

Neuro-fuzzy systems to estimate reference evapotranspiration

Mousaab Zakhrouf¹, Hamid Bouchelkia¹ and Madani Stamboul²

¹URMER Laboratory, Faculty of Technology, Department of Hydraulic, University of Tlemcen, Tlemcen, Algeria

²Faculty of Engineering Science, Department of Civil Engineering, Laghouat University, Laghouat, Algeria

ABSTRACT

Routine and rapid estimation of evapotranspiration (ET) at regional scale is of great significance for agricultural, hydrological and climatic studies. A large number of empirical or semi-empirical equations have been developed for assessing ET from meteorological data. The FAO-56 PM is one of the most important methods used to estimate evapotranspiration. The advantage of FAO-56 PM is a physically based method that requires a large number of climatic parameter data. In this paper, the potential of two types of neuro-fuzzy system, including ANFIS based on subtractive clustering (S_ANFIS), ANFIS based on the fuzzy C-means clustering method (F_ANFIS), and multiple linear regression (MLR), were used in modelling daily evapotranspiration (ET₀). For this purpose various daily climate data – air temperature (T), relative humidity (RH), wind speed (U) and insolation duration (ID) – from Dar El Beidain Algiers, Algeria, were used as inputs for the ANFIS and MLR models to estimate the ET₀ obtained by FAO-56 based on the Penman-Monteith equation. The obtained results show that the performances of S_ANFIS model yield superior to those of F_ANFIS and MLR models. It can be judged from results of the Nash-Sutcliffe efficiency coefficient (EC) where S_ANFIS (EC = 94.01%) model can improve the performances of F_ANFIS (EC = 93.00%) and MLR (EC = 92.12%) during the test period, respectively.

Keywords: modelling, FAO-56 PM evapotranspiration, S-ANFIS, F-ANFIS, MLR, semi-arid regions, Algeria

INTRODUCTION

Evapotranspiration is an important component of the hydrological cycle and influenced by many meteorological parameters. It is a combination of water loss from the soil surface by evaporation and from crops by transpiration (ET₀) (Aytek, 2008). Models for predicting ET₀ range from deterministically based combined energy balance–vapour transfer approach to empirical relationships based on climatological variables, or to evaporation from a standard evaporation pan. A large number of empirical or semi-empirical equations have been developed for assessing ET₀ from meteorological data. The Penman–Monteith (PM) equation for estimating reference evaporation was used in this study. The PM is a physically based method; comparative studies strongly suggest that the PM formula is preferred to other empirical models (Jensen et al., 1990). But the use of PM requires a large number of climatic parameters, such as daily maximum temperature and minimum temperature, solar radiation, relative humidity and wind speed. However, records for such weather variables are often incomplete or not always available for many locations, which is a weakness of the PM model (Allen et al., 1998). Neuro-fuzzy provides an efficient way of handling the uncertainty for complex systems without sufficient data or with only vague information (Ross, 1995; Cox, 1999), and clustering processes can be a very effective technique to identify natural groupings in data from a large dataset, thereby allowing concise representation of relationships embedded in the data. Therefore, two neuro-fuzzy systems, (i) fuzzy c-means clustering (FCM) based fuzzy inference system, named F-ANFIS, and (ii) subtractive clustering based fuzzy inference system, named S-ANFIS, are used in mapping the nonlinear relationship between the daily reference evapotranspiration

and meteorological data and also for investigating the effectiveness of the parameters. These NF use daily climatic data (air temperature, relative humidity, wind speed and the insolation duration) as inputs, and ET₀ values estimated by the Penman-Monteith formula as outputs.

FAO-56: Penman Monteith method

The Penman–Monteith (PM) model was recommended as the sole standard method of calculating ET₀ by the United Nations Food and Agriculture Organization (FAO) and the World Meteorological Organization (WMO), whenever the sufficient necessary data are available. The PM equation for estimating reference evaporation used in this study is based on most recent FAO-56 PM model described in the FAO's Irrigation and Drainage Paper No. 56 (Allen et al., 1998).

