Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)
Available online at www.inia.es/sjar

Spanish Journal of Agricultural Research 2010 8(S2), S162-S171
ISSN: 1695-971-X
eISSN: 2171-9292

Development of relative humidity models by using optimized neural network structures

A. Martinez-Romero*, J. F. Ortega, J. A. de-Juan, J. M. Tarjuelo and M. A. Moreno

Centro Regional de Estudios del Agua (CREA). Universidad de Castilla-La Mancha.
Ctra. de Las Peñas, km 3,4. 02071 Albacete. Spain

Abstract

Climate has always had a very important role in life on earth, as well as human activity and health. The influence of relative humidity (*RH*) in controlled environments (*e.g.* industrial processes in agro-food processing, cold storage of foods such as fruits, vegetables and meat, or controls in greenhouses) is very important. Relative humidity is a main factor in agricultural production and crop yield (due to the influence on crop water demand or the development and distribution of pests and diseases, for example). The main objective of this paper is to estimate *RH* [maximum (RH_{max}), average (RH_{ave}), and minimum (RH_{min})] data in a specific area, being applied to the Region of Castilla-La Mancha (*C-LM*) in this case, from available data at thermo-pluviometric weather stations. In this paper Artificial neural networks (*ANN)* are used to generate *RH* considering maximum and minimum temperatures and extraterrestrial solar radiation data. Model validation and generation is based on data from the years 2000 to 2008 from 44 complete agroclimatic weather stations. Relative errors are estimated as 1) spatial errors of 11.30%, 6.80% and 10.27% and 2) temporal errors of 10.34%, 6.59% and 9.77% for RH_{min}, RH_{max} and RH_{ave}, respectively. The use of *ANNs* is interesting in generating climate parameters from available climate data. For determining optimal *ANN* structure in estimating *RH* values, model calibration and validation is necessary, considering spatial and temporal variability.

Additional key words: artificial neural networks; climate data; limited data.

Resumen

Desarrollo de modelos de humedades relativas mediante el uso de redes neuronales optimizadas

El clima ha tenido siempre un gran impacto en la vida en la Tierra, la salud, actividad y cultura del hombre. La influencia de la humedad relativa (*HR*) tiene especial relevancia en ambientes controlados (por ejemplo, procesos industriales de elaboración y conservación de productos agroalimentarios, o el control en invernaderos). En el ámbito agrario también es un elemento fundamental, ya que puede modificar el rendimiento de los cultivos y la incidencia de plagas y enfermedades, sin olvidar su influencia directa en la demanda evaporativa de la atmósfera. El objetivo principal de este estudio es estimar valores de *HRs* ambientales (máximas, medias, y mínimas) en un territorio concreto, en este caso aplicado a Castilla-La Mancha (España) a partir de registros disponibles en estaciones termopluviométricas. Para ello, se ha planteado el uso de herramientas basadas en redes neuronales artificiales (*RNAs*), que permitan generar datos de *HRs* a partir de registros de temperaturas máximas y mínimas, y datos de radiación solar extraterrestre. Para generar y validar estos modelos se utilizan registros de los años 2000-2008, procedentes de 44 estaciones agroclimáticas completas ubicadas en esta comunidad autónoma. La metodología propuesta ofrece errores relativos espaciales del 11,30%, 6,80% y 10,27%, y temporales del 10,34%, 6,59% y 9,77% para las *HRs* mínimas, máximas y medias, respectivamente. El uso de las *RNAs* es efectivo para la generación de parámetros climáticos a partir de otros datos disponibles. La determinación de una estructura óptima de *RNA* para estimar valores de *HRs* requiere la calibración y validación de los modelos considerando la variabilidad espacial y temporal de los datos.

Palabras clave adicionales: datos climáticos; limitación de datos; red neuronal artificial.

* Corresponding author: angel.mromero@uclm.es
Received: 15-10-09; Accepted: 07-07-10.

J. M. Tarjuelo is member of the SEA and J. A. de Juan is member of the SECH.

Abbreviations used: ANN (artificial neural netwok), C-LM (Castilla-La Mancha), ET_0 (reference evapotranspiration), MLP (multi-layer perceptron), P (precipitation), Ra (extraterrestial solar radiation), RE (relative error), RH (relative humidity), RH_{ave} (average relative humidity), RH_{max} (maximum relative humidity), RH_{min} (minimum relative humidity), RMSE (root mean square error), R^2 (determination coefficient), SIAR (Servicio de Información Agroclimática para el Regadío, Agroclimatic Information Service for Irrigation), SR (similarity rate), T_{ave} (average temperature), T_{max} (maximum temperature), T_{min} (minimum temperature.

Introduction

Climate has always had a very important role in human health and lifestyle. It forms an integral part of the criteria determining the location of agricultural production sites, recreational areas, urban development, and industrial areas. The availability of data on different climatic parameters allows for characterization of regional climate, and a large amount of climate records are needed to carry out a complete study. Several studies have been based on the use of different climate data or indicators calculated from basic climatic parameters. However, these works usually presents limitations due to problems with data availability and quality (Elías and Ruiz-Beltrán, 1981; De León *et al.,* 1988; Allen *et al*., 1998; Fount, 2000).

Different authors have proposed algorithms and techniques to perform data generation across several disciplines (Allen *et al*., 1998; Allison, 2001). Some have aimed to estimate climatic parameters, such as global radiation and reference evapotranspiration (ET_0), from basic climate data. However, estimation of relative humidity (RH) is less common (De la Casa *et al*., 2003; Singh *et al*., 2004; Popova *et al*., 2006), despite the importance of RH as a climatic factor in the agricultural sector. In fact, many plant pathologists consider RH the most important environmental factor in the development of plant diseases (Villalobos *et al*., 2002). RH contributes to determining final crop yield, affects stomata opening and has a direct influence on atmospheric evaporative demand (De Juan and Martín de Santa Olalla, 1993).

Artificial Neural Networks (*ANN*) have been used in generating various types of climate data: air temperature, RH, vapour pressure, drew point temperature, and reference evapotranspiration (Shank, 2003; Trajkovic *et al.,* 2003; Zanetti *et al.,* 2007). In this paper, Artificial Neural Networks are used to generate RH considering maximum and minimum temperatures and extraterrestrial solar radiation data. The main objective of this paper is to estimate RH [maximum (RH_{max}), average (RH_{ave}), and minimum (RH_{min})] data in the Region of Castilla-La Mancha (*C-LM*) from available data at thermo-pluviometric weather stations.

