Kohonen Self-Organizing Map as a software sensor estimator of reference crop evapotranspiration

ADEBAYO ADELOYE[1] **& RABEE RUSTUM**[2]

1 *School of the Built Environment, Heriot-Watt University, Edinburgh, UK*
a.j.adeloye@sbe.hw.ac.uk

2 *College of Engineering, University of Dammam, Saudi Arabia*

Abstract Reference crop evapotranspiration (ETo) estimation is of importance in irrigation water management for the calculation of crop water requirements and its scheduling, rainfall–runoff modelling and numerous other water resources studies. This paper developed the Kohonen Self-Organizing Map (KSOM), unsupervised artificial neural networks software sensors to predict the ETo. This was achieved by using the powerful clustering capability of the KSOM to analyse the multi-dimensional data array of estimated Penman-Monteith ETo and different subsets of input climatic variables. Data obtained at two locations, Edinburgh (UK) and Udaipur (India), were used in order to demonstrate the versatility of the approach. The findings indicate that the KSOM-based ETo estimates were in good agreement with those obtained using the conventional FAO Penman-Monteith formulation at both locations. A further comparison of the KSOM estimates with other commonly used empirical methods for the ETo, e.g. Thornthwaite, Priestley-Taylor and Hargreaves, showed that the former were far superior. This offers significant potential for accurate estimation of the ETo in regions of the world where the needed climatic data are unavailable for the implementation of the full Penman-Monteith formulation.

Key words reference crop evapotranspiration; crop water requirements; Kohonen Self-Organizing Map; neural networks; FAO Penman-Monteith method

INTRODUCTION

Knowledge of the reference crop evapotranspiration is very important in various fields of water resources, such as estimation of crop water requirements, scheduling of irrigation water application, rainfall–runoff modelling, land suitability evaluation, and general catchment water balance studies. Because of its significance, several direct methods based on field measurement using different kinds of lysimeters have been employed, but they are time-consuming and require precision in the experimental set-up in order to achieve a good result. Consequently, significant research activities are being carried out to develop reliable and accurate methods that use observed weather data as inputs for the estimation of the reference crop evapotranspiration, ETo (George *et al.*, 2002; Itenfisu *et al.*, 2003).

Of the numerous estimation methods available for the reference crop evapotranspiration, the Penman-Monteith (PM) model has been recommended by the FAO based on extensive comparative studies in different climatic conditions. However, there are practical difficulties with applying the PM model in practice, notably the complexity of the associated expressions and more importantly the huge number of climatic data required as input, most of which are not routinely measured in several countries. A way of avoiding the complexity and difficulties associated with the PM-ETo estimation is to use data-driven artificial intelligence (AI) modelling techniques, such as artificial neural networks (ANNs), which are able to map any non-linear relationship among variables, no matter how complex, without the need to specify explicitly the mathematical form of the model. Zanetti *et al*. (2007) present an extensive review of feed-forward back-propagation ANN modelling of evapotranspiration. In most of these studies, minimum climatological data were used as inputs to produce relatively accurate estimates of the PM-ETo. However, a problem with the feed-forward back-propagation ANN is that it can be affected by missing values and outliers. Indeed, where such noise is available in the data, feed-forward ANNs have been known to give unrealistic results (Rustum *et al*., 2007). However, unsupervised ANNs, typified by the Kohonen Self Organizing Map (KSOM), perform a clustering of a large dimensional array into a smaller one of essential features (or best matching units, BMUs), thus making any inherent correlations between the vectors in the array much more visible. These features then become the basis of

predictions, implying that missing values, outliers or other types of noise are no hindrance when using the KSOM for prediction.

Despite this versatility of the KSOM, however, it has rarely seen any widespread application in hydrological modelling, especially in the modelling of the PM-ETo, although recent applications in water quality and wastewater treatment plant modelling have been reported (Rustum & Adeloye, 2007; Rustum *et al.*, 2007, 2008). Therefore, the aim of this study is to use the KSOM to develop and test intelligent models to estimate the PM-ETo using easily observed weather data as inputs. Once the model learns to predict the PM-ETo using the training data set, the model will be stored and used for prediction using new input vectors. We will also test the KSOM models against established approximate ETo models such as Thornthwaite, Blaney-Criddle, Hargreaves, etc. All of this will use data observed for two climatic conditions: a small experimental catchment in the temperate, sub-humid Edinburgh (UK) and a lake catchment in the arid Udaipur region of India, in order to test the versatility of the methodology.