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{\gamma + 273} U_2 (e_a - e_d)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where ET₀ = reference evapotranspiration (mm-day⁻¹); Δ = slope of the saturation vapour pressure function (kPa°C⁻¹); R_n = net radiation (MJ-m⁻²-day⁻¹); G = soil heat flux density (MJ m⁻²-day⁻¹); γ = psychrometric constant (kPa°C⁻¹); T = mean air temperature (°C); U₂ = average 24 h wind speed at 2 m height (m-s⁻¹), e_a is the saturation vapour pressure (kPa), e_d is the actual vapour pressure (kPa). Trajkovic and Kolakovic (2009) examined the potential of the FAO-56 PM equation in estimating the ET₀ under humid conditions from limited weather data. They concluded that the minimum and maximum of air temperature and local default wind speed value were the minimum data requirements in order to use the FAO-56 PM equation successfully under humid conditions.

Adaptive network-based fuzzy inference system (ANFIS)

Neuro-fuzzy systems are fuzzy systems formed by a learning algorithm inspired by the theory of neural networks. The

* To whom all correspondence should be addressed.
e-mail: bouchelkiahamid@gmail.com

Received 16 December 2017, accepted in revised form 8 March 2019.

learning technique operates according to local information and a unique production of local changes in the original fuzzy system. So the interest is to establish a predictive system based on the integration of neural networks and fuzzy inference systems (FIS). Fuzzy inference systems is based on the linguistic rules of the type (IF-THEN) to translate knowledge on the dynamic system (Dastorani et al., 2010) (Fig. 1). ANFIS is a fuzzy inference (FIS) 'Sugeno' type supervised learning 'Takagi', implemented in the framework of adaptive neural networks; the first introduction of this system was by Jang (Jang, 1993; Takagi and Sugeno, 1985).

In the literature a wide variety of rule generation techniques can be found. Clustering is one of the most important methods used to generate fuzzy rules. The purpose of these algorithms is to organize data in homogeneous groups, to account for the membership of an element in several classes.

A wide variety of rule generation techniques are found in the literature by the fuzzy clustering method, and among the most used are subtractive clustering, and fuzzy C-means clustering.

Subtractive clustering

The idea is to find regions in the feature space with high densities of data points. The point with the highest number of neighbours is selected as centre of the cluster. The data points within a prespecified fuzzy radius are then removed, and the algorithm looks for a new point with the highest number of neighbours (Cobaner, 2011). Consider a collection of n data points in an m -dimensional space. Without loss of generality, the data points are assumed to have been normalized within

a hypercube. Since each data point is a candidate for cluster centres, a density measure at data point x_i is defined as:

$$D_i = \sum_{j=1}^n e^{-\|x_i - x_j\|^2 / (r_a/2)^2} \quad (2)$$

where cluster radius r_a is a positive constant. Thus, a data point that has many neighbouring data points will have a high potential of being a cluster centre. The radius r_a defines a neighbourhood. Data points outside this radius have little effect on the density measure.

After the density measure of each data point has been calculated, the first cluster centre is chosen to be the data point with the highest density measure. Suppose x_{c1} is the point selected and D_{c1} is its density measure. Then, the density measure for each data point x_i is revised by the formula:

$$D_i = D_i - D_{c1} e^{-\frac{\|x_i - x_{c1}\|^2}{(r_b/2)^2}} \quad (3)$$

where r_b is a positive constant. Note that the data points near the first cluster centre x_{c1} will have significantly reduced density measures, so that they are unlikely to be selected as the next cluster centre.

The constant r_b defines a neighbourhood to be reduced in density measure. It is normally larger than r_a to prevent closely spaced cluster centres; typically, $r_b = 1.5 \cdot r_a$. After the density measure for each point is revised, the next cluster centre x_{c2}