Material and methods

The proposed methodology can be summarized in Figure 1; the set of variables that best estimates RH

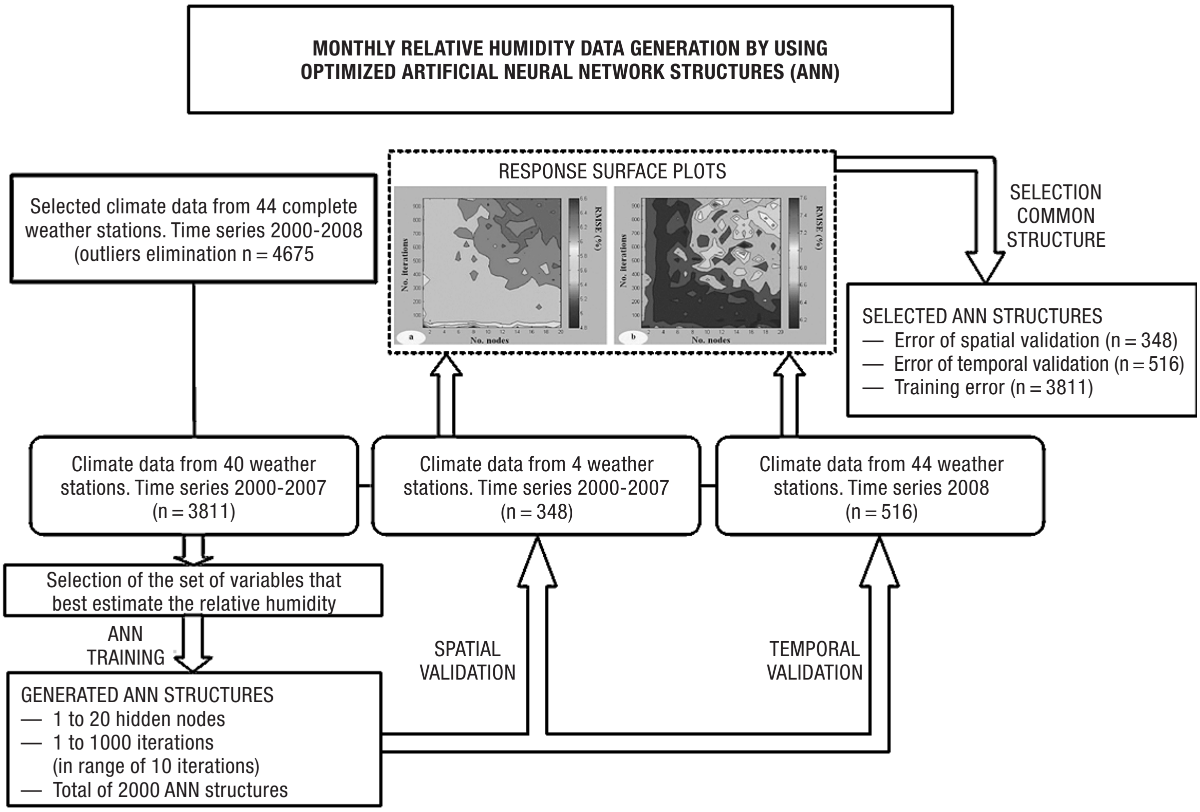

Figure 1. Diagram of the proposed methodology.

from the data series at 44 complete weather stations is selected. An improved structure is determined by analyzing the training error and the errors in spatial and temporal validation (defined by 10 nodes and 2,500 iterations) from a reference structure by using response surface plots.

The case study

The study location is in Castilla-La Mancha, Spain (*C-LM*; Fig. 2), which is a semiarid region with an area of 79,462 km^2. In this region, 132 weather stations are available with historical time series of monthly averages, maximum and minimum temperature data (T_{ave}, T_{max}, and T_{min}) and precipitation data, which is useful in characterizing local climate. The data from these weather stations have been checked and validated using different techniques for detecting data quality (Alexanderson, 1984, 1986).

Models to estimate RH from monthly temperature data are generated by using 44 agroclimatic weather stations with complete daily data from 2000 to 2008. These are included in the Agroclimatic Information Service for Irrigation (*SIAR*, «Servicio de Información Agroclimática para el Regadío») network (Fig. 2).

Artificial neural network models. Calibration and validation

Artificial Neural Networks (*ANN*s) have been used (Schalkoff, 1997; Nabney, 2002) to generate RH in the *C-LM* Region by using temperature, precipitation, and

