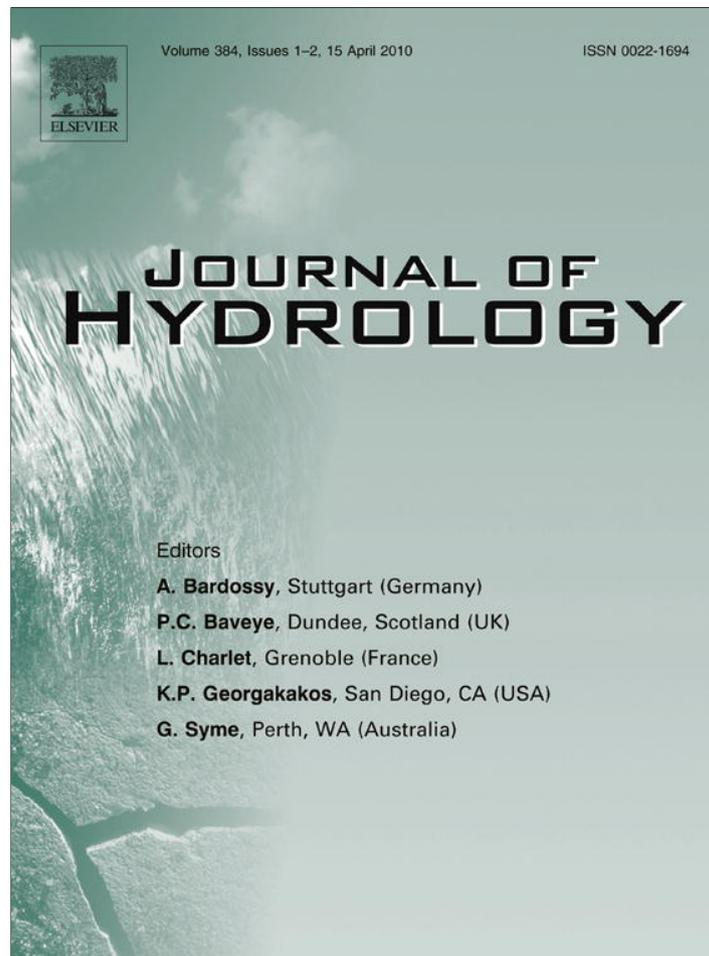


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network

Fi-John Chang^{a,*}, Li-Chiu Chang^b, Huey-Shan Kao^a, Gwo-Ru Wu^a

^aDepartment of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, Taiwan, ROC

^bDepartment of Water Resources and Environmental Engineering, Tamkang University, Taipei, Taiwan, ROC

ARTICLE INFO

Article history:

Received 8 August 2009

Received in revised form 30 December 2009

Accepted 15 January 2010

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Taha Ouarda, Associate Editor

Keywords:

Artificial neural network

Evaporation

Meteorological variables

Self-organizing map

SUMMARY

The phenomenon of evaporation affects the distribution of water in the hydrological cycle and plays a key role in agriculture and water resource management. We propose a self-organizing map neural network (SOMN) to assess the variability of daily evaporation based on meteorological variables. The daily meteorological data sets from a climate gauge were collected as inputs to the SOMN and then were classified into a topology map based on their similarities to investigate their multi-collinear relationships to assess their effort in the evaporation. To accurately estimate the daily evaporation based on the input pattern, the weights that connect the clustered centers in a hidden layer with the output were trained by using the least square regression method. In addition, we compared the results with those of back propagation neural network (BPNN), modified Penman and Penman–Monteith formulas. The results demonstrated that the topological structures of SOMN could give a meaningful map to present the clusters of meteorological variables and the networks could well estimate the daily evaporation. By comparing the performances of these models in estimating daily and long-term (monthly or yearly) cumulative evaporation, the SOMN provides the best performance.

© 2010 Elsevier B.V. All rights reserved.