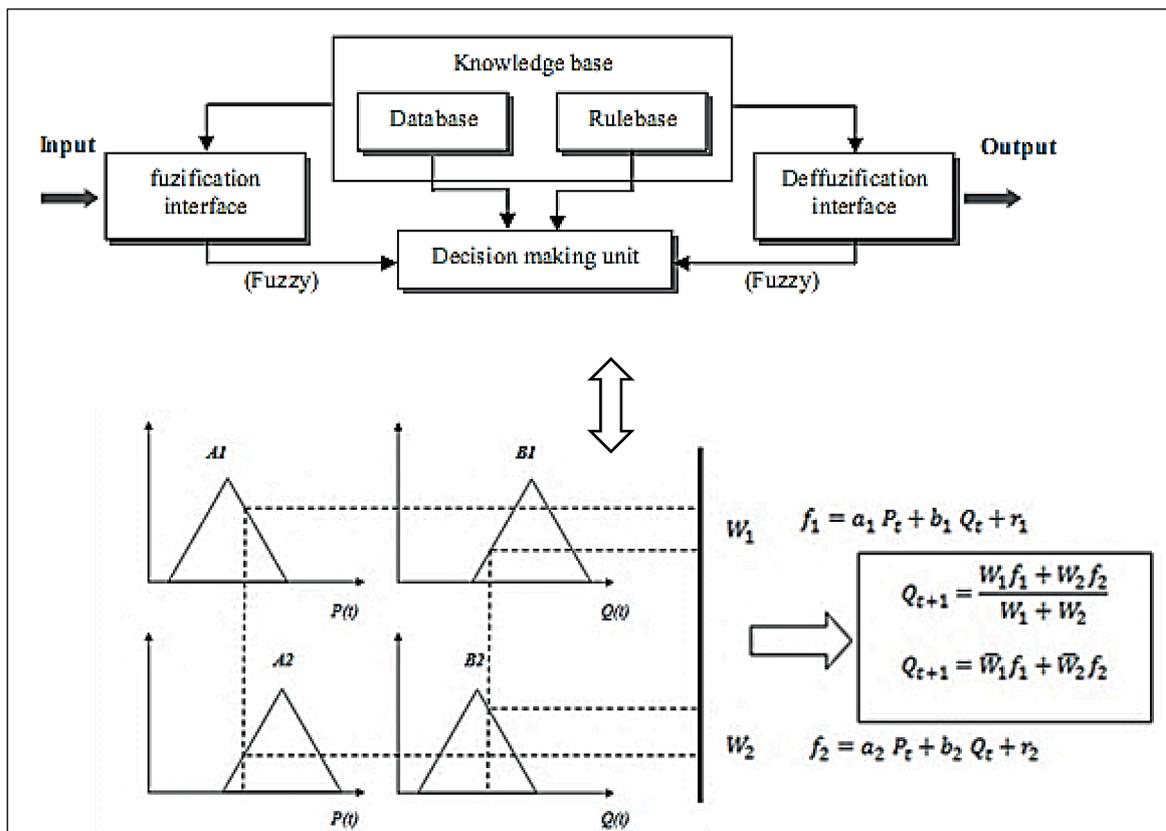


Figure 1
Mechanism of inference fuzzy system of type 'Sugeno'

is selected and all the density measures are revised again. The process is repeated until a sufficient number of cluster centres are generated.

Fuzzy C-means clustering

Bezdek (1981) introduced the fuzzy C-means clustering method in 1981, with the development of the fuzzy theory. This algorithm is used for analysis based on distance between various input data points. The clusters are formed according to the distance between data points and the cluster centres are formed for each cluster. In fact, FCM is a data-clustering technique (Chen and Zhang, 1998; Hui, 2008) in which a dataset is grouped into n clusters with every data point in the dataset related to every cluster; it will have a high degree of belonging (connection) to that cluster and another data point that lies far away from the centre of a cluster will have a low degree of belonging to that cluster. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2 \quad (4)$$

where m is any real number greater than 1; it was set to 2 by Bezdek (Bezdek, 1981), u_{ij} is the degree of membership of x_i in the cluster j ; x_i is the i^{th} of d -dimensional measured data; c_j is the d dimension centre of the cluster and $\|\cdot\|$ is any norm expressing the similarity between any measured data and the centre. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the c_j cluster centres by:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (5)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (6)$$

The operation principle of adaptive neuro-fuzzy inference system (ANFIS) is presented in (Fig. 2). It has 6 functional blocks (input layer, based on rules database, a unit of decision-interface fuzzification and defuzzification interface) which are generated using 6 layers of neurons:

- **Layer 1:** input layer.
- **Layer 2:** consists of a number of nodes whose activation functions are membership functions (MFs)
- **Layer 3:** the output layer provides the minimal value of its inputs
- **Layer 4:** Layer 4, normalized with respect to the other input, where the output node i is equal to the input i divided by the sum of the inputs
- **Layer 5:** the output of the node i , is a linear function of the output and the signal input of the controller
- **Layer 6:** is the summation of the signal input

Performance criteria

The statistical parameters used in this work are: mean absolute relative error (MARE), mean square error (MSE), mean

error (ME) and the Nash-Sutcliffe efficiency coefficient (EC) (Abraham, 2004). These parameters are given by the following relationships:

$$MARE = \frac{\sum_{i=1}^N \left| \frac{Et0_i - \widehat{ET}0_i}{Et0_i} \right|}{N} \cdot 100C \quad (7)$$

$$MSE = \frac{\sum_{i=1}^N (Et0_i - \widehat{ET}0_i)^2}{N} \quad (8)$$

$$ME = \frac{\sum_{i=1}^N (Et0_i - \widehat{ET}0_i)}{N} \quad (9)$$

$$EC = \left(1 - \frac{\sum_{i=1}^N (Et0_i - \widehat{ET}0_i)^2}{\sum_{i=1}^N (Et0_i - \overline{ET}0)^2} \right) \cdot 100 \quad (10)$$

where $Et0_i$ is the measured evapotranspiration rate value, $\widehat{ET}0_i$ is the evapotranspiration rate calculated by the model, $\overline{ET}0$ is the average evapotranspiration measured, $\widehat{ET}0$ is the average evapotranspiration simulated, and N is the number of data.

CASE STUDY

In the present study, the aforementioned fuzzy inference systems are applied to the daily climatic data of Dar El Beida automated weather station (36°43' N; 03°15' E) in the north-centre of (Algiers) Algerian County (Fig. 3). Dar El Beida weather station is located 24 m amsl. The climate in Algiers is classified as a Mediterranean Sea climate.

In this part of the study, different input combinations, including various daily climatic data, that is, (air temperature, relative humidity, wind speed and the insolation duration) were used in NF models to estimate reference evapotranspiration. The available dataset consisting of 1 673 days was separated into 3 parts, the training set containing 1 095 days, validation set containing the third set of 365 days of the dataset and the testing set containing 213 days of the dataset. The daily statistical parameters of climatic data and FAO-56 PM ET_0 values are shown in (Table1).

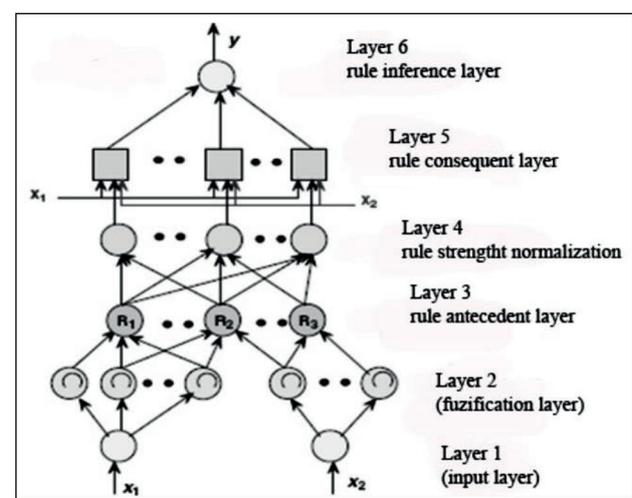


Figure 2
Operation principle of the ANFIS network based on fuzzy inference system of type 'Sugeno'

	Training				Validation				Testing			
	Min	Max	Mean	Std	Min	Max	Mean	Std	Min	Max	Mean	Std
T	4.3	35.7	18.31	6.12	3.5	35.6	17.67	6.47	0.5	31.8	16.38	6.93
RH	0.3	0.9	0.73	0.10	0.4	1	0.74	0.09	0.5	0.9	0.73	0.09
U	0	9.2	2.69	1.61	0	8.2	2.29	1.56	0	9.8	2.65	1.65
ID	0	14	7.71	3.74	0	13.2	7.47	3.49	0	13.6	7.69	4.11
ET_0	0.3	10.2	3.53	2.11	0.5	8.1	3.22	1.95	0.4	7.9	3.62	2.31

Reference evapotranspiration estimation using adaptive NF techniques

To realize the approximation of evapotranspiration using a neuro-fuzzy network, two types of neuro-fuzzy system including ANFIS based on subtractive clustering (S_ANFIS) and ANFIS based on the fuzzy C-means clustering method (F_ANFIS) were used.

