based on the seasons. Thus in this study the first of the year was allocated to the first day of winter, which is on December 21 and 90 is the day the March 21 final day of winter.

As fuzzy sets and linguistic variable label proposed season: Winter (I), Spring (P), Summer (V) and Autumn (O). The membership function of the variable season shown in section B of Figure 2. A cross-linking was carried out considering the behavior of the weather because in the changing seasons of the year in the days of the start of a variable and the last days of the other variable behavior is similar. Also, the trapezoidal function was used as in the seasons there is a gap where the behavior of the variables is stable and maintained for a period.

The ranges for the four linguistic variables are: wind speed [0, 27] in km / h, temperature [0, 26] in C °, relative humidity [0, 100] and solar radiation in% [0, 900] in wm2. The ranges are average values obtained each season. Fuzzy sets with their respective linguistic labels that make up each of the variables were: wind speed [Weak (D), Moderate (M) and Strong (F)] see subsection C of Figure 2; for temperature [Frio (F), temperate (T), Warm (C)] see subsection D of Figure 2; for relative humidity [Low (B), Moderate (M) and High (A)] see subsection E in Figure 2 and solar radiation [Low (B), Moderate (M) and High (A)] see subsection F Figure 2.

The shape of the rules used for wind speed is as follows: IF Station is x AND THEN VELS Hora and is z. See Table I.

Table I. inference rules for wind speed.

# Regla	Estación	Hora	VELS	# Regla	Estación	Hora	VELS
1	I	M1	D	14	O	M1	D
2	I	M1	M	15	O	M1	M
3	I	M2	D	16	O	M2	D
4	I	M2	M	17	O	M2	M
5	I	M3	D	18	O	M3	D
6	I	M3	M	19	O	M3	M
7	I	A1	D	20	O	A1	D
8	I	A1	M	21	O	A1	M
9	I	A2	D	22	O	A2	D
10	I	A2	M	23	O	A2	M
11	I	N	D	24	O	N	D
12	I	N	M	25	O	N	M
13	I	N	F	26	O	N	F

To temperature is the form IF station is x AND THEN TEMP Time is and z is as shown in Table II.

Table II. Temperature inference rules.

# Regla	Estación	Hora	TEMP	# Regla	Estación	Hora	TEMP
1	I	M1	F	15	O	M1	T
2	I	M1	T	16	O	M1	C
3	I	M1	C	17	O	M2	T
4	I	M2	F	18	O	M2	F
5	I	M2	T	19	O	M3	T
6	I	M3	F	20	O	M3	F
7	I	M3	T	21	O	A1	F
8	I	A1	F	22	O	A1	T
9	I	A1	T	23	O	A1	C
10	I	A1	C	24	O	A2	C
11	I	A2	T	25	O	N	C
12	I	A2	C	26	O	N	T
13	I	N	T				
14	I	N	F				

Similarly, for Relative humidity it is as follows: IF Station is x AND THEN Hora and HR is z. See Table III.

Table III. Inference rules for Relative Humidity

# Regla	Estación	Hora	HR	# Regla	Estación	Hora	HR
1	I	M1	B	16	O	M1	O
2	I	M1	O	17	O	M1	A
3	I	M1	A	18	O	M2	O
4	I	M2	B	19	O	M2	A
5	I	M2	O	20	O	M3	A
6	I	M2	A	21	O	A1	O
7	I	M3	O	22	O	A1	A
8	I	M3	A	23	O	A2	B
9	I	A1	B	24	O	A2	O
10	I	A1	O	25	O	A2	A
11	I	A1	A	26	O	N	A
12	I	A2	B	27	O	N	O
13	I	A2	O	28	O	N	B
14	I	N	B				
15	I	N	O				

And finally, solar radiation is as follows: IF Station is x AND THEN Hora and RAD-SOL is z. See Table IV.