Introduction

Evaporation is a key input to hydrological models. Estimating evaporation is fundamental for agricultural irrigation, water balance studies, water supply, and land resource planning. However, it is difficult to accurately estimate because of complex interactions between the components of the land and atmosphere systems. Temperature, wind speed, atmospheric pressure, and solar radiation could all influence the amount of evaporation. In hydrological practice, the estimation can be achieved by direct or indirect methods. One of the direct methods for evaporation measurements is pan evaporation. A practical means of estimating the amount of pan evaporation is of considerable significance to hydrologists and agriculturists. Indirect methods, such as mass transfer and water budget methods, based on meteorological data have been used to estimate evaporation on a water body by many researchers (Burman, 1977; Coulomb et al., 2001; Gavin and Agnew, 2004). The FAO Penman–Monteith equation is recommended as the standard method for computation of daily reference evapotranspiration (Allen et al., 1998). Warnaka and Pochop (1988) compared six equations to estimate evaporation using climatologic data and stated that the equations vary greatly in their ability to

define the variability of evaporation. The accuracy of the evaporation estimate is highly dependent on the reliability and precision of a number of measurements involved in radiation measurements. The wide range of data type and the expertise needed to correctly use the various equations make it difficult to select the most suitable equation to use for a given study.

Many researchers have emphasized the need for accurate estimates of evaporation in hydrologic modeling studies (Sudheer et al., 2002; Szilagyi and Jozsa, 2009). This requirement could be addressed through better models that will address the inherent non-linearity in the process. The artificial neural network (ANN) often serves as a viable alternative to physical modeling in real time applications. It is used to characterize mapping of the input to the output directly with less emphasis on the internal structure driving the physical process. Neural network approaches have been successfully applied to a number of diverse fields. In the hydrological context, recent experiments have reported that the ANN may offer a promising alternative (Chang and Chang, 2001; Tayfur, 2002; Cancelliere et al., 2002; Chiang et al., 2004; Supharithid, 2003; Kumar et al., 2004; Kisi, 2005; Cigizoglu and Kisi, 2005; Chau, 2006; Chen et al., 2008; Wu et al., 2008; Chen and Chang, 2009).

The application of ANN in the fields of the evaporation and evapotranspiration has also been the subject of intense research (Bruton et al., 2000; Sudheer et al., 2003; Trajkovic et al., 2003;

* Corresponding author. Tel.: +886 2 23639461; fax: +886 2 23635854.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

Keskin and Terzi, 2006a; Kisi, 2006; Parasuraman et al., 2007). Odhiambo et al. (2001) stated the optimized fuzzy-neural-model is reasonably accurate, and is comparable to the FAO Penman–Monteith equation. Sudheer et al. (2002) used the neural network model for the evaporation process using proper combinations of the observed climatic variables such as temperature, relative humidity, sunshine duration, and wind speed. Kisi (2006) used the neuro-fuzzy model to estimate the daily PE using observed climatic variables such as air temperature, solar radiation, wind speed, pressure, and relative humidity for the neuro-fuzzy model. Kisi and Ozturk (2007) used the neuro-fuzzy model to estimate the evapotranspiration using the observed climatic variables. An uncertainty based on the neural network model can be ascribed not only to the modeling process but also to the limited data used for the training performance of the neural network model. Keskin and Terzi (2006b) used ANN model as an alternative approach to evaporation estimation and demonstrated the ANN estimations of daily pan evaporation is better than the Penman estimations. Kim and Kim (2008) proposed the generalized regression neural network model embedded with the genetic algorithm to estimate the pan evaporation (PE) and the alfalfa reference evapotranspiration (ET_r), and claimed that PE and ET_r maps could be constructed to provide the reference data for a drought analysis and an irrigation network system.

