IOP Conference Series: Earth and Environmental Science

PAPER • OPEN ACCESS

Reference evapotranspiration estimation using adaptive neuro-fuzzy inference system with limited meteorological data

To cite this article: M Y Chia *et al* 2020 *IOP Conf. Ser.: Earth Environ. Sci.* **612** 012017

Reference evapotranspiration estimation using adaptive neuro-fuzzy inference system with limited meteorological data

M Y Chia[1], Y F Huang[1,2] and C H Koo[1]

[1]Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Kajang, Selangor, Malaysia.

[2]Email: huangyf@utar.edu.my

Abstract. Machine learning tools are extremely useful for the estimation and modelling of hydrological processes such as evapotranspiration (ET). In this study, reference evapotranspiration (ET_0) in Labuan located in the East Malaysia was estimated using an artificial neuro-fuzzy inference system (ANFIS). In order to investigate the feasibility of the ANFIS model for a wide temporal range, daily meteorological data collected at Station 96465 (Labuan) from year 2014 to 2018 were divided on an annual basis. ANFIS models were trained using data from different years as well as varying combinations of one climatic parameter with solar radiation. The study revealed that the ANFIS model was capable of performing accurate estimation when only one year of training data were used where errors of less than 5 % and NSE above 0.950 were achieved. This finding could be useful for new meteorological stations where data are limited. Furthermore, solar radiation and minimum temperature were deemed to be the best input combination because of their distinguishable characteristics. Maximum temperature which highly overlaps solar radiation in nature was found the worst complementary input. However, it is important to note that the importance of climatic parameters could be affected by extreme weather conditions.

1. Introduction

Evapotranspiration (ET) is considered as one of the most important component to sustain the water balance in the hydrological cycle [1]. Evapotranspiration data are useful in many fields such as environmental science, irrigation and water resources management [2]. Lysimeter is the most direct way to measure evapotranspiration. However, its application had been restricted due to the high operational cost as well as narrow geographical representation [3]. Therefore, numerous empirical models and equations were developed and improvised over the years as an approach to fill in the gap left by the disadvantages of the lysimetric measurement [4-6]. To date, the Penman-Monteith (PM) model stands supreme and is regarded as the standard for the calculation of reference evapotranspiration (ET_0) and is well recognised by the Food and Agriculture Organisation of the United Nations [7]. Nevertheless, challenges such as the need of at least six meteorological parameters have to be overcome when using the PM model to estimate ET_0. Hence, the focus of current research started to shift to a new direction where artificial intelligent based models are sought as replacements for the PM model and other empirical models.

Machine learning tools are being regarded as one of the most promising solution to estimate ET_0 as proven by many available literatures [8]. To simplify, machine learning utilises certain algorithms to learn the relationship between inputs and outputs for a given training data set. Deduced relationship by selected algorithms will be used to compute ET_0 for the inputs provided in the future. Application of machine learning models to estimate or predict ET_0 has been studied extensively by researchers worldwide. One of the most commonly used machine learning models is the artificial neural network (ANN) [9-12]. However the black box nature of ANN's operation lacks in explanatory capability for researchers to understand the ET process [13].

Besides the ANN model, some researchers argued that ET_0 can be modelled in a more linguistic way using fuzzy logic algorithms so that it is easier to be interpreted by experts [14]. However, the construction and formation of fuzzy rules are tedious for high dimensional problems. Hence, researchers tend to integrate ANN-based computation into the fuzzy model – adaptive neuro-fuzzy inference system (ANFIS) so that optimised fuzzy rules can be determined using ANN. The application of ANFIS model to predict ET_0 is well reported in the literatures. Cobaner [15] and Kisi and Zounemat-Kermani [16] compared two ways of generating fuzzy rules of ANFIS model, namely the grid partition and subtractive clustering methods, to estimate ET_0. Both works proved that the two methods of generating fuzzy rules could yield estimation with similar accuracies.