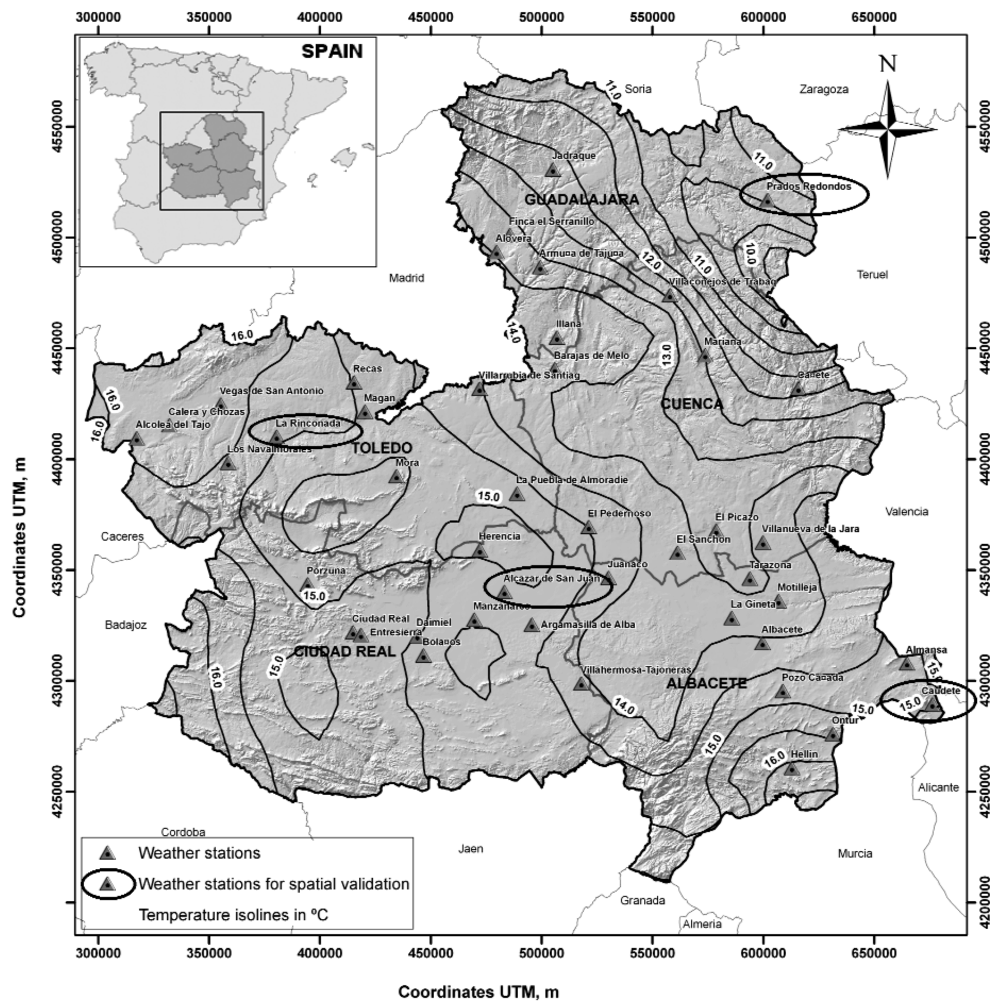

Figure 2. Location of the complete weather stations available and used in this study in Castilla-La Mancha (*C-LM*) during 2000-2008. Distribution of annual average temperature.

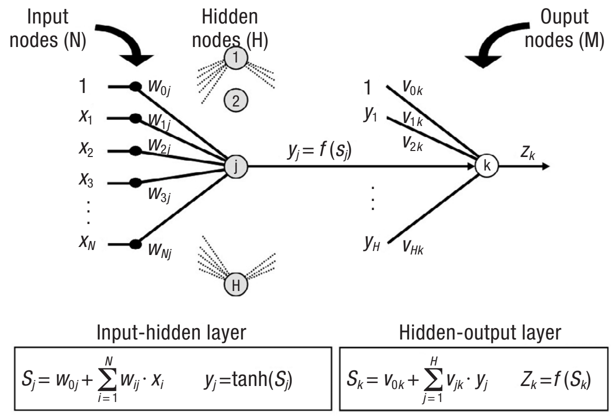

Figure 3. Artificial neural network structure used and associated equations. $S_j = w_{0j} + \sum_{i=1}^{N} w_{ij} \cdot x_i$, is the activation function; S_j, is the variable associated with each hidden unit; w_{0j}, are the bias parameters associated with the hidden units; w_{ij}, represents the elements of the first-layer weight matrix and x_i, are input values to the network; $y_j = \tanh(S_j)$, is the squashing function (hyperbolic tangent); y_j, are the outputs of the hidden unit; $S_k = v_{0k} + \sum_{j=1}^{H} v_{jk} \cdot y_j$, is the activation function; S_k, are second layer activation values; v_{0k}, are bias parameters associated with the hidden units; v_{jk}, represents the elements of the second-layer weight matrix and y_i, are input values to the network; $z_k = f(S_k)$, is the output-unit activation function; Z_k are output values; in regression problems, $Z_k = S_k$.

extraterrestrial solar radiation data (Ra) (values for *Ra* on the 15th day of the month, Allen *et al*., 1998). The general structure of the *ANN* is shown in Figure 3. Three layers are defined by three groups of neurons. Those layers are: input, hidden and output nodes. Each neuron implements a local function (Figs. 3 and 4). Input nodes and output nodes implement a linear function (Eq. [1] and Eq. [4], respectively), whereas hidden nodes implement a nonlinear function (Eq. [2]; Figs. 3 and 4).

The basic scheme of a neuron is composed by activation and squashing functions (Figs. 3 and 4). There are several activation functions that can be used, but the most common is the weighted sum function (Eq. [1], Nabney, 2002):

$$S_j = w_{0j} + \sum_{i=1}^{N} w_{ij} \cdot x_i \qquad [1]$$

A squashing function provides non-linearity to the structure. The hyperbolic tangent squashing function was used in this study (Eq. [2]):

$$y_j = \tanh(S_j) \qquad [2]$$

This squashing function follows the equation:

$$\frac{dy_j}{dS_j} = (1 - y_j^2) \qquad [3]$$

The hidden to output layer can have different activation functions depending on the goal of the problem. The logistic or softmax, among other possible functions, can be used for classification problems. In this case, the linear activation function is used (Eq. [4], Nabney, 2002) to solve a regression problem.

$$S_k = v_{0k} + \sum_{j=1}^{H} v_{jk} \cdot y_j \qquad [4]$$

Thus, the result of the *ANN* is described by Eq. [5]:

$$Z_k = f(S_k) \quad \text{for } k = 1 \ldots M \qquad [5]$$

In the case of regression problems, $Z_k = S_k$, and in classification problems, Z_k is the result of a non-linear transformation of S_k (Fig. 3).

The *ANNs* were trained (calibration process) under supervision using a scaled conjugate gradient algorithm (Nabney, 2002). This algorithm was selected because the multi-layer perceptron (*MLP*) has a non-linear structure, so training is best conducted using a general purpose, non-linear optimization method.