Feature maps have the property of representing regions of high signal density on the topological structure, preserving neighborhood relations, and displaying important statistical characteristics of input patterns. Users could visually label clusters on the map and gain an idea over the structure of the data. This distinct feature of the human brain motivates the development of the self-organizing maps (SOM). Introduced by Kohonen (1982), the SOM algorithm is capable of generating mappings from high-dimensional signal spaces to lower dimensional topological structure, and has been widely applied to pattern recognition, classification, speech processing and geometric models (Barhak and Fischer, 2002). There are also a number of studies which use the SOM to solve various water resource problems. For instance, Hsu et al. (2002) presented the self-organizing linear output map (SOLO) for hydrologic modeling and analysis. Richardson et al. (2003) claimed the SOM-algorithm succeeds in providing insight in the groundwater quality data set, highlighting the main differences between groups of samples and pointing out anomalous wells and well screens. Regis et al. (2005) investigated the suitability of the self-organizing map neural network for patterning habitat invasion by exotic fish species. Tadeusz et al. (2006) applied the self-organizing maps to revealing variation in non-obligatory riverine fish. Grieu et al. (2006) developed a data exploration technique for wastewater treatment monitoring based on self-organizing map. Chang et al. (2007) presented an SOM network for flood forecasting and stated that it has great efficiency for clustering, especially for peak flow, and super capability of modeling flood forecasts.

The purpose of this study is to assess the effort of meteorological variables for evaporation through the constructed self-organizing map neural network (SOMN). Our interest lies on building an artificial topographic map to demonstrate that the spatial location of an output in the map would correspond to a particular feature of data drawn from the input space and gain ideas over the formulated map of input–output patterns. The 6 years' (2001–2006) daily meteorological data of the Hengchun weather station in South Taiwan are the focus of this analysis. The effectiveness of the SOMN for building the relationships among meteorological variables with evaporation are examined and discussed, and the reliability of the constructed SOMN for estimating the evaporation based on the meteorological variables is compared with those of the commonly used modified Penman and Penman–Monteith formula and the back propagation neural network (BPNN).

Methods

The SOM neural network (SOMN)

Kohonen (1982) self-organizing feature map (SOM) algorithm is a simple yet powerful learning process and an effective clustering method. A self-organizing map consists of components called nodes. Associated with each node are a weight vector of the same dimension as the input data vectors and a position in the map space (Fig. 1). Self-organizing maps use a neighborhood function to preserve the topological properties of the input space. It can transform high dimensional input patterns into the responses of two-dimensional arrays of neurons and perform this transformation adaptively in a topologically ordered fashion based on similarity, thus facilitates the detection of the inherent structure and the interrelationship of data. The training algorithm is summarized as follows:

- Step1. Initialization:* Choose random values for the initial weights.
- Step 2. Winner-finding:* Find the winning neuron using the minimum distance Euclidean criterion. The neuron with weight vector most similar to the input is called the winner.
- Step3. Weight-updating:* Adjust the weights of the winner and its neighborhood neurons towards the input vector. The magnitude of the change decreases with time and with distance from the winner.

This process is repeated for a large number of cycles until the map is well unfolded. After a certain number of iterations (e.g., 2000), if the map has still not unfolded, it is suggested to restart the training process with a different set of initial weights.

The LMS algorithm for optimizing output weights

Like most artificial neural networks, SOMs operate in two cases: training and mapping. As mentioned above, training builds the map using input examples, while mapping automatically classifies a new input vector. During mapping, there will be one single winning neuron: i.e. the neuron whose weight vector lies closest to the input vector. Because the SOM uses an unsupervised learning algorithm for clustering the input patterns into groups of similar patterns, the output of the SOM, like other cluster methods, to the input vectors can only recall the classifying results specified in the network. This way is only good for recognizing problems and cannot be effectively used for continuous function approximation. Because similar input patterns could have various outputs, one easy way to determine the output for a given input pattern is to use the average output value as the clustered input patterns to the correspondent neuron and then directly use the output of the closest (most similar) neuron for the given input pattern (Kohonen, 1990). To be more generalized in extending the use of the topological structure of SOM, a weighted sum of the outputs is commonly used. In this study, we implement the least-mean-square (LMS) method to search the connected weights between neurons in the hidden layer and output layer. The learning algorithm is summarized as follows.