It is well agreed that in order to develop a powerful machine learning model, the quality and quantity of training data play important roles. However, in most cases, data can be insufficient, or in some extreme cases, certain data could not be collected due to a host of reasons. Therefore, the aim of this study is to overcome these barriers and restrictions to formulate an efficient ET_0 prediction application. The specific objectives are: (1) to study the robustness of the ANFIS model in temporal context, where meteorological data of different years were used as training data and (2) to investigate the effect of different input combinations on the performance of ANFIS model.

2. Methods

2.1. Materials and study area

Meteorological data was collected for the Station 96465 (Labuan, Malaysia), which is located west of Sabah state in East Malaysia (5°18' N, 115°15' E). The location of the station is shown in Figure 1. Labuan is the targeted area of interest in this study due to the presence of Bukit Kuda water dam on the island. Terrestrial water storage is strongly affected by ET and therefore precise estimation of ET_0 in this region shall be given attention so that the decision makers can draw appropriate policies based on the predictions.

Daily meteorological data from 1st January 2014 to 31st December 2018 was provided by the Malaysian Meteorological Department (MMD). These data included maximum temperature (T_{max}, °C), minimum temperature (T_{min}, °C), daily mean temperature (T_{mean}, °C), relative humidity (RH, %), wind speed at 2 m elevation (u_2, m/s) and solar radiation (R_s, MJ/m^2). The details of the data are presented in Table 1.

Table 1. Statistical analysis of meteorological data obtained from MMD.

	Overall		**2014**		**2015**		**2016**		**2017**		**2018**	
	μ[a]	**σ**[b]	**μ**	**σ**	**μ**	**σ**	**μ**	**σ**	**μ**	**σ**	**μ**	**σ**
T_{max} (°C)	31.39	1.32	31.33	1.28	31.55	1.30	31.86	1.34	31.12	1.23	31.13	1.30
T_{min} (°C)	25.18	1.06	25.23	1.00	25.31	1.08	25.58	1.02	25.02	0.94	24.79	1.07
T_{mean} (°C)	27.92	0.97	27.91	0.97	28.10	0.97	28.23	0.88	27.74	0.90	27.63	0.98
RH (%)	81.77	4.50	80.50	3.67	79.37	4.38	81.10	4.17	83.12	4.21	84.88	3.37
u_2 (m/s)	1.57	0.82	1.31	0.60	1.64	0.85	1.84	1.08	1.48	0.66	1.57	0.75
R_s (MJ/m^2)	18.71	4.23	17.95	4.30	18.33	4.05	19.25	3.89	18.58	4.56	19.46	4.13

[a] mean of all data
[b] standard deviation of data

2.2. Penman-Monteith model

In order to train the ANFIS model, ET_0 values calculated using PM model is used as the training target. The overall equation of PM model is provided in (1):

$$\mathrm{ET_0} = \frac{0.408\Delta(R_n - G) + \gamma(\frac{900}{T+273})u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where ET_0 is daily reference evapotranspiration (mm/day), R_n is net radiation ($MJm^{-2}day^{-1}$), G is soil heat flux ($MJm^{-2}day^{-1}$), T is daily mean temperature (°C), u_2 is wind speed at 2 m height (m/s), e_s is mean saturation vapour pressure (kPa), e_a is actual vapour pressure (kPa), Δ is slope of vapour pressure curve (kPa/°C) and γ is psychrometric constant [7].