Table IV. Inference Rules for Solar Radiation

# Regla	Estación	Hora	RAD-SOL	# Regla	Estación	Hora	RAD-SOL
1	I	M1	B	12	O	M1	B
2	I	M2	B	13	O	M1	B
3	I	M3	B	14	O	M3	B
4	I	A1	B	15	O	A1	B
5	I	A1	M	16	O	A1	M
6	I	A1	A	17	O	A1	A
7	I	A2	M	18	O	A2	B
8	I	A2	A	19	O	A2	M
9	I	N	B	20	O	A2	A
10	I	N	M	21	O	N	B
11	I	N	A	22	O	N	M
				23	O	N	A

The rules are stored in the inference engine also known as a control center, for in it are orders that must operate. To model ANFIS used the Matlab Fuzzy Toolbox. Select the values that will be processed at the entrances time and season of the year, passing the fuzzifier, this translates into a language understandable to the inference engine, which associates each input to a fuzzy set returning as an output a numerical value and not diffuse.

Experiment and results

For the membership functions of the fuzzy sets of wind speed, temperature, relative humidity and solar radiation proceeded to collecting daily data on the above variables in a given by CONAGUA period, through the National Weather Service and the use of weather indicators of the study area where the predominant climate is temperate semi-dry. This also yielded annual average of each of the variable values to generate the ranges of each fuzzy set, and thus discard outliers.

The EMAs perform collection and monitoring meteorological variables to generate files on average every 10 minutes from all the variables, this information is sent via satellite at intervals of 1-3 hours per season. The station data are presented in three formats in accordance with the frequency of data collection stations. The report corresponds to one hour 10 minute period, the period of 24 hours 60 minutes 90-day report is the corresponding 24-hour period, which get 144 records each weather variable.

For research is considered to solve the problem of missing reports in an hour meteorological data, since it is they who have higher absence of data, and then to calculate the ETo for periods times.

The neuro network training is performed for each of the weather variables, which consists of a minimum of times (number of iterations) for processing the chosen samples. These samples vary in size according to the number of rules for each weather variable, which were mentioned in the previous section, as well as the season in question. The evaluation of each of the networks are based on data from one day to the winter season (8th) corresponding to 28 December and for the autumn season (day 279) corresponding to 25 September. The data for training were selected according to the formulation of each rule during the day evaluated.

Tables V and VI show some peculiarities that were watching the neuro network training. The data in the course of the day are 144 per meteorological variable; however, this number may vary according to the season failures; fuzzy rules as mentioned above vary with respect to the course of each season. The total training data is obtained for the number of times by the number of fuzzy rules for each weather variable. Training elements correspond to those duplicates or filler before or after day that enabled us to meet the fuzzy rules on records evaluated. The actual data indicate the number of accurate data that satisfies the neuro network training in the corresponding number of iterations, also can be seen that are not evenly distributed within each fuzzy set of the time variable.

Table V. Training Network neuro 5 times

5 Épocas								
	Día 279 Otoño				Día 8 Invierno			
Variable Meteorológica	VELS	TEMP	HR	RAD-SOL	VELS	TEMP	HR	RAD-SOL
Datos en el transcurso del día	144	144	144	144	144	144	144	144
Reglas Difusas	13	12	13	12	13	14	15	11
Elementos en el Entrenamiento (datos faltantes rellenados por duplicidad o relleno de datos anterior o posterior)	17	27	25	18	25	35	45	15
Datos reales (satisficieron las reglas)	48	33	39	42	38	35	30	40
Total de Datos de entrenamiento (época * n de reglas)	65	60	65	60	65	70	75	55

Table VI. Neuro training network with 10 Times

10 Épocas								
	Día 279 Otoño				Día 8 Invierno			
Variable Meteorológica	VELS	TEMP	HR	RAD-SOL	VELS	TEMP	HR	RAD-SOL
Datos en el transcurso del día	144	144	144	144	144	144	144	144
Reglas Difusas	13	12	13	12	13	14	15	11
Elementos en el Entrenamiento (datos faltantes rellenados con relleno de datos anterior o posterior)	47	55	61	51	60	72	92	39
Datos reales (satisficieron las reglas)	83	65	69	69	70	68	58	71
Total de Datos de entrenamiento (época * n de reglas)	130	120	130	120	130	140	150	110