Once the centers are determined, the output of the network to the input vectors can be computed as

$$\hat{y}(q) = \sum_{i=1}^m \alpha_i \phi(X(q), W_i), \quad q = 1, 2, \dots, Q \quad (1)$$

$$\phi(X(q), W_i) = \phi(\|X(q) - W_i\|_2), \quad q = 1, 2, \dots, Q \quad (2)$$

$$\phi(a) = \exp(-a^2)$$

where $\hat{y} \in \mathfrak{R}^{1 \times 1}$ is the actual network output, $X \in \mathfrak{R}^{n \times 1}$ is the input vector, Q is the number of samples, $\phi_i(\cdot)$ is a Euclidean function,

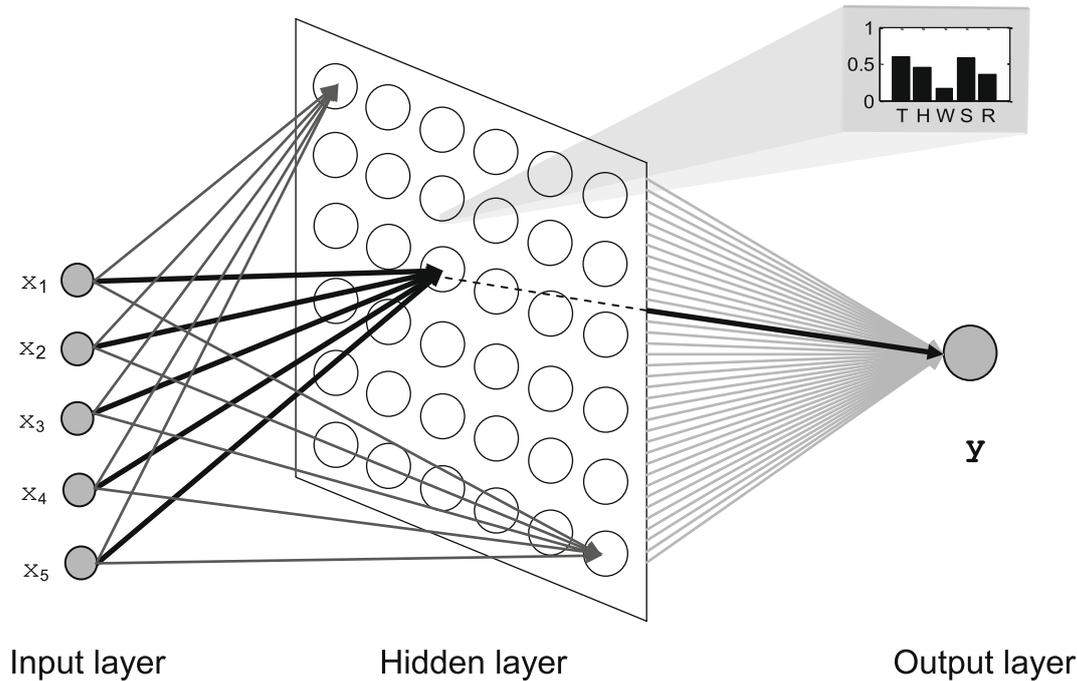


Fig. 1. The architecture of self-organizing map network.

α_i are the weights in the output layer, m is the number of neurons in the hidden layer, and $W_i \in \mathbb{R}^{n \times 1}$ are the SOM centers in the hidden layer.

The optimal set of weights minimizes the performance measures

$$J(w) = \frac{1}{2} \sum_{q=1}^Q [y_d(q) - \hat{y}(q)]^2 \quad (3)$$

where $y_d \in \mathbb{R}^{1 \times 1}$ denotes the desired network output. The weights (α_i) can then be easily obtained by using the training sets (input–output patterns) through the least-mean-square algorithm (LMS). Fig. 1 presents the prototype of SOMN used in this study. It includes the input variables, the winner clustered hidden node, and the weighted sum of the outputs.