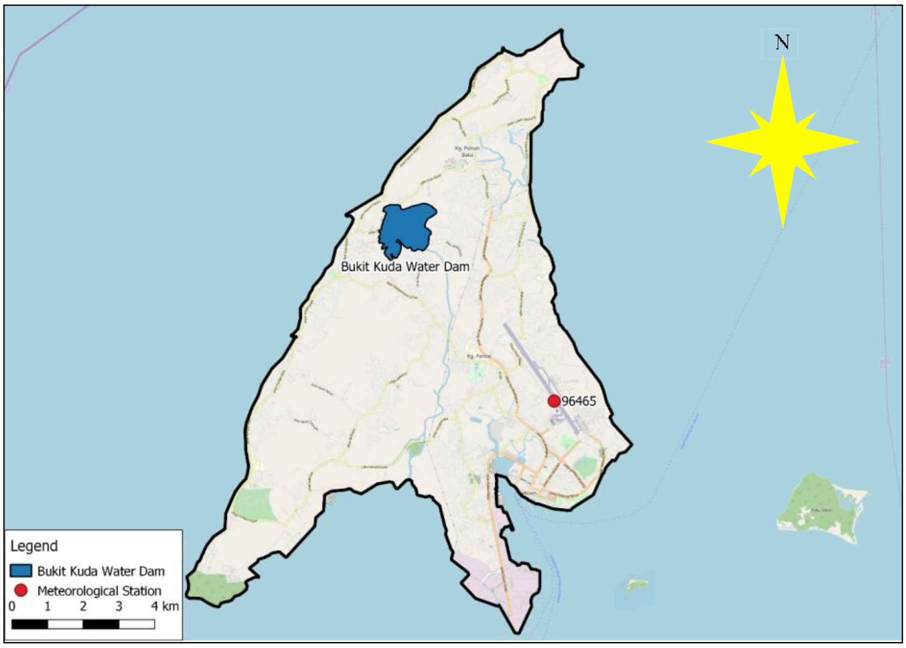

Figure 1. Targeted Study Area – the Labuan Island.

2.3. ANFIS model

The model used to estimate ET_0 in this study is an ANFIS model. Generally, an ANFIS model consists of fixed nodes and adaptive nodes which are responsible for the computation of weights based on membership functions and application of fuzzy rules, respectively. In this study, fuzzy rules generated are based on the Sugeno type fuzzy rule, which can be expressed as the following:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$ (2)

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$ (3)

On top of that, subtractive clustering method is integrated into the ANFIS model as an approach to optimise the model. Subtractive clustering method is favoured over grid partition method due to the ability of the former to treat data points as clusters in order to reduce the overall complexity of the problem. Prior to the training of the ANFIS models, the training data were normalised using min-max normalisation as shown in (4):

$$x_{norm} = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \quad (4)$$

where x_{norm} is the normalised data, x_0 is the raw data, x_{max} is the maximum value of raw data and x_{min} is the minimum value of raw data.

In this study, the training strategy was divided into two parts. In the first part, different combinations of input climatic parameters will be used to train the model. In the second part, data from each combination are to be divided on a yearly basis to train the ANFIS model. For example, data from year 2014 will be used as training data while data from 2015 to 2018 will be used for testing and verification purpose. Selected combinations of input climatic parameters are shown in Table 2. Several studies had shown that radiation is the main driver and best predictor of ET_0 in warm regions [13, 17, 18]. Hence, in this study, the authors would like to compare the performance of different combinations of R_s with another climatic parameter in estimating ET_0. This assumption is logical and reasonable as the East Malaysia is also located in proximity to the Equator Line, where the climate is warm as well. A full set of climatic parameters (C1) was also tested as a controlled experiment.

Table 2. Different combinations of input climatic parameters.

Combinations	Climatic Parameters
C1	R_s, T_{max}, T_{min}, T_{mean}, RH, u_2
C2	R_s, T_{max}
C3	R_s, T_{min}
C4	R_s, T_{mean}
C5	R_s, RH
C6	R_s, u_2

2.4. Performance evaluation

In order to assess the performance of the ANFIS models, several performance indicators were used in this study. The Mean absolute error (MAE) is used to determine the deviation of the models' estimations from the actual value. The Root mean square error (RMSE) is used to detect if there are any extreme errors occurred during the computation of the train models. The Nash-Sutcliffe efficiency (NSE) is used to evaluate the stability of the models' estimations [19]. The equations of MAE, RMSE and NSE are provided in (5), (6) and (7), respectively:

$$\text{MAE} = \frac{1}{N}\sum_{i}^{N} | y_{actual,i} - y_{predicted,i} | \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N} (y_{actual,i} - y_{predicted,i})^2} \tag{6}$$

$$\text{NSE} = 1 - \frac{\sum_{i}^{N} (y_{actual,i} - y_{predicted,i})^2}{\sum_{i}^{N} (y_{actial,i} - y_{mean})^2} \tag{7}$$

where N is the number of data points, y_{actual} is the actual value and $y_{predicted}$ is the value predicted by the ANFIS models.