KOHONEN SELF-ORGANISING MAP (KSOM)

Basics of the Kohonen Self-Organizing Map

The KSOM (also called feature map or Kohonen map) is one of the most widely used artificial neural networks algorithms (Kohonen *et al.*, 1996) and its principal goal is to transform an incoming signal pattern of arbitrary dimension into a two-dimensional discrete map. It involves clustering the input patterns in such a way that similar patterns are represented by the same output neurons, or by one of its neighbours (Back *et al.*, 1998).

The KSOM consists of two layers: the multi-dimensional input layer and the competitive or output layer; both of these layers are fully interconnected as illustrated in Fig. 1(a). The output layer consists of M neurons arranged in a two-dimensional grid of nodes. Each node or neuron i (i = 1,2,…,M) is represented by an n-dimensional weight or reference vector $W_i = [w_{i1},\ldots,w_{in}]$, where n is the dimension of each input vector, i.e. the maximum number of variables in the input vector. In training, an input vector is randomly selected and its most similar reference vector is identified by determining the closest to it in terms of the Euclidian distance, D_i:

$$D_i = \sqrt{\sum_{j=1}^{n} m_j \left(x_j - w_{ij}\right)^2}; \quad i = 1,2,\ldots,M \tag{1}$$

where x_j is the j-th element of the current input vector; w_{ij} is the j-th element of the code vector i;

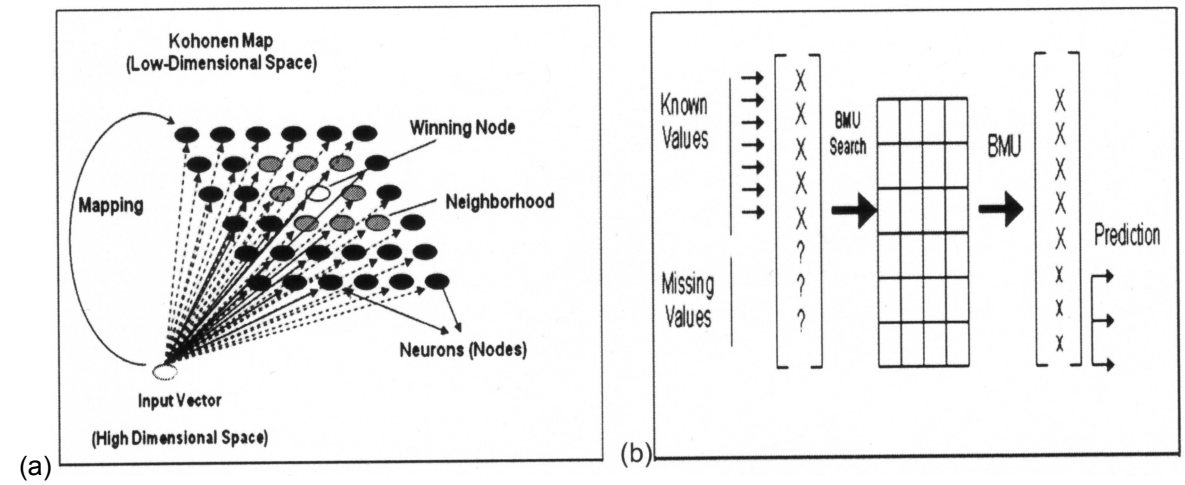

Fig. 1 (a) Illustration of the winning node and its neighbourhood in the Kohonen Self-Organizing Map; and (b) prediction of missing components of the input vector using the Kohonen Self-Organizing Map (BMU = best matching unit).

and m_j is the so called "mask" which is used to include in ($m_j = 1$), or exclude from ($m_j = 0$), the calculation of the Euclidian distance, the contribution of a given element x_j of the input vector. Once identified, the elements of the reference vector are updated and the process continues until some stated stopping criteria are reached. Rustum & Adeloye (2007) present additional comprehensive information about training the KSOM, including various indices for assessing its performance.