As stated above, an *ANN* structure was proposed with three layers: the input layer, the hidden layer with different numbers of hidden nodes, and the output layer, which shows the results. The input layer is composed by basic climatic data such as temperature (average, maximum, and/or minimum temperature), precipitation, and/or extraterrestrial solar radiation (Fig. 5). A total of 2,000 *ANN* structures were trained with monthly data from 40 of the 44 *SIAR* weather stations for the period 2000 to 2007 (Figs. 2 and 6). The three improved structures (for estimating RH_{min}, RH_{max} and RH_{ave}) were selected after the validation process (Fig. 1 and 6) by focusing on spatial and temporal validation. During the spatial and temporal calibration process, other data (validation data) were applied on the previously calibrated *ANN* structures using the training

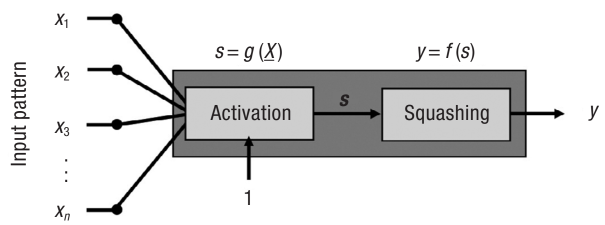

Figure 4. Scheme of a neuron.

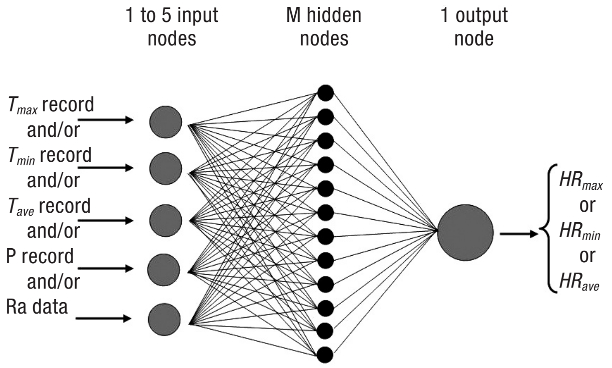

Figure 5. General artificial neural network structure used for estimating relative humidity. T_{max}: maximum temperature. T_{min}: minimum temperature. T_{ave}: average temperature. P: precipitation. Ra: extraterrestrial solar radiation. RH_{max}: maximum relative humidity. RH_{min}: minimum relative humidity. RH_{ave}: average relative humidity.

data. Four weather stations not used in the calibration process were used for the spatial validation. These stations are located in different areas of the region (Fig. 2) and were previously selected as representative of the different climatic zones (despite the homogeneity of thermal records, Tables 1 and 2, Fig. 2). Tables 1 and 2 present maximum and minimum temperature average values in weather stations in the Region of Castilla-La Mancha. Characterization of mean temperatures is shown graphically in Fig. 2. Table 1, which gives information on the variability detected among the 44 weather stations for monthly and annual averages corresponding

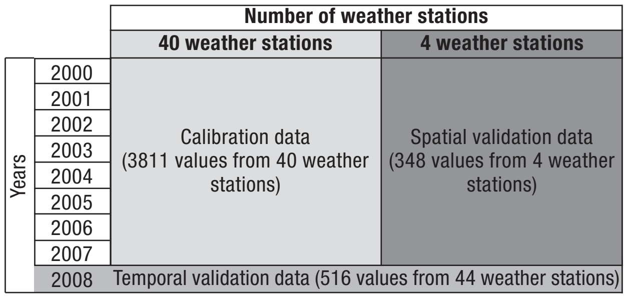

Figure 6. Scheme of data distribution for calibration and validation.

to T_{max} and T_{min} over the period 2000-2008, uses descriptive statistical analysis (Fernández-Fernández *et al.*, 2002). In Table 2 the average monthly and annual data of the four weather stations used for spatial validation can be observed. These data completes the temperature information for those selected stations useding the validation process. The correlation coefficient matrix between temperature data in the 44 weather stations are high (Martínez-Romero, 2010). Correlations between records from the stations is important when applying *ANN* techniques (Nabney, 2002). To perform temporal validation, the year 2008 was used (data from the 44 weather stations), as it was not included in the calibration process (Fig. 6).

To determine the variables (input nodes) that permit to better estimate the RH, an *ANN* structure with 10 hidden nodes and 2,500 iterations were used. Thus, different combinations of temperatures (T_{max}, T_{min}, and

Table 1. Statistical values for maximum and minimum temperatures of the 44 weather stations in Castilla-La Mancha region for average values in the series 2000-2008.

	Average		Maximum		Minimum		Range		Variance		Standard deviation	
	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}
January	13.8	3.4	10.8	–0.6	8.5	–3.7	2.3	3.2	0.9	2.2	1.0	1.5
February	15.5	4.1	12.7	0.2	9.4	–4.0	3.3	4.2	1.2	2.6	1.1	1.6
March	18.6	8.6	16.2	6.1	12.7	2.4	3.5	3.7	1.4	1.6	1.2	1.2
April	20.7	6.0	18.8	2.9	15.1	–1.2	3.7	4.2	1.4	2.3	1.2	1.5
May	25.0	7.4	23.0	4.9	19.1	0.3	3.9	4.6	1.6	2.1	1.3	1.4
June	32.6	10.9	30.4	8.5	26.4	3.8	4.1	4.6	1.9	2.4	1.4	1.6
July	34.9	16.4	33.0	13.2	29.2	8.0	3.8	5.2	1.5	3.5	1.2	1.9
August	34.0	18.2	32.2	14.8	28.5	8.7	3.8	6.1	1.4	4.5	1.2	2.1
September	29.5	18.1	27.3	14.8	23.7	8.6	3.6	6.2	1.7	4.7	1.3	2.2
October	22.8	15.2	20.8	11.9	17.9	5.7	2.9	6.2	0.9	4.2	1.0	2.0
November	16.4	6.0	14.0	2.2	11.2	–2.0	2.8	4.2	0.8	2.6	0.9	1.6
December	13.7	3.5	10.7	0.0	8.0	–4.4	2.7	4.5	0.7	2.6	0.9	1.6
Annual	22.7	10.0	20.8	6.8	17.5	2.0	3.4	4.8	1.1	2.8	1.0	1.7

T_{max}: maximum temperature (°C). T_{min}: minimum temperature (°C).