3. Results and discussion

3.1. Suitability of different training periods

The results of this study are shown in Table 3. The performance of the ANFIS models are sorted according to different training periods. From the results, it can be seen that the ANFIS model is capable of delivering good predictions even in the case of limited training data. For example, when the training period was set to be the data from year 2014, the ANFIS model with subtractive clustering optimisation method was able to produce estimations with MAE ranging from 0.035 mm/day to 0.115 mm/day. Taking the mean of ET_0 from year 2015 to 2018 as 4.018 mm/day, the percentage of error was only 0.871 % to 2.862 %. Although there is no universal threshold to indicate the acceptability of ANFIS estimations, however, errors less than 5 % are considered as accurate, coupled with a minimum NSE of 0.950. These could be considered as rather accurate predictions. Besides, the NSE, which represented the stability as well as the reliability of ANFIS model, ranged from 0.959 to 0.993 which also indicated that ET_0 at Station 96465 (Labuan) can be well modelled by an ANFIS. In fact, this provided a strong basis to suggest that a one-year collection of daily data (minimum 365 data points) was sufficient for developing a good ANFIS model for long term ET_0 prediction. This phenomenon was also observed when the annual dataset from year 2015, 2016, 2017 and 2018 were individually used as training data.

Table 3. Performance of ANFIS models using different training periods and input combinations.

Combinations	**Performance**		
	MAE (% error)	RMSE (% error)	NSE
Training Period: 2014			
C1	0.035 (0.871)	0.072 (1.792)	0.993
C2	0.115 (2.862)	0.168 (4.181)	0.959
C3	0.093 (2.315)	0.139 (3.460)	0.972
C4	0.108 (2.688)	0.160 (3.982)	0.963
C5	0.107 (2.663)	0.154 (3.833)	0.971
C6	0.114 (2.837)	0.157 (3.908)	0.964
Training Period: 2015			
C1	0.021 (0.528)	0.033 (0.829)	0.999
C2	0.156 (3.921)	0.193 (4.851)	0.962
C3	0.128 (3.217)	0.161 (4.047)	0.976
C4	0.148 (3.720)	0.184 (4.624)	0.967
C5	0.099 (2.488)	0.136 (3.418)	0.973
C6	0.126 (3.167)	0.172 (4.323)	0.968
Training Period: 2016			
C1	0.022 (0.558)	0.033 (0.838)	0.998
C2	0.137 (3.478)	0.173 (4.392)	0.964
C3	0.107 (2.716)	0.136 (3.453)	0.976
C4	0.135 (3.427)	0.168 (4.264)	0.967
C5	0.124 (3.148)	0.157 (3.986)	0.974
C6	0.120 (3.046)	0.161 (4.087)	0.968
Training Period: 2017			
C1	0.031 (0.776)	0.049 (1.226)	0.996
C2	0.112 (2.803)	0.173 (4.329)	0.959
C3	0.095 (2.378)	0.139 (3.479)	0.973
C4	0.103 (2.578)	0.161 (4.029)	0.964
C5	0.108 (2.703)	0.149 (3.729)	0.968
C6	0.121 (3.028)	0.168 (4.205)	0.960
Training Period: 2018			
C1	0.046 (1.160)	0.070 (1.732)	0.993
C2	0.114 (2.875)	0.186 (4.691)	0.959
C3	0.096 (2.421)	0.147 (3.707)	0.974
C4	0.111 (2.799)	0.174 (4.388)	0.962
C5	0.123 (3.102)	0.161 (4.060)	0.965
C6	0.136 (3.430)	0.187 (4.716)	0.967