In order to investigate the effect of daily climatic data on ET_0 , the final architecture of the ANFIS models was determined via an extensive trial-and-error process. The number of clusters of the F-ANFIS model was found based on this approach. In order to generate fuzzy rules using S-ANFIS, it is critical to determine the proper cluster radius (radii).

The cluster radius indicates the range of influence of a cluster when you assume the data space to be a unit hypercube (range [0, 1]). Generally, specifying a small cluster radius will result in many small clusters in the data (with many rules), and specifying a large cluster radius will result in a few large clusters in the data, (with fewer rules). Good values for radii are usually between 0.2 and 0.5 (Samhouri, 2009). In this paper, the optimum number of cluster and radii values were determined according to the MARE criterion in the testing phase.

RESULTS AND DISCUSSION

The effect of the value of radii changing on the quality of results of S_ANFIS is shown in Fig. 4. From Fig. 4 it can be deduced that the optimal radius value of 0.45 gives us the best results. The effect of number of cluster changes on the quality of the results from F_ANFIS is shown in Fig. 5; from Fig. 5 it was observed that the F_ANFIS model having 7 clusters estimated the minimum value of MARE.

Results obtained for the S_ANFIS, F_ANFIS and MLR approaches proposed in this paper are presented in Table 2.

The three models are evaluated based on the indices described. The NASH criterion (EC) provides an overall assessment of the quality of estimation. Models with NASH values close to 0.9 are generally of good quality, while models with NASH values close to 1 are deemed to produce near-perfect estimation. The NASH values for the S_ANFIS and F_ANFIS models are all over 90%, which indicates that both types of models achieved a good quality result. The overall quality of estimation of the S_ANFIS model is better than the F_ANFIS and multiple linear regression (MLR) models, which indicates that the NASH values of the S-ANFIS model are higher than those of the F_ANFIS and MLR models.

The magnitude of systematic overestimation or underestimation of a model is evaluated using the ME and MARE indices. The results indicate that the three models generally underestimate evapotranspiration quantiles. Based

on the MSE index, the S_ANFIS and F_ANFIS models are the least biased models, while the MLR is the most biased model. The MARE index provides the measure of bias in the absolute relative scale. Based on the MARE index, the S_ANFIS model has slightly higher biases than the F_ANFIS and MLR models.