As shown in the above tables, the number of elements of the training does not compare with the total data for training in both tables, this means that although the number of data for training is increased, a fraction corresponding to actual data . Similarly, although it went from 5-10 times the actual data in the table V no VI corresponding to 10 times doubled in the table, this due to two main factors: the day does not fully satisfy the fuzzy rules in addition to the range comprising each fuzzy set variable time (Dawn, Morning, Noon, Sunset, Dusk and Night) it is different, resulting in variation in the actual data on training in each iteration. Therefore, space fillers that are not covered in each iteration of the training are made taking into account data that is closer to the day evaluated, which may be later or earlier days as applicable, or equally a fact repeat in one iteration.

Therefore, training and validation of the neuro network took 279 days for the fall season and day 8 for the winter season. The performance of each network was done with the square root of the mean square error (RMSE, see Table VII) calculated by the following equation 6.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - x_i)^2} \qquad (6)$$

Where N is the number of considered observations, xi is the real value and yi is the value estimated by the model. Table VII can observe the data obtained with the neuro network training for 5 to 10 times.

Table VII: results from the neuro network

	VELS	TEMP	HR	RAD-SOL
Datos evaluados en el transcurso del día	144	144	144	144
Rango de Hora de día	0:00-23:50	0:00-23:50	0:00-23:50	0:00-23:50
Valor para estación del año	otoño	otoño	otoño	otoño
Rango de valores para la variable	0-27 Km/h	0-26 º C	0-100 %	0-900Wm^2
	Día 279 Otoño			
RMSE con 5 épocas de entrenamiento	2.06	0.95	1.97	381.61
RMSE con 10 épocas de entrenamiento	2.05	0.53	1.89	92.53
Diferencia en precisión	0.01	0.42	0.08	289.08
	Día 8 Invierno			
RMSE con 5 épocas de entrenamiento	2.37	2.98	3.60	46.48
RMSE con 10 épocas de entrenamiento	1.43	0.88	2.35	44.84
Diferencia en precisión	0.94	2.1	1.25	1.64

According to the results obtained for improved weather variables shown to increase the number of times it with 10 times the RMSE error is reduced. Regarding the behavior of the variables, you can clearly see an improvement in the adjustment of the estimated 10 times with respect to the actual data as the lines on the graph show several sections of coincidence data. See Figure 3.