The application of the KSOM for prediction purposes is illustrated in Fig. 1(b) (Rustum & Adeloye, 2007). First, the model is trained using the available data set. Then the depleted vector, i.e. with the predictand either missing or deliberately removed, is presented to the KSOM to identify its BMU using the computed D_i (equation (1)). The values for the missing variables are then obtained as their corresponding values in the BMU.

METHODOLOGY

Data base

Daily data for both Edinburgh (UK) and Udaipur (India) were used. The Edinburgh weather station is located at 55°57′N, 3°13′W and at an altitude of 25 m a.s.l. and has a good quality record of weather data—air temperature, relative humidity, wind speed and incident solar radiation—spanning 1997–2006, i.e. 2282 daily records. The arid data were obtained from the meteorological observatory in Udaipur, India, at 24°35′N, 73°42′E, recording similar daily data to the Edinburgh station over five years, 2002–2006, i.e. a total of 1825 data points. Summary statistics of the weather variables together with the estimated PM-ETo are shown in Table 1 from which it is clear that the temperature, relative humidity and incident solar radiation at Udaipur are generally higher than those observed at Edinburgh. Indeed, the relative humidity at Udaipur, being close to saturation, is an indication that the potential evapotranspiration process will be less influenced by the aerodynamic term in comparison to the radiation term. A further observation in Table 1 is that PM-ETo estimates at the Udaipur site are much higher than similar estimates at the Edinburgh site, which is not surprising given the relatively temperate conditions at the latter.

Table 1 Statistical characteristics of the data used for training and (validation).

Variables:			Min.		Max.		Average		Std	
Symbols	Description	Units	UK	India	UK	India	UK	India	UK	India
T_{max}	Max. air temperature	°C	1.26 (2.65)	14.5 (19.6)	23.98 (21.55)	43.20 (42.40)	12.2 (12.20)	31.75 (31.60)	4.65 (4.54)	4.75 (4.749)
T_{min}	Min. air temperature	°C	–6.21 (–3.1)	–1.5 (1.0)	15.28 (13.10)	32.30 (30.50)	6.2 (6.01)	16.24 (17.20)	4.07 (3.99)	7.96 (7.81)
Rh_{max}	Max. relative humidity	%	65.49 (62.9)	14.0 (23.0)	93.41 90.39)	98.00 (100.0)	83.39 (79.64)	69.27 (77.71)	4.87 (4.77)	20.20 (17.91)
Rh_{min}	Min. relative humidity	%	18.56 (26.5)	5.0 (10.0)	85.97 79.13)	98.00 (100.0)	60.21 (52.16)	35.52 (41.18)	11.61 (11.81)	21.15 (21.38)
U_2	Wind speed	m/s	0.3 (0.4)	0.0 (0.2)	5.31 (3.54)	5.22 (11.39)	1.26 (1.15)	1.03 (1.18)	0.83 (0.65)	0.79 (0.86)
Rs	Solar radiation	$MJ/m^2/d$	0.8 (1.0)	6.36 (8.54)	18.48 (13.91)	28.32 (28.18)	4.9 (4.71)	19.48 (19.12)	3.87 (3.24)	4.61 (4.99)
N	Maximum possible sunshine hours	hours	6.7 (6.7)	1.0 (1.0)	17.32 (17.32)	12.20 (12.10)	12.11 (12.23)	8.56 (8.21)	3.61 (3.66)	2.56 (2.85)
Ra	Extraterrestrial radiation	$MJ/m^2/d$	4.1 (4.1)	22.9 (22.9)	41.50 (41.50)	40.40 (40.40)	22.29 (22.79)	32.71 (33.29)	13.29 (13.50)	6.20 (6.41)
PM-ETo	Penman-Monteith reference group evapo-transpiration	mm/d	0.2 (0.3)	1.1 (1.7)	2.70 (2.6)	8.89 (8.51)	1.14 (1.18)	4.10 (4.08)	0.56 (0.55)	1.72 (1.79)