Table 2. Mean temperature values from 44 weather stations in the Region of Castilla-La Mancha (series 2000-2008). Mean temperature values from the four weather stations used for spatial validation (series 2000-2008)

Data period	Prados Redondos		La Rinconada		Alcázar de San Juan		Caudete	
	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}	T_{max}	T_{min}
January	8.5	–3.7	11.8	0.1	10.6	–0.3	13.1	1.1
February	9.4	–4.0	13.7	1.0	13.5	0.2	13.7	2.3
March	12.7	–1.2	17.9	3.4	16.6	3.0	17.4	4.9
April	15.1	0.3	20.5	5.2	19.2	4.9	19.5	6.5
May	19.1	3.8	24.7	9.1	23.4	8.8	23.5	9.6
June	26.4	8.0	32.2	14.1	31.4	13.6	30.0	13.9
July	29.2	8.8	34.4	14.9	33.5	15.2	32.7	16.5
August	28.5	8.6	33.5	14.8	32.7	14.9	32.0	16.7
September	23.7	5.7	29.1	12.1	27.9	12.4	26.9	14.2
October	17.9	3.7	21.5	8.9	20.7	8.5	22.1	10.2
November	11.2	–2.0	15.1	3.2	14.0	2.5	15.5	4.4
December	8.0	–4.4	11.3	0.5	10.6	0.3	12.3	2.1
Annual	17.5	2.0	22.1	7.3	21.2	7.0	21.6	8.5

T_{ave}), precipitation, and atmospheric solar radiation were analyzed and a set of variables that best estimates RH was selected.

After determining this «best» set of variables, an *ANN* structure optimization process was performed, with a main objective of minimizing estimation error. The optimization process consists of calibrating and validating *ANN* structures with different numbers of nodes (1 to 20) and different numbers of iterations (1 to 1,000 in a range of 10). Thus, 2,000 structures were calibrated and then validated with the above proposed methodology, resulting in 2,000 error estimates for spatial validation and 2,000 errors for temporal validation. In order to select an improved *ANN* structure, response surface plots were used to detect the areas with lower error, which coincides in both figures.

To maintain consistency among RH data (maximum, minimum and average), it is advisable to generate two of them by using *ANN* models and calculate one RH considering the other two *RHs*, using Eq. [6] as the relation between RHs.

$$RH_{ave} = \frac{RH_{max} + RH_{min}}{2} \qquad [6]$$

Goodness of fit of the *ANN* models

Statistical parameters were used to determine the goodness of fit of the *ANN* models (Willmott, 1982), such as coefficient of determination (R^2), root mean square error (*RMSE*), relative error (*RE*), and the similarity rate (*SR*).

$$RMSE = \left[n^{-1}\sum (P_i - O_i)^2\right]^{1/2} \qquad [7]$$

RMSE is the root mean square error, *n* is the number of observations, and P_i and O_i are the predicted and observed values, respectively.

$$RE = (RMSE / O_{med}) \cdot 100 \qquad [8]$$

RE is the relative error, estimated as a percentage of the average value of the variable and O_{ave} is the average value of the variable observed.

$$SR = 1 - \left[\sum (P_i - O_i)^2 / \sum \left((P_i - O_{ave}) + (O_i - O_{ave})\right)^2\right] \qquad [9]$$

SR is the similarity rate and is expressed as a relative measure of the difference between variables; if *SR* = 1, there is perfect agreement between P_i and O_i.

Results

The set of variables (thermo-pluviometric) that obtain good fit with the RH data for a structure of 10 nodes and 2,500 iterations are T_{max}, T_{min}, and *Ra*. As a first approximation, the RH_{min}, RH_{max} and RH_{ave} can be estimated with training errors of 3.88%, 4.77% and 5.02%, respectively (Table 3). However, the validation errors are perceptibly higher and can be estimated with the following *RMSE*, respectively: 1) for spatial validation 5.06%, 6.91% and 7.04%; and 2) for temporal validation 4.68%, 5.99% and 6.31%. These errors have been obtained by contrasting *ANN* model results with the average monthly values estimated using the daily RH records.

Table 3. Estimation of relative humidity by using artificial neural networks (*ANN*) models (defined structure by 10 nodes and 2,500 iterations) in Castilla-La Mancha (*C-LM*) with average maximum, and minimum monthly temperatures and extraterrestrial solar radiation data as inputs of the model

Statistical parameters	*ANN*		
	RH_{min}	RH_{max}	RH_{ave}
Number of data	3,811	3,811	3,811
Adjusted R^2	0.93	0.74	0.88
RMSE (%)	3.88	4.77	5.02
RE (%)	10.30	5.53	7.91
SR	0.98	0.92	0.97

RH_{max}: maximum relative humidity. RH_{min}: minimum relative humidity. RH_{ave}: average relative humidity. *RMSE*: root mean square error. *RE*: relative error. *SR*: similarity rate.

The response surface plots show the *RMSE* of all *ANN* structures (2,000 in this case) as a function of the number of iterations and hidden nodes. As an example, Figure 7 shows the response surface plots of the spatial (Fig. 7a) and temporal (Fig. 7b) validation of the estimation of RH_{max}, using T_{max}, T_{min}, and *Ra* as input variables. The common minimum *RMSE* in the estimation of RH_{max} is obtained with a structure of more than 3 nodes and 50-250 iterations. In the error analysis shown by the 2,000 *ANN* structures for spatial and temporal validation, there are signs of overfeeding in the network, which increases the *RMSE* when the number of nodes and iterations are simultaneously increased (Fig. 7).

The structures selected for estimating RHs are: 10 hidden nodes and 410 iterations for H_{min} (*RMSE* of 4.21% for spatial validation and 3.99% for temporal validation), 12 hidden nodes and 210 iterations in the estimation of RH_{max} (*RMSE* of 5.84% for spatial validation and 5.69% for temporal validation), 14 nodes and 160 iterations for the estimation of RH_{ave} (*RMSE* of 6.17% for spatial validation and 5.70% for temporal validation). The training errors for these three structures are 3.85%, 4.92% and 5.14% for H_{min}, H_{max} and H_{ave}, respectively, values slightly below the validation errors.

Table 4 summarize the statistical parameters analyzed after estimating RH data for spatial and temporal validation from average monthly T_{max}, T_{min}, and *Ra* data using improved *ANN* structures. To maintain the consistency between the RH data, two should be generated using *ANN* models and by calculating each RH value considering the other two, which considers the relationships among RHs. Thus, RH_{min} estimated by the improved *ANN* model shows better fit ($R^2 = 92\%$) than the fit from obtaining it using the average of RH_{max} and RH_{ave} ($R^2 = 90$ and 91% for spatial and temporal validation, respectively). In addition, the relative error is doubled when RH_{max} and RH_{ave} are used. RH_{max} estimation shows similar trends (Table 4). RH_{ave} estimation using the selected *ANN* shows smaller differences, maintaining fit (R^2 of 82-83%) and slightly increasing the relative error (*RE* of 9.79% *vs.* 10.27% for spatial validation and *RE* of 8.88% *vs.* 9.77% for temporal validation) (Table 4).