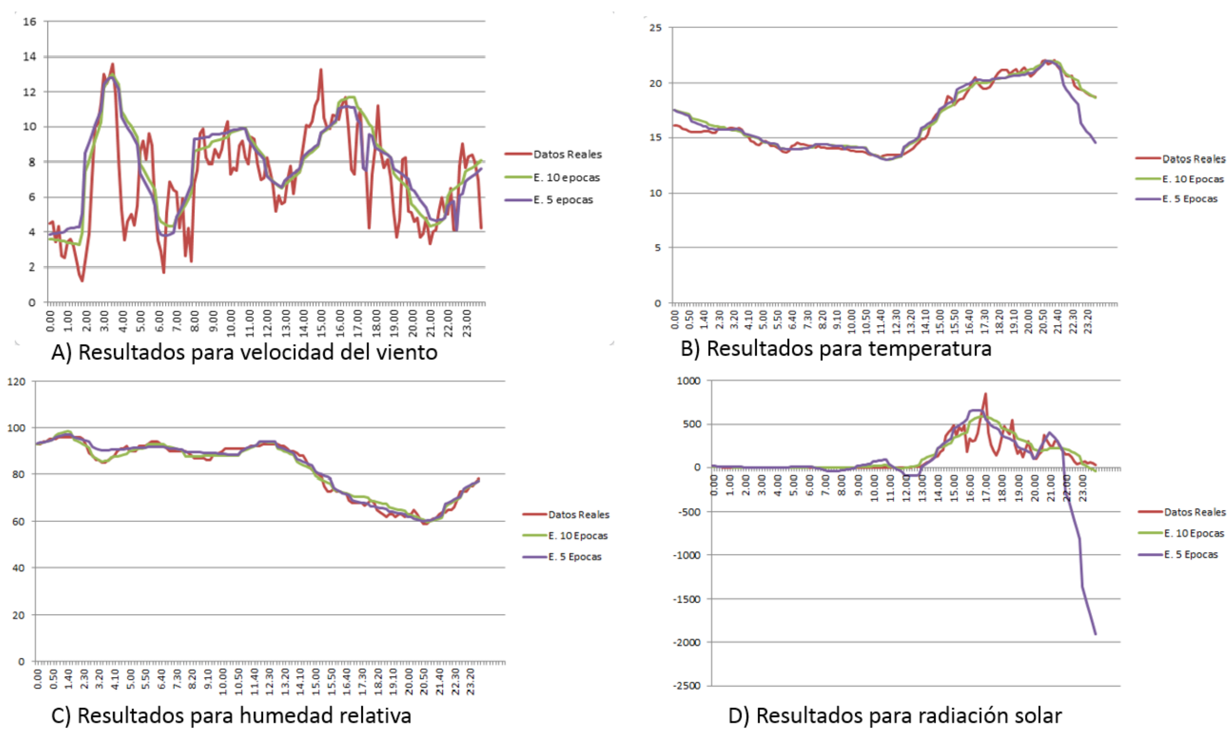


Figure 3: Results for: A) wind speed, b) temperature, c) relative humidity, D) solar radiation

To assess the usefulness of the neuro network in estimating reference evapotranspiration proceeded to its estimate following the procedures for periods of times times pointing Allen (2006), who explains that in implementing the FAO Penman-Monteith for periods time hours or less, the equation and some meteorological data to calculate procedures should be adjusted to these periods, so the FAO Penman-Monteith equation for hourly calculations are amended as follows:

$$ETo= (0.408 \cdot \Delta \cdot (R_n\text{-}G)+y \cdot 37/(T_{hr}+273) \cdot u_2 \cdot (e° (T_{hr})\text{-}e_a))/(\Delta+ \gamma \cdot (1+0.34 \cdot u_2) \qquad (7)$$

Where:

ETo reference evapotranspiration [mm $hora^{-1}$]
Rn net radiation at the surface of the culture [MJ m^{-2} $hora^{-1}$],
G Soil heat flow [MJ m^{-2} $hora^{-1}$]
T_{hr} average air temperature every hour [°C]
Δ slope of the curve of saturation vapor pressure in T_{hr},
γ psychrometric constant [kPa $°C^{-1}$],
$e°(T_{hr})$ saturation vapor pressure at air temperature T_{hr},
e_a average hourly actual vapor pressure [kPa],
u^2 average hourly wind speed [m s^{-1}]

The modified formula for hourly periods ETo with hourly data is calculated in periods of light from 8:00 to 19:00 hours on 25 September (279 autumn) and December 28, 2013 (8th winter) in Chapingo located at 19 ° 50 'north latitude and 98 ° 88' west longitude and 2800 meters above sea level.