KSOM analysis

Three different combinations of weather data and PM-ETo were considered for the KSOM analysis, as detailed in Table 2. Model 1 contains the full complement of input weather data required for the PM-ETo, while models 2 and 3 contain sub-sets of these and are required to test the performance against some of the incomplete ETo methods in the literature. For each location, the available data record was partitioned into two, with a part being used for model training and the other for independent validation. Thus at Edinburgh, the first 1500 data vectors were used for training and the remaining 782 vectors went into validation. The corresponding figures at Udaipur were 1215 and 610, respectively.

Table 2 Input weather variables for the KSOM models.

Model	Input variables (see Table 1)
1	T_{max}, T_{min}, Rh_{max}, Rh_{min}, U2, Rs
2	T_{max}, T_{min}, N, Ra
3	T_{max}, T_{min}, N

RESULTS AND DISCUSSION OF RESULTS

KSOM models

Due to lack of space, only the results of model 1 (see Table 2) are detailed, although as to be expected given the number of input variables used, model 1 was better than the other two. The component planes, a major feature of KSOM analysis, are shown in Fig. 2 and help to illustrate the associations or correlations between variables. The component planes show the values of the variables in each map unit that can be used to estimate the data variable of the input spaces (Vesanto *et al.*, 2000). The planes are filled using coloured or grey shades and the way the gradients of these colours or greys relate is an indication of the correlation. Variables exhibiting parallel colour or grey gradients will have high positive correlations; anti-parallel gradient is indicative of negative correlations. For example in Fig. 2(a), grey gradients for the PM-ETo and Rs are parallel whereas those between the PM-ETo and RH_{min} are strongly anti-parallel at Edinburgh. Both these suggest strong positive correlation of PM-ETo with Rs and strong negative correlation with RH_{min}, which in turn indicates that both the mass transfer and solar radiation effects are equally significant in determining ETo at the UK site. At Udaipur, however, while the grey gradients of the PM-ETo and Rs component planes are strongly parallel, there was no clear-cut pattern to the relationship between PM-ETo and RH_{min} grey gradients (see Fig. 2(b)). From the summary presented in Table 1, it is quite clear that the humidity deficit at Udaipur is very low and, hence, mass transfer is unlikely to have much influence on the ETo process, a fact now borne out by the component planes. Other correlations can be similarly visualized from the component planes and such information can become useful in identifying the most significant weather variables for the ETo in different locations so that available resources for data collection can be better targeted.

The performance of model 1 in predicting the PM-ETo during training is illustrated in the X–Y plots in Fig. 3 for both Edinburgh (R = 0.98) and Udaipur (R = 0.99), which is very good. Although the relevant plots are not shown here for lack of space, model 1 at both locations was equally good during validation, with the corresponding R = 0.97 at Edinburgh and R = 0.99 at Udaipur. Models 2 and 3, which used fewer weather input data, and whose full details have been omitted for lack of space, also recorded satisfactory performances in predicting the PM-ETo during both training and validation. For example at Edinburgh, the R values for model 2 were 0.96 and 0.94, respectively, for training and validation, while the corresponding values for model 3 were 0.97 and 0.95. At Udaipur, models 2 and 3 correlations during training and validation were both 0.98 and 0.99, respectively.