3.2. Effect of different input combinations

From Table 3, it can be seen that generally, C1 gave better accuracy and stability than C2 to C6, which could be well explained by the different number of input climatic parameters. However, among C2 to C6, their performance differed due to the inclusion of different climatic parameters. Except for ANFIS model trained with data from year 2015, other ANFIS models suggested that in the case of very limited parameters (only two in this context), C3 is the most suitable combinations of parameters to be used. C3 had R_s and T_{min}, and this could explain the major drivers of ET in Station 96465 (Labuan) which had warm and humid climate. R_s which was contributed by the sunshine duration was responsible for the ET in the day time, whereas T_{min} was usually achieved in the night time. Hence, for estimation of ET_0 in the study area, the collection of data required in C3 would be recommended. The discrepancy that occurred in year 2015 could be interpreted as a form of different interactions between each climatic parameter. The lowest RH was registered in that particular year, which resulted in the shift of importance from T_{min} to RH and C5 was considered as the best input combination.

On the other hand, the majority of the ANFIS models with different training periods (year 2014, 2015 and 2016) suggested that C2 was the worst among the input combinations. This could be due to the nature of T_{max}, which was likely to be recorded in the day time, had high overlapping nature with R_s. Therefore, inputting the two climatic parameters concurrently could not well explain the ET which could also take place in the night time. Nonetheless, the ANFIS models trained with data from year 2017 and 2018 suggested otherwise. According to Table 1, 2017 and 2018 had the lowest mean T_{max}, which means the overlapping effect of it with R_s was reducing. The ANFIS model could easily distinguish the trend T_{max} and R_s which led to the improvement of models trained with C2. As such, the accuracy of ANFIS models trained with C2 using data from 2017 and 2018 had relatively better performance as compared to C5 and C6.

T_{mean}, RH and u_2 did not have significant contribution to the accuracy of ET_0 estimation with the ANFIS model. Firstly, T_{mean} does not have outstanding characteristics that are well suited to explain ET phenomenon in tropical climate region where the weather condition is hot and humid. In particular, the usefulness of T_{mean} was strongly over-shadowed by T_{max} and T_{min}. On the other hand, the high humidity at Station 96465 (Labuan) limited its contribution to ET of that area and therefore the effect of RH parameter was not prominent. However, in the case of low annual average RH, such as in year 2015, the importance of RH could then be observed. As for u_2, due to the low wind speed (1.31 m/s to 1.84 m/s) of the warm study area, its effect was marginal [20, 21].

4. Conclusions

The ANFIS model which use fuzzy rules that are easily interpretable by experts were investigated in this study. The scope of this study was focused on the determination of ability of ANFIS model to estimate daily ET_0 at Station 96465 (Labuan) under circumstances of limited data, in terms of data points and meteorological data. ANFIS models were trained with one-year data (annual basis from year 2014 to year 2018) of two climatic parameters: a combination of R_s and either T_{max}, T_{min}, T_{mean}, RH or u_2. The results of the study showed that the ANFIS model was able to provide accurate estimation with only one year training data, thus reducing the difficulty of performing long term prediction in places where meteorological stations are newly set up. Besides, the study also suggested that T_{min} could be the decisive parameter to be combined with R_s due to the clear distinction (responsible for night time and day time ET, respectively) between the two parameters. This argument was well supported by the relatively poorer performance when combination of T_{max} and R_s was used, where high similarity (both day time) was exhibited. However, the importance of climatic parameter could be shifted or affected when extreme cases occur such as high RH and low T_{max}. Overall, this study helped to cement the belief that the ANFIS model is appropriately applicable for ET_0 estimation with limited meteorological data. Interpretation of experts can be translated into effective policies for the decision makers in order to utilise scarce water resources effectively.

Acknowledgments

This research was funded by Universiti Tunku Abdul Rahman (UTAR), Malaysia through Universiti Tunku Abdul Rahman Research Fund under project number IPSR/RMC/UTARRF/2018-C2/K03.