Table VIII the results obtained with real data and estimated data on 25 September, covering the period 8:00 to 18:00 hours a day at intervals of one hour show.Tabla VIII. Cálculo del RMSE de la ETo real y estimada del 25 de septiembre y 8 de diciembre

Hora del día	Eto Real 25 sep	Eto Estimada 25 sep	RMSE	Eto Real 8 dic	Eto Estimada 8 dic	RMSE
08:00	0.029	0.023	2.6807E-06	0.031	0.044	1.3919E-05
09:00	0.027	0.039	1.288E-05	0.030	0.068	0.00013122
10:00	0.030	0.053	4.6105E-05	0.038	0.412	0.01269089
11:00	0.029	0.024	2.2721E-06	0.036	0.027	7.598E-06
12:00	0.031	0.028	6.0344E-07	0.047	0.079	9.5851E-05
13:00	0.087	0.154	0.00040424	0.221	0.383	0.00240965
14:00	0.296	0.293	1.2983E-06	0.702	0.679	4.9402E-05
15:00	0.340	0.431	0.0007513	0.927	0.990	0.00035727
16:00	0.677	0.707	8.1985E-05	1.198	1.254	0.00028123
17:00	0.260	0.288	7.042E-05	0.734	0.942	0.00395536
18:00	3.654	3.034	0.03500723	9.098	9.518	0.01601642
		RMSE =	0.19073808		RMSE =	0.18975991

This calculation is important because with its help you can get other important estimates for water management, such as the calculation of the net irrigation for different crops.

For example, it is the estimate of the net irrigation spinach with data on 25 September (279 tests) and a rainfall of 110.95 mm. The net irrigation is estimated with the formula 8:

$$ETc = ETo * Kc \quad (8)$$

To calculate the net irrigation requirements (Nn) will use the following formula:

$$(Nn) = ETc - Pe \quad (9)$$

Dónde:

Kc= Crop coefficient (determined by the stage of crop development)
ETo= Reference Evapotrasnpiración
ETc = Daily needs irrigation farming.
effective rainfall (Pe)= 0.8 P – 25
Pe = 0.8 (110.95) -25 = 63.76 mm

Table IX shows the estimated ETo and ETc with Kc2 because spinach for September 25 is in its second phase of development and has a value of 1 and Table X shows the estimate of the need for net irrigation along with RSME.

Table IX. Calculation of crop evapotranspiration

	Datos Reales			Datos Estimados		
Hora	**ETo**	**K_{c2}**	**ETc**	**ETo**	**K_{c2}**	**ETc**
08:00	0.031	1	0.031	0.044	1	0.044
09:00	0.03	1	0.03	0.068	1	0.068
10:00	0.038	1	0.038	0.412	1	0.412
11:00	0.036	1	0.036	0.027	1	0.027
12:00	0.047	1	0.047	0.079	1	0.079
13:00	0.221	1	0.221	0.383	1	0.383
14:00	0.702	1	0.702	0.679	1	0.679
15:00	0.927	1	0.927	0.99	1	0.99
16:00	1.198	1	1.198	1.254	1	1.254
17:00	0.734	1	0.734	0.942	1	0.942
18:00	9.098	1	9.098	9.518	1	9.518

Hora	**Nn con datos reales**	**Nn con datos estimados con la Red Neurodifusa**	**RSME**
08:00	0.031	0.044	1.5364E-05
09:00	0.03	0.068	0.00013127
10:00	0.038	0.412	0.012716
11:00	0.036	0.027	7.3636E-06
12:00	0.047	0.079	9.3091E-05
13:00	0.221	0.383	0.00238582
14:00	0.702	0.679	4.8091E-05
15:00	0.927	0.99	0.00036082
16:00	1.198	1.254	0.00028509
17:00	0.734	0.942	0.00393309
18:00	9.098	9.518	0.01603636
		RSME	0.03601236

Conclusions

Satisfactory results were obtained in the estimation of missing data for the Chapingo season with the implementation of neuro-fuzzy, network with which various scenarios were designed; for example, the design of a neuro-fuzzy network formulating fuzzy sets for the period of one day, one month and one year. The latter produced better results (presented in the research) because the number of fuzzy sets and data for training can be tailored with anterior and posterior days is reduced. For the period of one month, a greater number of fuzzy sets should be created, resulting in the problem of ordering data to meet each of the rules also should consider that within months no similarity between the behavior of the variables.