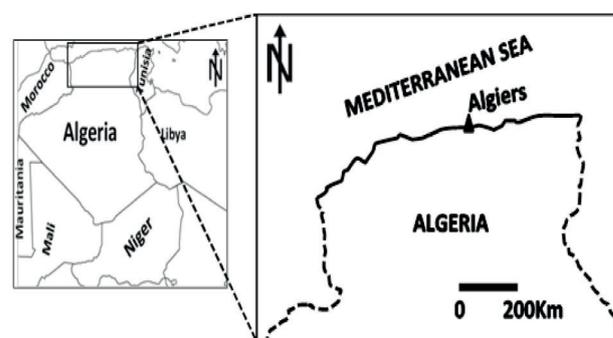


Figure 3
Location of Dar El Beida automated weather station

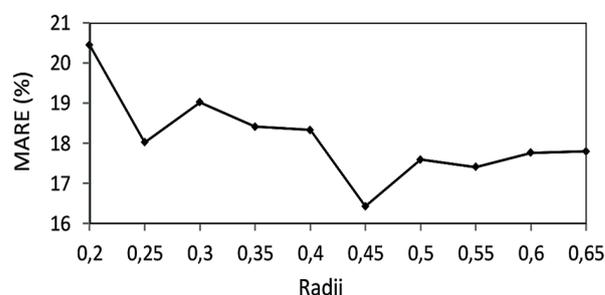


Figure 4
Optimal value of radii in the testing phase of S_ANFIS

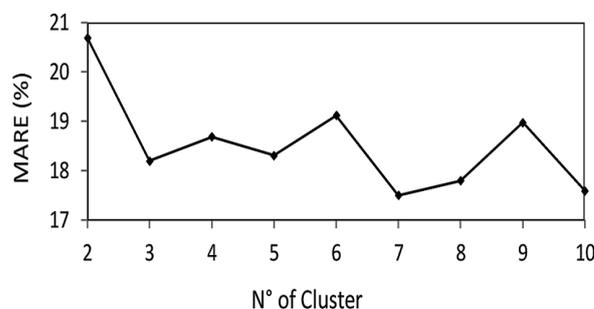


Figure 5
Optimal number of clusters in the testing phase of F_ANFIS

Figure 6 shows the scatter plot between the observed and modelled flows by S_ANFIS, F_ANFIS and MLR models. It was observed that the flow forecast by the ANFIS model was close to the 45° line. From Rindex, it is worth mentioning that the performance of S_ANFIS was better than that of F_ANFIS and MLR models.

The distribution of relative error, computed by S_ANFIS, F_ANFIS, and MLR models during the testing phase is presented in Figs 7(b), 7(c), and 7(d), in which it can be observed that the estimation of ET_0 was better using S_ANFIS than the F_ANFIS and MLR models.

		S_ANFIS	F_ANFIS	MLR
Training	MARE %	16.27	16.98	29.54
	MSE mm ²	0.233	0.240	0.512
	ME mm	-2×10^{-9}	-5×10^{-9}	-0.0306
	EC %	94.73	94.58	88.43
Validation	MARE %	21.52	21.96	34.97
	MSE mm ²	0.415	0.417	0.626
	ME mm	0.12	0.11	0.01
	EC %	89.01	88.98	83.44
Testing	MARE %	16.43	17.50	29.30
	MSE mm ²	0.316	0.370	0.416
	ME mm	0.30	0.36	0.39
	EC %	94.01	93.00	92.12

	S_ANFIS	F_ANFIS	MLR
ET_0 sim (mm)	706.52	693.25	688.37
ET_0 obs (mm)	770.80	770.80	770.80
ARE (%)	8.34	10.06	10.69

The estimation of total FAO-56 PM ET_0 obtained from the estimated ET_0 values was also considered for comparison due to its importance in irrigation systems (in determining the irrigation channel and/or tube, and total required irrigation water demand and its storage capacity, etc.). The total estimated ET_0 amounts in the test period are given in Table 3. The S-ANFIS, F-ANFIS and MLR models whose input parameters are T , RH, U and ID, estimate the total FAO-56 PM ET_0 value of 770.80 mm as 706.52 mm and 693.25 mm, with an underestimation of 8.34%, 10.06% and 10.69%, respectively.

CONCLUSION

The estimation of evapotranspiration using climatic variables was predicted by the adaptive NF techniques based on subtractive clustering and fuzzy c-means clustering-based fuzzy inference systems. The developed models were validated using the FAO-56 PM ET_0 values of Dar El Beida, in Algiers, Algeria. The validations of the developed models with 4 input parameters, including daily climatic data, namely, temperature, relative humidity, wind speed and the insolation duration.

The results obtained in this study showed the effectiveness of subtractive clustering algorithm based fuzzy inference systems to estimate the FAO-56 PM ET_0 . The use of this hybrid method is an alternative fully justified for good water management. These encouraging results introduce a number of perspectives; it would be interesting to try hybrid models by coupling wavelet transform with neuro-fuzzy systems, and simultaneously optimizing by genetic algorithm: membership functions, scaling factors, and conclusions of fuzzy rules.