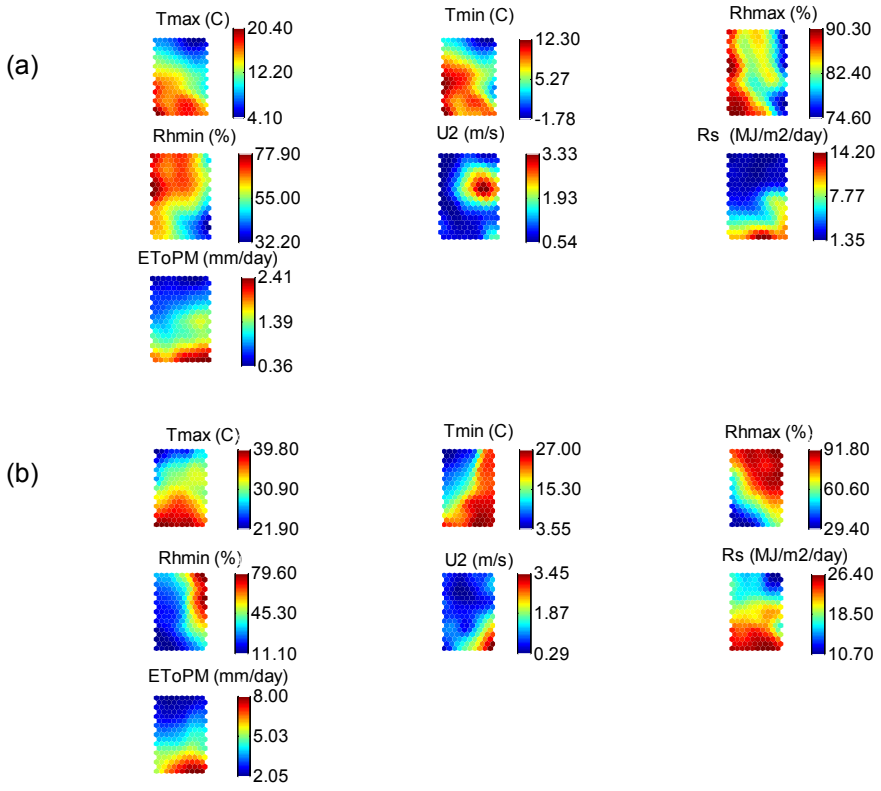

Fig. 2 SOM component planes (grey) for model 1 at (a) Edinburgh (UK) and (b) Udaipur (India).

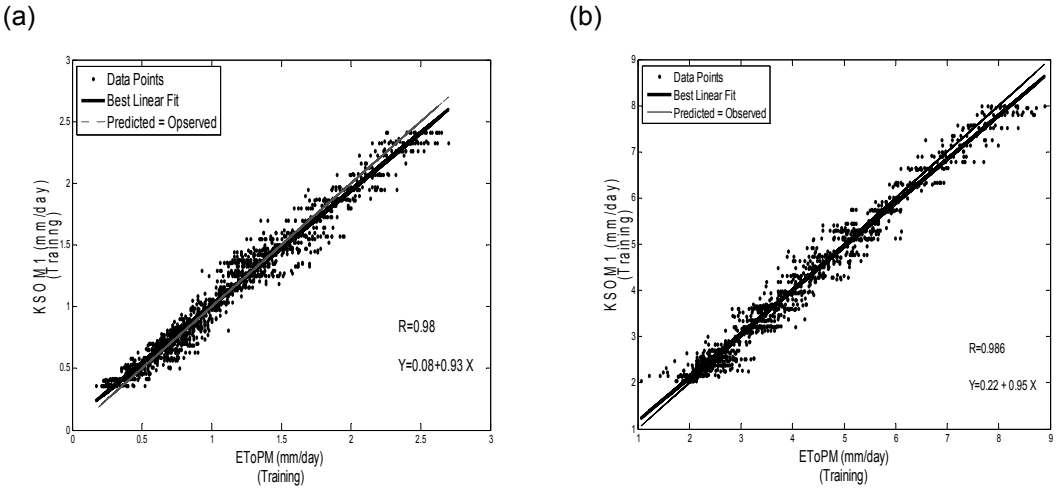

Fig. 3 X–Y scatter plot of observed and KSOM-predicted PM-ETo during training: (a) Edinburgh and (b) Udaipur.