Determining the structure of the networks

The general architecture of the SOMN includes the input layer, clustering layer and output layer. There are five input variables and one output in Table 1. Determining the appropriate size of clustering neural networks is important for validity and efficiency of clustering. Since there is no systematic or standard method for finding the optimal number of clusters in the clustering algorithms,

Table 1 The statistics results of meteorological variables in the Hengchun weather station.

Meteorological variables	Mean	SD ^a	CC ^b	Maximum	Minimum
Temperature (°C)	25.30	3.22	0.250	30.4	15.0
Humidity (%)	71.50	7.47	-0.387	94	49.0
Wind speed (m/s)	3.52	1.91	0.187	11.2	0.7
Sunshine hour (h)	6.51	3.32	0.564	12.5	0
Solar radiation (MJ/m ² /day)	11.59	5.89	0.577	25.43	0
Evaporation (mm/day)	4.50	1.53	1	9.5	0.8

^a SD: standard deviation.
^b CC: coefficient of correlation.

the optimal network size is based on trial-and-error in the circumstance. To determine a suitable size of the SOM network, a number of networks (i.e., 3 × 3, 4 × 4, 5 × 5, 6 × 6, 7 × 7, and 8 × 8) were chosen to train, validate, and test the constructed networks based on training, validation and testing data sets, respectively. The results are presented in Fig. 2. In the training case, as expected, the estimation error (root mean square error) is decreasing as the size of network is increased, while in the validation case, the 6 × 6 network has the minimum estimation error. Consequently, the 6 × 6 network is chosen to be applied to the testing case and further interpolation.

Back propagation neural networks

The back propagation neural network (BPNN) is the most popular and widely used neural network in use today. Details on the

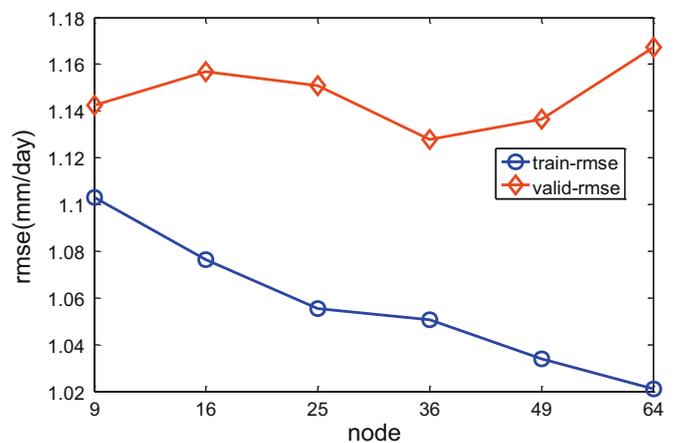


Fig. 2. The number of SOM hidden nodes and their corresponding errors in the training and validation cases.

standard BP algorithm can be found in the literature (Rumelhart et al., 1986). For the purpose of comparison, the BPNN is investigated through a different number of hidden nodes and a great number of initial settings for the connected weights. The input–output patterns are the same for building the SOMN. To determine a suitable BPNN for the given input–output patterns, we also investigated several network architectures. Because the input vector and output are set the same as previously, i.e., five input variables and one output, only the number of nodes in the hidden layer needs be determined. The networks with various numbers of hidden nodes, from 1 to 10, are performed; furthermore, each network is executed with a number of initial sets of connected weights to make sure the final results are reason-

able. We find when the hidden layer has two nodes, both training and validation sets have the best performance. This simple structure of the BPNN is then used for the purpose of comparison only.