Discussion

The availability of complete RH data series as a climate variable that directly acts on living beings is very important in various disciplines, among them

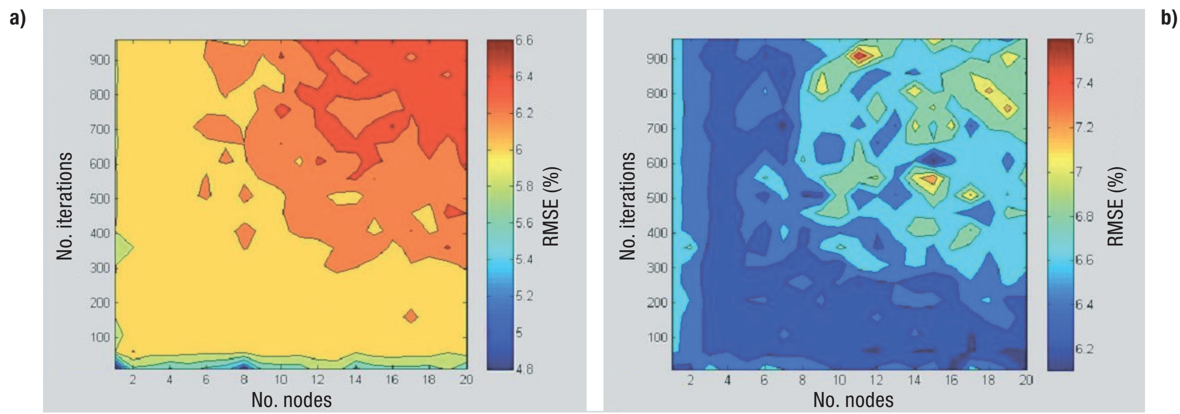

Figure 7. Response surface plots of the evolution of the root mean square error (*RMSE*) based on the Artificial Neural Network structure used for estimating maximum relative humidity in Castilla-La Mancha (*C-LM*): a) spatial validation for 4 weather stations (time series 2000-2008), b) temporal validation for the year 2008 (44 weather stations).

Table 4. Estimated monthly average relative humidity in Castilla-La Mancha (*C-LM*) from optimized artificial neural networks (*ANN*) using averages of maximum and minimum temperatures and extraterrestrial solar radiation. Error of spatial an temporal validation

Statistical parameters[a]	Error of validation	Optimized *ANN* (10 nodes, 410 iterations)	Optimized *ANN* (12 nodes, 210 iterations)	Optimized *ANN* (14 nodes, 160 iterations)	Considering the relation: $RH_{ave} = (RH_{min} + RH_{max}) / 2$		
		RH_{min}	RH_{max}	RH_{ave}	RH_{min}[b]	RH_{max}[b]	RH_{ave}[b]
Number of data	Spatial	348	348	348	348	348	348
	Temporal	516	516	516	516	516	516
Average x	Spatial	37.24	85.80	63.03	38.01	88.14	64.88
	Temporal	38.55	86.39	64.16	38.12	86.17	64.16
Average y	Spatial	38.01	88.14	64.88	40.25	88.81	61.52
	Temporal	38.12	86.17	64.16	41.93	89.78	62.47
Average x /Average y	Spatial	1.02	1.03	1.03	1.06	1.01	0.95
	Temporal	0.99	1.00	1.00	1.10	1.04	0.97
R^2	Spatial	0.92	0.60	0.83	0.90	0.62	0.82
	Temporal	0.92	0.60	0.83	0.91	0.56	0.83
Adjusted R^2	Spatial	0.92	0.60	0.83	0.90	0.61	0.82
	Temporal	0.92	0.60	0.83	0.91	0.56	0.83
RMSE	Spatial	4.21	5.84	6.17	8.88	10.16	6.66
	Temporal	4.21	5.84	6.17	8.88	10.16	6.66
RE	Spatial	11.30	6.80	9.79	23.36	11.53	10.27
	Temporal	10.34	6.59	8.88	22.60	11.37	9.77
SR	Spatial	0.98	0.84	0.95	0.93	0.75	0.92
	Temporal	0.98	0.87	0.95	0.93	0.79	0.93

[a] *x*: observed values. *y*: predicted values. R^2: determination coefficient. *RMSE*: root mean square error. *RE*: relative error. *SR*: similarity rate. RH_{min}: minimum relative humidity. RH_{max}: maximum relative humidity. RH_{ave}: average relative humidity. [b] Each relative humidity value is estimated from the other two, generated from *ANN* techniques using the ratio that determines average relative humidity as the average of minimum and maximum humidity.

agricultural applications (De Juan and Martín de Santa Olalla, 1993; Capel, 2000; Matsushita *et al*., 2004). It is considered one of the most important factors in the development of pests and diseases (Huber and Gillespie, 1992; Laurence *et al.,* 2002), and directly influences evaporation by affecting stomata. Therefore, it has an important influence on final crop yield (Villalobos *et al*., 2002). RH values are necessary for applying models with a strong physiological basis for estimating ET_0 (Allen *et al.,* 1998), including models for generating the test reference year (TRY) from measured meteorological variables (De Miguel and Bilbao, 2005).

Weather generators have been used extensively in agricultural and hydrological applications where high spatial resolution and/or long sequential series of records are required to solve common problems (Wilby and Wilks, 1999), but estimating this climate parameter when there are no records has not been studied. Missing value estimation is extremely difficult for locations with high spatial variance (Cano *et al*., 2004). Standard techniques for this problem use regression or interpolation models associated with neighboring stations with complete records of RH (Eischeid *et al*., 1995; Allen *et al*., 1998; Xia *et al*., 1999). There are climate data generators which are more complex, based on Bayesian Networks or other stochastic models, neural networks, clustering methods, canonical correlation analysis, among others. These models also require values for the variable to be estimated from nearby stations (Wilby and Wilks, 1999; Basak *et al.,* 2004; Cano *et al*., 2004; Hruschka *et al*., 2007).