Comparison of KSOM and established empirical models

To demonstrate the performance of the KSOM models against other widely-used empirical ETo models, Fig. 4 compares the mean monthly evapotranspiration estimates of nine alternative models with the PM-ETo estimates for the training data sets. As seen in Fig. 4, whilst the three KSOM models nearly match the PM-ETo perfectly, none of the four empirical models tested: Hargreaves, Thornthwaite, Blaney-Criddle and Priestley-Taylor, in their basic forms were able to do so, especially during the summer months. As noted previously, empirical models are location-

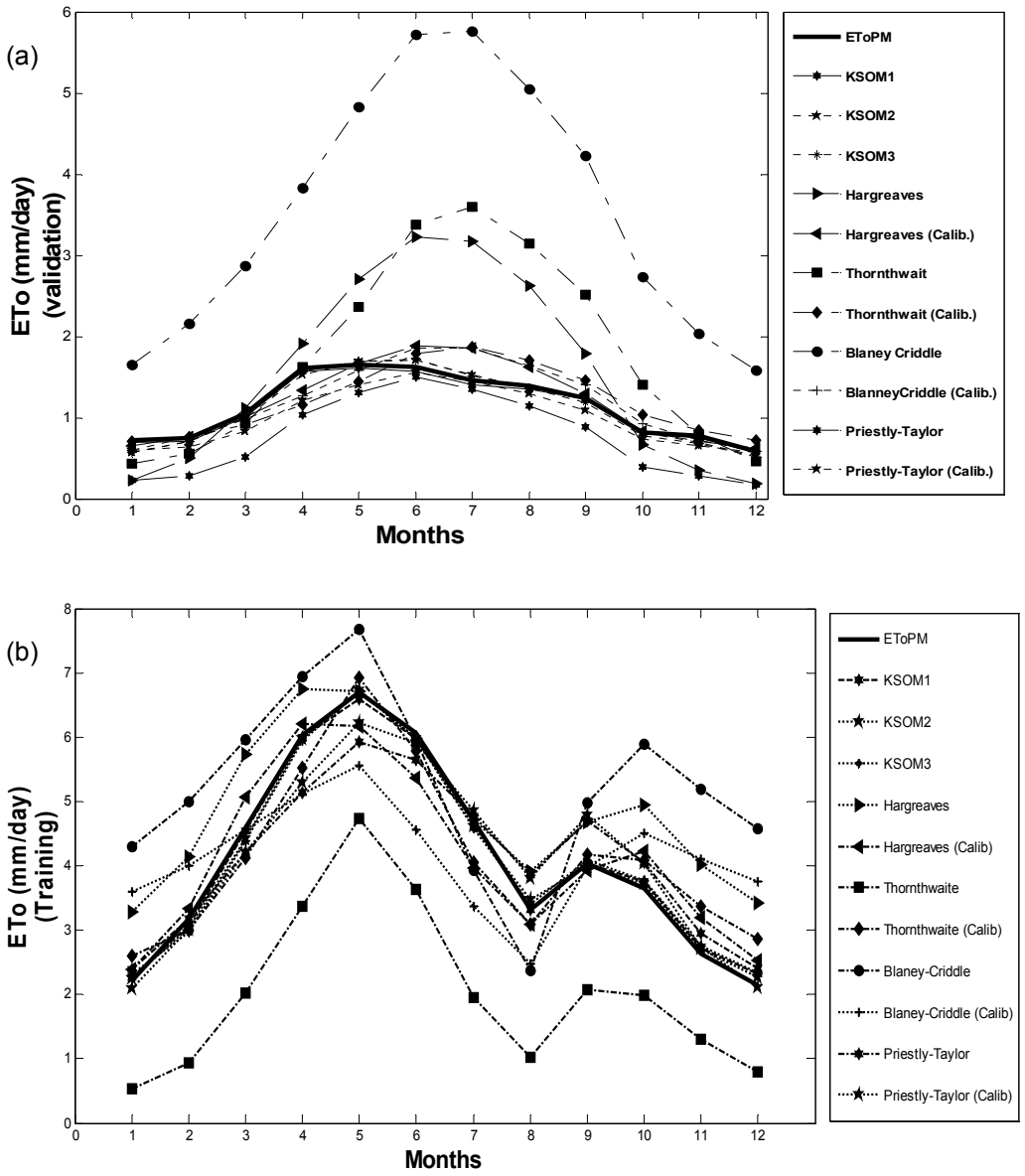

Fig. 4 Comparison of estimated mean monthly ETo for nine different model assumptions with the training data set: (a) Edinburgh, and (b) Udaipur.