After implementing the neuro-fuzzy network accurate results were obtained with increasing training iterations in the network; however, it is clear that there is a possibility that not only was the number of iterations but to pass each iteration training input data satisfy each of the fuzzy rules. Thus the duplication of data in subsequent iterations are avoided, which leads to a good distribution of data within the training.

The proposal of this work is to test based neuro-fuzzy for filling data networks because of their ability to solve problems related to the uncertainty of the information or expert knowledge model. Thus, the neuro-fuzzy network is considered a good option depending on the type of data available and the format in which (Excel file format) are published.

Our model can be improved if there is a history of records from previous years as well as new entry would annex the year under review in the network, which would compare data from previous years or predict future data.

You can also improve network performance through a system that can generate the training data in an automated manner and that the investigation was carried out manually, this would facilitate the feedback to the network and establish a point at which he can overtraining occur. Although fuzzy logic has a short history, it is a promising technique in the field of recovery of meteorological information.

Bibliography

Alfaro, J., & Javier, F. (2008). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*, 60-758.

Alfaro, J., & Soley, J. (2009). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*.

Arca, B. B. (2001). Evaluation of neural network techniques for estimating evapotranspiration. En *Evolving Solution with Neural Networks,* Italia: Baratti, R. y De Canete, pp. 62-97.

Campos, A., Quispe, S., & Tatiana, S. (2012). Gestión de datos meteorológicos. *XXII Congreso Nacional de Hidráulica.* Acapulco Guerrero.

Chen, H. (1995). *Machine learning for information retrieval: neural networks, symbolic learning, and genètic algorithms. Journal ofthe American Soclety for Information Science.*

Cruz, H. (2012). *Estimación de la evapotranspiración de referencia en regiones con datos climáticos limitados.* Texcoco.

Doorenbos, J., & Pruitt, W. (1997). *Guidelines for predicting crop water requirements.* Rome: Irrigation and Drainage Paper.

Elizondo, D. G. (1994). Development of a neural network model to predict daily solar radiation. Agricultural and Forest Meteorology, pp. 115-132.

Enke W, S. A. (1997). Downscaling climate model outputs into local and regional weather elements by classification and regression. *Climate Research*, pp. 195-207.

Ferreira, M. (2003). Metodologías de análisis e imputación de datos faltantes en series de velocidad del viento. *VI Congreso Galego de Estatística e Investigación de Operaciónes*, pp. 5-7.

Jang., J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems*, pp. 665-685.

JE, K. (2001). Empirical eigenvectors of sea-level pressure, surface temperature and precipitation complexes over North America. *Cagliari*.

Lin, C., Lee, T., & CS, G. (1996). *Neural Fuzzy System: A Neuro-Fuzzy Synergism to Intelligent Systems.* Prentice Hall.

M. del Brio Bonifacio, S. M. (2010). *Redes Neuronales y Sistemas Borrosos.* Madrid. España: Alfaomega.

Pitarque, A., & Roy, J. (1998). Redes neuronales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificacíon, Valencia, España: Universidad de Valencia, pp. 387-400..

R. Alfaro, R. P., Alfaro, R., & Pacheco, R. (2000). Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica. *Redalyc*.

Rivera, M. (2008). *Estimación estadística de valores faltantes en series históricas de lluvia.* Pereira: Facultad De Ingeniería Industrial, Escuela de Posgrados.

Saba, I., & Ortega, J. (2008). Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales.

Solana, H., & Bote, G. (1998). *La aplicación de Redes Neuronales Artificiales (RNA): a la recuperación de la información*.

Valesani, E., & Quintana, P. (2009). Imputación de datos con redes neuronales. *XI Workshop de Investigadores en Ciencias de la Computación*, Argentina, pp. 281-285.

Villada F, & García, E. (s.f.). Pronóstico del Precio de la Energía Eléctrica usando Redes Neuro-Difusas. *Redalyc*.

Wilby R, W. T. (1997). Downscaling general circulation model output: a review of methods and limitations. *Progress in Physical Geography*, pp. 530-548.