References

[1] Xiang K, Li Y, Horton R and Feng H 2020 Similarity and difference of potential evapotranspiration and reference crop evapotranspiration – A review *Agr. Water Manage.*

232 106043

[2] Saggi M K and Jain S 2020 Application of fuzzy-genetic and regularization random forest (FG-RRF): Estimation of crop evapotranspiration (ET) for maize and wheat crops *Agr. Water Manage.* **229** 105907

[3] Stanhill G 2005 *Evapotranspiration* (Amsterdam: Elsevier) pp 502-6

[4] Thornthwaite C W 1948 An Approach toward a Rational Classification of Climate *Geogr. Rev.* **38**(1) 55

[5] Hargreaves G H and Samani Z A 1985 Reference Crop Evapotranspiration from Temperature *Appl. Eng. Agric.* **1**(2) 96-9

[6] Priestley C H B and Taylor R J 1972 On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters *Mon. Weather Rev.* **100**(2) 81-92

[7] Allan R G, Pereira L, Raes D and Smith M 1998 *Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and Drainage Paper* (Rome: Food and Agriculture Organization of the United Nations) p 561998

[8] Chia M Y, Huang Y F, Koo C H and Fung K F 2020 Recent Advances in Evapotranspiration Estimation Using Artificial Intelligence Approaches with a Focus on Hybridization Techniques—A Review *Agronomy* **10**(1) 101

[9] Ladlani I, Houichi L, Djemili L, Heddam S and Belouz K 2012 Modeling daily reference evapotranspiration (ET0) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study *Meteorol. Atmos. Phys.* **118**(3-4) 163-78

[10] Traore S, Luo Y and Fipps G 2016 Deployment of artificial neural network for short-term forecasting of evapotranspiration using public weather forecast restricted messages *Agr. Water Manage.* **163** 363-79

[11] Abdullah S S, Malek M A, Abdullah N S, Kisi O and Yap K S 2015 Extreme Learning Machines: A new approach for prediction of reference evapotranspiration *J. Hydrol.* **527** 184-95

[12] Traore S, Wang Y M and Kerh T 2010 Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone *Agr. Water Manage.* **97**(5) 707-14

[13] Pinos J, Chacón G and Feyen J 2020 Comparative analysis of reference evapotranspiration models with application to the wet Andean páramo ecosystem in southern Ecuador *Meteorologica* **45**(1) 25-45

[14] Kisi O 2013 Applicability of Mamdani and Sugeno fuzzy genetic approaches for modeling reference evapotranspiration *J. Hydrol.* **504** 160-70

[15] Cobaner M 2011 Evapotranspiration estimation by two different neuro-fuzzy inference systems *J. Hydrol.* **398**(3-4) 292-302

[16] Kisi O and Zounemat-Kermani M 2014 Comparison of Two Different Adaptive Neuro-Fuzzy Inference Systems in Modelling Daily Reference Evapotranspiration *Water Resour. Manage.* **28**(9) 2655-75

[17] Yang Y, Chen R, Song Y, Han C, Liu J and Liu Z 2019 Sensitivity of potential evapotranspiration to meteorological factors and their elevational gradients in the Qilian Mountains, northwestern China *J. Hydrol.* **568** 147-59

[18] Gao Z, He J, Dong K and Li X 2017 Trends in reference evapotranspiration and their causative factors in the West Liao River basin, China *Agr. Forest Meteorol.* **232** 106-17

[19] Nash J E and Sutcliffe J V 1970 River flow forecasting through conceptual models part I — A discussion of principles *J. Hydrol.* **10**(3) 282-90

[20] Valipour M 2014 Analysis of potential evapotranspiration using limited weather data *Appl. Water Sci.* **7**(1) 187-97

[21] Huang G, Wu L, Ma X, Zhang W, Fan J and Yu X 2019 Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions *J. Hydrol.* **574** 1029-41