specific and may not therefore be expected to do well at locations different from where they were developed. This is why it is recommended to calibrate these models against local climate before using them (Allen *et al*., 1998). However, as shown in Fig. 4, even the calibrated versions of these empirical models could not match the performance of any of the KSOM models based on the training data sets. Similar plots for the validation sets are available but are not shown here because of their striking similarity to Fig. 4(a) and (b). However, Table 3 contains the biases, i.e. model estimate minus PM-ETo, for the validation set, from which it is clear that the KSOM models are also far superior to the empirical models.

CONCLUSIONS

The current work presents a methodology based on the use of Kohonen Self-Organizing Map (KSOM) models to predict the PM-ETo for sub-humid and arid catchments. Extensive testing and

Table 3 Bias (mm) in estimated mean daily ETo at Edinburgh and Udaipur by different model assumptions during validation. Udaipur values are in parenthesis.

Month	1	2	3	4	5	6	7	8	9	10	11	12
KSOM1	0.03 (–0.09)	0.05 (0.05)	0.05 (0.25)	0.05 (0.10)	0.04 (0.27)	0.07 (0.16)	0.06 (–0.02)	0.04 (–0.07)	0.01 (–0.06)	0.01 (–0.03)	0.04 (0.02)	0.01 (–0.21)
KSOM2	0.1 (–0.07)	0.05 (0.12)	0.05 (0.13)	0.08 (0.02)	–0.04 (0.15)	–0.07 (0.11)	–0.07 (–0.08)	0.03 (0.02)	–0.01 (–0.08)	0.00 (–0.07)	0.09 (–0.01)	0.06 (–0.12)
KSOM3	0.15 (–0.14)	0.04 (–0.07)	0.04 (0.34)	0.04 (0.17)	–0.04 (0.19)	–0.08 (0.21)	–0.07 (0.01)	0.03 (0.13)	0.06 (–0.10)	0.04 (–0.12)	0.10 (–0.02)	0.06 (–0.20)
Hargreaves	0.49 (–1.17)	0.25 (–1.40)	–0.06 (–1.31)	–0.30 (–1.21)	–1.05 (0.19)	–1.59 (0.27)	–1.71 (–0.25)	–1.23 (–0.40)	–0.55 (–0.57)	0.14 (–1.16)	0.43 (–1.33)	0.39 (–1.12)
Hargreaves calibrated	0.08 (–0.28)	0.00 (–0.66)	0.04 (–0.62)	0.28 (–0.67)	–0.01 (0.72)	–0.24 (0.85)	–0.40 (0.54)	–0.24 (0.45)	–0.05 (0.23)	–0.02 (–0.41)	0.08 (–0.50)	–0.05 (–0.21)
Thornthwaite	0.29 (1.66)	0.19 (1.96)	0.13 (2.51)	–0.01 (2.69)	–0.71 (2.37)	–1.75 (2.35)	–2.13 (2.79)	–1.74 (2.44)	–1.27 (1.97)	–0.60 (1.70)	–0.01 (1.51)	0.13 (1.46)
Thornthwaite calibrated	0.02 (–0.40)	0.00 (–0.13)	0.15 (0.42)	0.46 (0.56)	0.19 (0.18)	–0.16 (0.17)	–0.41 (0.71)	–0.31 (0.37)	–0.23 (–0.13)	–0.23 (–0.40)	–0.06 (–0.56)	–0.14 (–0.60)
Blaney-Criddle	–0.92 (–2.14)	–1.40 (–2.39)	–1.82 (–1.63)	–2.22 (–1.14)	–3.18 (–0.79)	–4.09 (–0.27)	–4.30 (1.52)	–3.66 (0.90)	–2.99 (–0.53)	–1.92 (–2.31)	–1.25 (–2.44)	–1.00 (–2.20)
Blaney-Criddle calibrated	0.12 (–1.40)	0.00 (–1.04)	0.08 (–0.35)	0.34 (0.58)	0.07 (1.40)	–0.22 (1.57)	–0.41 (1.46)	–0.25 (0.87)	–0.16 (0.16)	–0.13 (–0.92)	0.06 (–1.42)	0.00 (–1.52)
Priestley-Taylor	0.50 (–0.27)	0.47 (–0.40)	0.53 (–0.44)	0.57 (0.30)	0.34 (0.95)	0.13 (0.68)	0.11 (–0.09)	0.25 (–0.44)	0.36 (–0.64)	0.42 (–0.56)	0.50 (–0.38)	0.41 (–0.23)
Priestley-Taylor calibrated	0.12 (0.02)	0.11 (–0.34)	0.22 (–0.50)	0.40 (0.11)	0.24 (0.62)	0.07 (0.36	0.01 (–0.07)	0.10 (–0.40)	0.15 (–0.68)	0.08 (–0.55)	0.13 (–0.18)	0.01 (0.10)