This innovative research permits to estimate RH values in this area to develop *ANNs* that estimate mean monthly RH values from temperature data in a specific location. Authors like Randhir *et al.* (2004) used *ANNs* to estimate surface specific humidity using microwave brightness temperature observation where the global error (*RMSE*) differences were 1.1 g kg^{-1}.

The results are of application in diverse, climate-based models of dynamic simulation in plant growth (Acock, 1991), with climate as a direct influence on gross photosynthetic rate at a given time (Pereira and Machado, 1986; Dourado-Neto *et al*., 1998). A direct application of monthly RH data can be performed by the program ClimGen (Stöckle *et al*., 1999), an extension of the crop growth simulator Cropsyst. This program generates daily values from a series of monthly values for their use in simulations.

The use of an *ANN* structure similar to that used in other studies to estimate climatic parameters (Kifli and Öztürk, 2007; Dai *et al*., 2008; Kumar *et al*., 2008) does not guarantee good fit, as the models can be too simple or too complex to yield good results, or generate overfeeding problems. The number of hidden layers and the number of nodes in each hidden layer are usually determined by a trial-and-error procedure (Ritchie *et al.*, 2003; Xu and Chen, 2008), and there are no rules for determining the exact number of layers or hidden nodes (Dawson and Wilby, 2001; Xiong and O'Connor, 2002). Therefore, the response surface plots are useful tools in this sense.

The results obtained in the present study are a good starting point in research on generating *RH* values when data is limited, which can be used in databases (metadata from weather stations) and can be made available for a variety of uses (Prohom and Herrero, 2008).

The use of *ANNs* is interesting in generating climate parameters from available climate data. For determining optimal *ANN* structure in estimating *RH* values, model calibration with some of the available data is necessary, and validation must be performed on the results with climatic data from stations not used in the calibration process, considering spatial and temporal variability.

It is recommended estimating RH_{min} and RH_{max} using the improved *ANN* model and estimating RH_{ave} as the average of RH_{min} and RH_{max}. This maintains coherence among the three RHs.

In the Region of *C-LM*, the estimated environmental *RH* data from basic temperature and *Ra* data were estimated with relative errors of: 1) spatial error of 11.30%, 6.80% and 10.27% for RH_{min}, RH_{max} and RH_{ave}; 2) temporal errors of 10.34%, 6.59% and 9.77% for RH_{min}, RH_{max} and RH_{ave}.

The methodology developed in this paper reduces the *RE* to aproximately 10% to estimate the *RHs* and it could be used for different purposes in future research. Thus, this methodology could be implemented in areas with lack of climatic data and could improve the process of crop water requirements forecast, as well as the management of water resources. On the other hand, in other applications, *ANN* models could be optimized for different frequencies in the climate records (*e.g.* hourly, daily, etc.), which could be interesting in agricultural or industry sector.

Acknowledgements

We would like to thank to the CICYT for funding the Spanish national project (AGL2007-66716-C03-03), and the Regional Education Government of *C-LM* for funding the project PCI08-0117, in which this work is included.

References

ACOCK B., 1991. Potential for using long-term field research data to develop and validate crop simulators. Agron J 83, 56-61.

ALEXANDERSSON H., 1984. A homogeneity test based on ratios and applied to precipitation series. Report 79, Dept Meteorology, Uppsala Univ, Uppsala. 55 pp.

ALEXANDERSSON H., 1986. A homogeneity test applied to precipitation data. J Climatol 6, 661-675.

ALLEN R.G., PEREIRA L.S., RAES D., SMITH M., 1998. Crop evapotranspiration. guidelines for computing crop water requirements. FAO Irrig and Drain paper No. 56. Rome, Italy. pp. 57-64.

ALLISON P.D., 2001. Missing data. Series: quantitative applications in the social sciences. Sage Publ Inc, Univ Pennsylvania, Philadelphia, PA, USA.

BASAK J., SUDARSHAN A., TRIVEDI D., SANTHANAM M.S., 2004. Weather data mining using independent component analysis. J Mach Learn Res 5, 239-253.

CANO R., SORDO C., GUTIÉRREZ J.M., 2004. Applications of bayesian networks in meteorology. In: Advances in bayesian networks (Gámez J.A, Moral S., Salmerón A., eds). Springer, London, UK. pp. 309-327.

CAPEL J.J., 2000. El clima en la Península Ibérica. Ariel SA, Barcelona, Spain. 281 pp. [In Spanish].

DAI X., SHI H., LI Y., OUYANG Z., HUO Z., 2008. Artificial neural network models for estimating regional reference evapotranspiration based on climate factors. Hydrol Process 23, 442-450.

DAWSON D.W., WILBY R., 2001. Hydrological modelling using artificial neural networks. Prog Phys Geogr 25, 80-108.

DE JUAN J.A., MARTÍN DE SANTA OLALLA F.J., 1993. La evapotranspiración. En: Agronomía del riego (Martín de Santa Olalla F.J., De Juan, J.A., eds). Mundi-Prensa, Madrid, España. pp. 239-298. [In Spanish].

DE LA CASA A., OVANDO G., RODRÍGUEZ A., 2003. Estimación de la radiación solar global en la provincia de Córdoba, Argentina, y su empleo en un modelo de rendimiento potencial de papa. Revista de Investigaciones Agropecuarias (RIA) 32, 45-62. [In Spanish].

DE LEÓN A., ARRIBA A., DE LA PLAZA M.C., 1988. Caracterización agroclimática de la provincia de Albacete. Ministerio de Agricultura, Pesca y Alimentación, Secretaría Técnica General, Madrid. 168 pp. [In Spanish].

DE MIGUEL A., BILBAO J., 2005. Test reference year generation from meteorological and simulated solar radiation data. Sol Energy 78, 695-703.

DOURADO-NETO D., TERUEL D.A., REICHARDT K., NIELSEN D.R., FRIZZONE J.A., BACCHI O.O.S., 1998. Principles of crop modelling and simulation: II. The implications of the objective in model development. Sci Agric 55, 51-57.