REFERENCES

- ABRAHART RJ, KNEALE PE and SEE LM (2004) *Neural Networks for Hydrological Modelling*. Taylor & Francis, Hoboken, NJ. 304 pp.
- ALLEN RG, PEREIRA LS, RAES D and SMITH M (1998) Crop evapotranspiration guidelines for computing crop water requirements. Paper No. 56. FAO Irrigation and Drainage, Food and Agriculture Organization of the United Nations, Rome.
- AYTEK A, GUVEN A, ISHAK YUCEM M and AKSOY H (2008) An explicit neural network formulation for evapotranspiration. *Hydrol. Sci. J.* 53 893–904. <https://doi.org/10.1623/hysj.53.4.893>

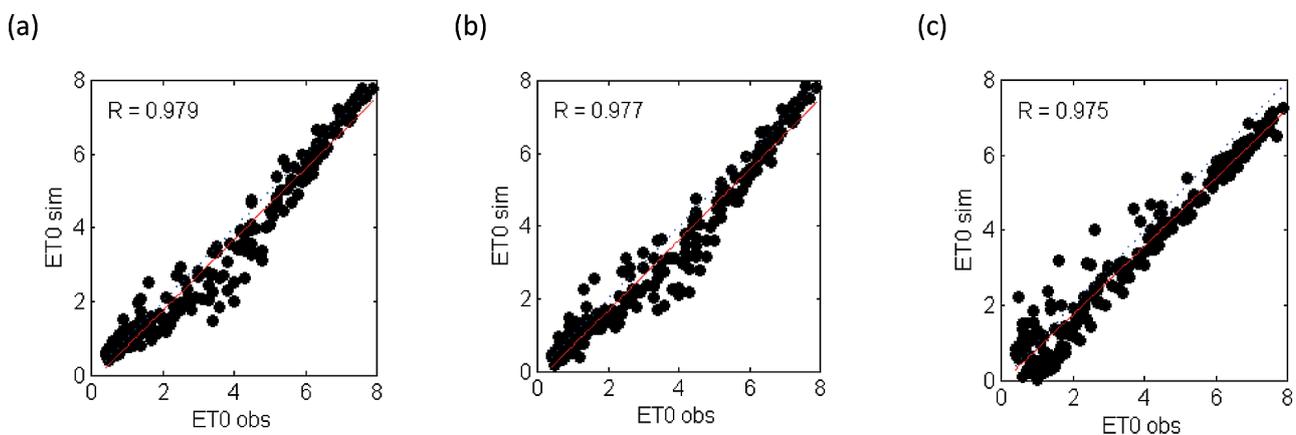


Figure 6
Scatter plots comparing observed and forecast ET_0 using (a) S-ANFIS, (b) F-ANFIS models in the test period