validation of the models showed that they can produce estimates comparable to the PM-ETo, even when forced with only a sub-set of the panoply of input weather data required to drive the full implementation of the PM-ETo. This was certainly the case for models 2 and 3 which, despite the fact that they used fewer input weather variables, still predicted the PM-ETo with correlations of at least 0.95. Additionally, for these two models only the minimum and maximum temperature (T_{min} and T_{max}) are required to be measured in order to estimate PM-ETo, making them a useful tool in poorly instrumented catchments. Similarly, the developed KSOM models outperformed some of the established empirical models for estimating evapotranspiration, both in their basic forms and when calibrated against the PM-ETo. This is a significant breakthrough for regions with inadequate data for evapotranspiration estimation.

A major feature of the KSOM modelling is that its predictive ability is unencumbered if some of its predictor variables are missing. Model 1, the best performing of the KSOM models, has the full complement of six input weather variables; however, unlike the PM-ETo calculations, the KSOM model will still predict the ETo if any of the six variables is missing. All of this offers promise for poorly instrumented catchments such as those in the developing world.

REFERENCES

Allen, R. G., Pereira, L. S., Rae, D. & Smith, M. (1998) Crop evapotranspiration: Guidelines for computing crop water requirements. *FAO Irrigation and Drainage paper 56*. FAO, Rome, 300pp.

Back, B., Sere, K. & Hanna, V. (1998) Managing complexity in large database using self organising map. *Accounting Management & Information Technologies* **8**, 191–210.

George, B. A., Reddy, B. R. S., Raghuwanshi, N. S. & Wallender, W. W. (2002) Decision support system for estimating reference evapotranspiration. *J. Irrig. & Drainage Engng* **128**(1), 1–10.

Itenfisu, D., Elliott, R. L., Allen, R. G. & Walter, I. A. (2003) Comparison of reference evapotranspiration calculations as part of the ASCE standardization effort, *J. Irrig. & Drainage Engng* **129** (60), 440-448.

Kohonen, T., Simula, O. & Visa, A. (1996) Engineering applications of the Self-organizing Map. *Proc. IEEE* **84**(10), 1358–1384.

Rustum R. & Adeloye, A. J. (2007) Replacing outliers and missing values from activated sludge data using Kohonen Self Organizing Map. *J. Environ. Engng* **133**(9), 909–916.

Rustum R., Adeloye, A. J. & Simala, A. (2007) Kohonen Self-Organising map (KSOM) extracted features for enhancing MLP-ANN prediction models for BOD5. In: *Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century* (ed. by B. W. Webb & D. de Boer), 181–178. IAHS Publ. 314. IAHS Press, Wallingford, UK.

Rustum R., Adeloye, A. J. & Scholz, M. (2008) Applying Kohonen self-organizing map as a software sensor to predict the biochemical oxygen demand. *Water Environ. Res.* **80** (1), 32–40.

Vesanto, J., Himberge, J., Alhoniemi, E. & Parhankangas, J. (2000) Self-organizing Map (SOM) Toolbox for Matlab 5. Report no. A57, Helsinki University of Technology, Laboratory of Computer and Information Science, Helsinki, Finland.

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. & Drainage Engng* **133**(2), 83–89.