EISCHEID J.K., BAKER C.B., KARL T.R., DÍAZ H.F., 1995. The quality control of long-term climatological data using objective data analysis. J Appl Meteorol 34, 2787-2795.

ELÍAS F., RUIZ-BELTRÁN L., 1981. Estudio agroclimático de la región de Castilla-La Mancha. Departamento de Agricultura de la Junta de Comunidades de Castilla-La Mancha, Toledo, Spain. 230 pp. [In Spanish].

FERNÁNDEZ-FERNÁNDEZ S., CORDERO J.M., CÓRDOBA-LARGO A., 2002. Estadística descriptiva, 2ª ed. Escuela Superior de Gestión Comercial y Marketing (ESIC), Univ Rey Juan Carlos, Madrid. 562 pp. [In Spanish].

FOUNT I., 2000. Climatología de España y Portugal, 2ª ed. Univ de Salamanca, Salamanca, Spain. 170 pp. [In Spanish].

HRUSCHKA E.R. Jr., HRUSCHKA E.R., EBECKEN N., 2007. Applying bayesian networks for meteorological data mining. In: Applications and innovations in intelligent systems XII (Macintosh A., Ellis R., Allen T., eds). Springer, London, UK. pp 122-133.

HUBER L., GILLESPIE T.J., 1992. Modeling leaf wetness in relation to plant disease epidemiology. Annu Rev Phytopathol 30, 553-577.

KIŞI Ö., ÖZTÜRK Ö., 2007. Adaptive neurofuzzy computing technique for evapotranspiration estimation. J Irrig Drainage Eng-ASCE 133, 368-379.

KUMAR M., BANDYOPADHYAY A., RAGHUWANSHI N.S., SINGH R., 2008. Comarative study of conventional and artificial neural netwoirk-based ET_0 estimation models. Irrig Sci 26, 531-545.

LAURENCE H., FABRY F., DUTILLEUL P., BOURGEOIS G., ZAWADZKI I., 2002. Estimation of the spatial pattern of surface relative humidity using ground based radar measurements and its application to disease risk assessment. Agric For Meteorol 111, 223-231.

MARTÍNEZ-ROMERO A., 2010. Parámetros agroclimáticos y su distribución espacial en Castilla-La Mancha. Doctoral thesis. Escuela Tecnica Superior de Ingenieros Agrónomos de Albacete, Albacete, Spain. [In Spanish].

MATSUSHITA B., XUB M., CHENC J., KAMEYAMAD S., TAMURA M., 2004. Estimation of regional net primary productivity (NPP) using a process-based ecosystem model: how important is the accuracy of climate data? Ecol Model 178, 371-388.

NABNEY I.T., 2002. Netlab. Algorithms for patter recognition, 4th ed. Springer Verlag, London, UK. 420 pp.

PEREIRA A.R., MACHADO E.C., 1986. Um simulador dinâmico do cresimento de uma cultura de cana-de-açúcar. Bragantia 45, 107-122. [In Portuguese].

POPOVA Z., KERCHEVA M., PEREIRA L.S., 2006. Validation of the FAO methodology for computing ET_0 with limited data. Application to South Bulgaria. Irrig and Drain 55, 201-215.

PROHOM M., HERRERO M., 2008. Hacia la creación de una base de datos climática de Cataluña (siglos XVIII a XXI). Tethys 5, 3-12. [In Spanish].

RANDHIR S., VASUDEVAN B.G., PAL P.K., JOSHI P.C., (2004). Artificial neural network approach for estimation of surface specific humidity and air temperature using multifrequency scanning microwave radiometer. Proc Indian Acad Sci (Earth Planet Sci) 113, 89-101.

RITCHIE M.D., WHITE B.C., PARKER J.S., HAHN L.W., MOORE J.H., 2003. Optimization of neural network architecture programming improves detection and modelling interactions in studies of human diseases. BMC Bioinformatics 7, 4-28.

SCHALKOFF R.J., 1997. Artificial neural networks. McGraw Hill, London, UK. 419 pp.

SHANK D., 2003. Drew point temperature prediction using artificial neural networks. Doctoral thesis. Artificial Intelligence Center, and the Department of Biological and Agricultural Engineering, Driftmier Engineering Center, University of Georgia, Athens, GA, USA.

SINGH R., VASUDEVAN B.G., PAL P.K., JOSHI P.C., 2004. Artificial neural network approach for estimation of surface specific humidity and air temperature using multifrequency scanning microwave radiometer. Proc Indian Acad Sci (Earth Planet Sci) 113, 89-101.

STÖCKLE C.O., CAMPBELL G.S., NELSON R., 1999. ClimGen manual. Biol Syst Eng Dept, Washington St Univ, Pullman, WA, 28 pp.

TRAJKOVIC S., TODOROVIC B., STANKOVIC M., 2003. Forecasting of reference evapotrans-piration by artificial neural networks. J Irrig Drain Eng-ASCE 129, 454-457.

VILLALOBOS F.J., MATEOS L., ORGAZ F., FERERES E., 2002. Fitotecnia. Bases y tecnología de la producción agrícola. Mundi-Prensa, Madrid. 496 pp. [In Spanish].

WILBY R.L., WILKS D.S., 1999. The weather generation game. A review of stochastic weather models. Progr Phys Geogr 23, 329-357.

WILLMOTT C.J., 1982. Some comments on the evaluation of model performance. Bulletin American Meteorological Society 63, 1309-1313.

XIA Y., FABIAN P., STOHL A., WINTERHALTER M., 1999. Forest climatology: estimation of missing values for Bavaria, Germany. Agric For Meteorol 96, 131-144.

XIONG L.H., O'CONNOR K.M., 2002. Comparison of four updating models for real-time river flow forecasting. Hydrol Sci J 47, 621-639.

XU S., CHEN L., 2008. A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining. Proc V Intnl Conference on Information Technology and Applications. Cairns, QLD, Australia, Jun 23-26. pp. 683-686.

ZANETTI S.S., SOUSA E.F., OLIVEIRA V.P.S., ALMEIDA F.T., BERNARDO S., 2007. Estimating evapotranspiration using artificial neural network and minimum climatological data. J Irrig Drain Eng-ASCE 33, 83-89.