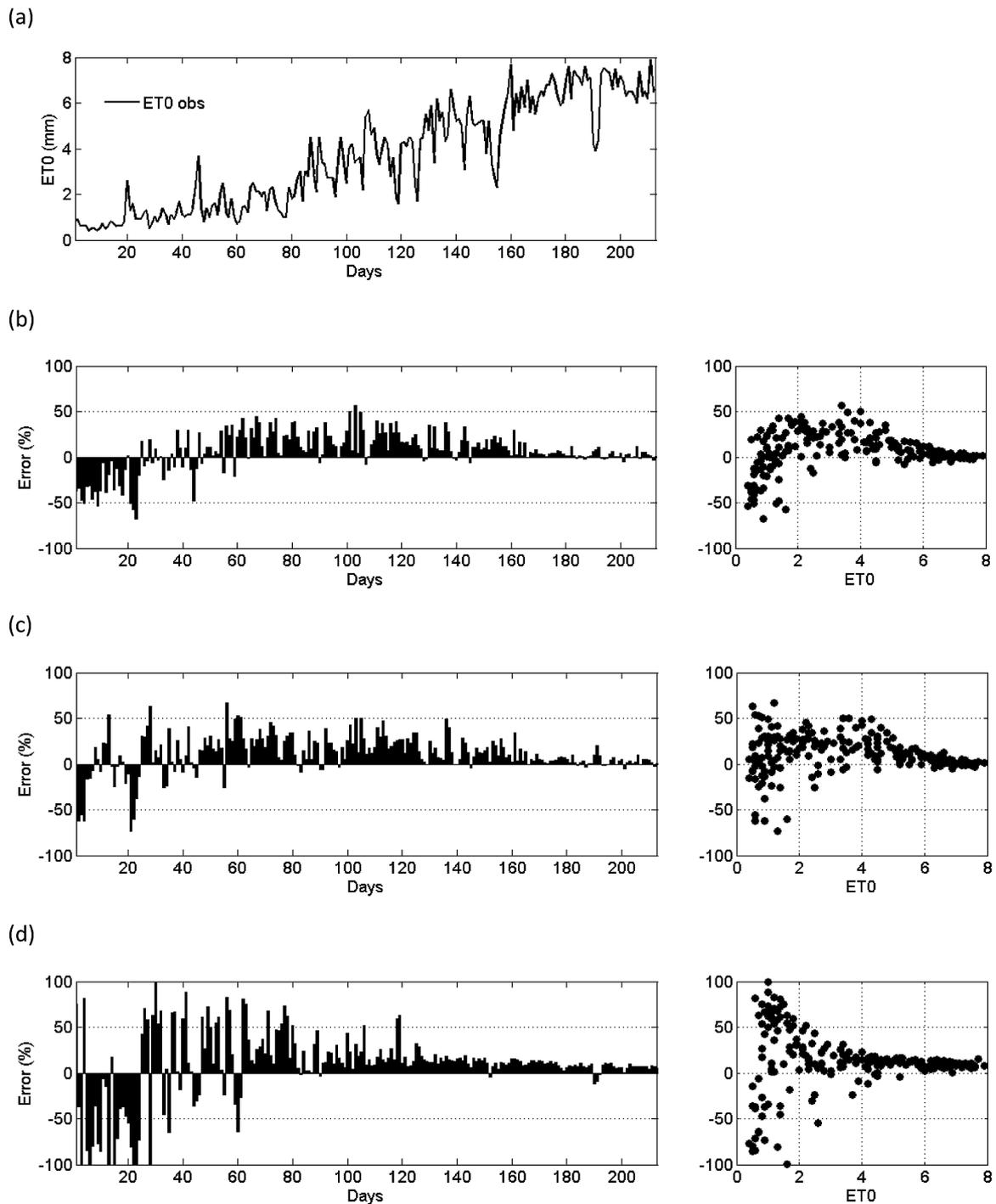


Figure 7

Distribution of relative error plots for (b) S_ANFIS model (c) F_ANFIS model and (d) MLR model during the testing period. (a) observed flow

BEZDEK JC (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York. <https://doi.org/10.1007/978-1-4757-0450-1>

CHEN S and ZHANG D (1998) Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybern.* **34** 1907–1916. <https://doi.org/10.1109/TSMCB.2004.831165>

COBANER M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. *J. Hyd.* **398** (2011) 292–302. <https://doi.org/10.1016/j.jhydrol.2010.12.030>

COX E (1999) *The Fuzzy Systems Handbook* (2nd edn). Academic Press, California.

DASTORANI M, MOGHADAMNIA A, PIRI J and RICO-RAMIREZ M (2010) Application of ANN and ANFIS models for reconstructing missing flow data. *Environ. Monit. Assess.* **166** 421–434. <https://doi.org/10.1007/s10661-009-1012-8>

JANG JS (1993) ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man Cybern.* **23** (1993) 665–685. <https://doi.org/10.1109/21.256541>

JENSEN ME, BURMAN RD and ALLEN RG (1990)

- Evapotranspiration and irrigation water requirements. *ASCE Manuals and Reports on Engineering Practices no. 70*. ASCE, New York.
- HUI L (2008) Method of image segmentation on high-resolution image and classification for land covers. *Fourth International Conference on Natural Computation*, vol. 5. 563–566.
- ROSS TJ (1995) *Fuzzy Logic with Engineering Applications*. McGraw Hill Co, New York.
- SAMHOURI M, ABU-GHOUSH M, YASEEN E and HERALD T (2009) Fuzzy clustering-based modeling of surface interactions and emulsions of selected whey protein concentrate combined to κ -carrageenan and gum arabic solutions. *J. Food Eng.* **91** 10–17. <https://doi.org/10.1016/j.jfoodeng.2008.07.029>
- TAKAGI T and SUGENO M (1985) Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **15** 116–132. <https://doi.org/10.1109/TSMC.1985.6313399>
-