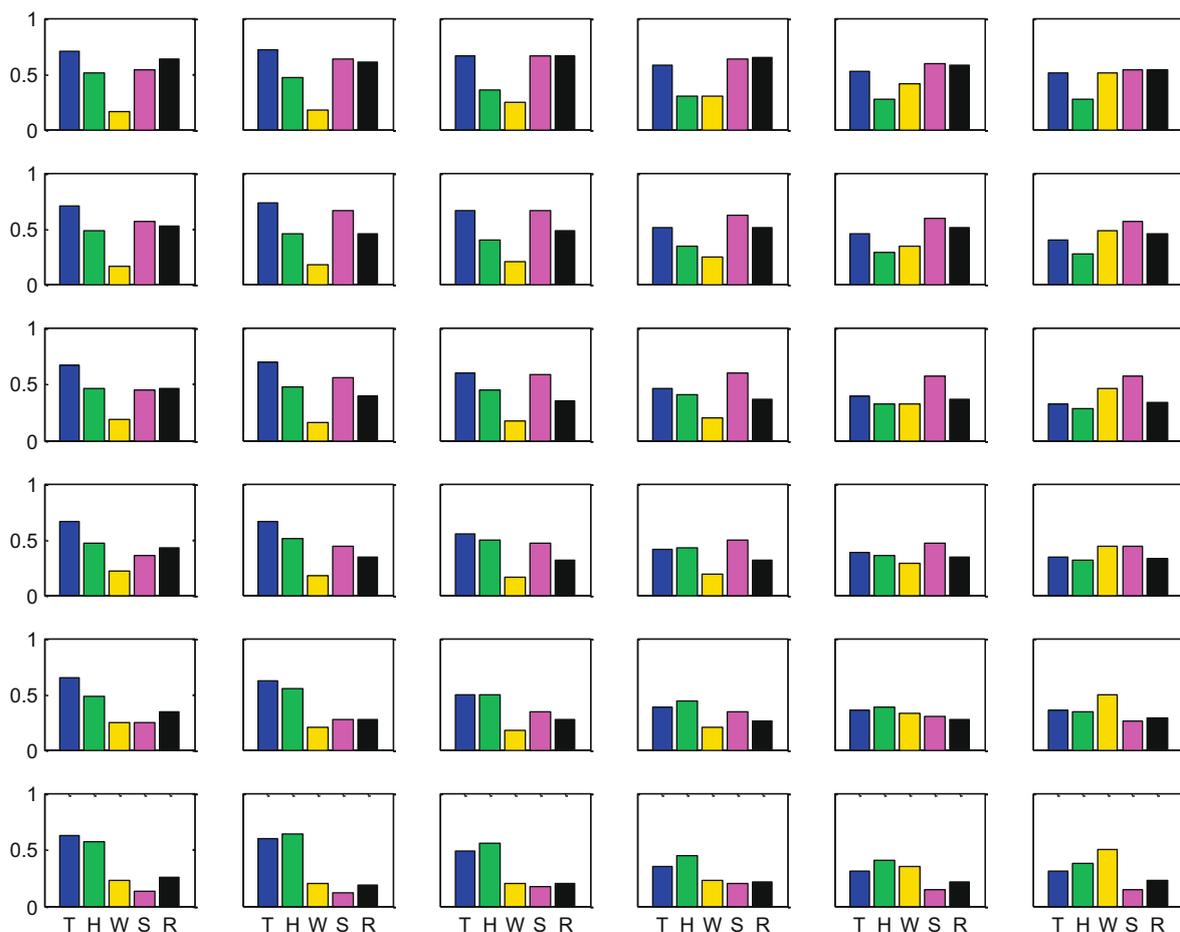
The modified Penman and Penman–Monteith estimators

Xu and Singh (1998) noted that the values of the Penman method agree most closely with the pan evaporation values. The FAO-modified Penman equation has gained acceptance as the standard method of estimating reference crop evapotranspiration. For the purpose of comparison, two popular evaporation estimation models, the modified Penman and Penman–Monteith, were used.

Table 2
The estimation errors in three cases by four different methods.

Models	Training case	Validation case	Testing case		
	RMSE (mm/day)	RMSE (mm/day)	RMSE (mm/day)	MAE (mm/day)	EC
BPNN	1.064	1.077	1.172	0.891	0.569
SOMN	1.048	1.121	1.162	0.881	0.572
Modified Penman	1.263	1.261	1.305	1.035	0.520
Penman–Monteith	1.137	1.252	1.307	1.014	0.519

RMSE: root mean square error; MAE: mean absolute error; EC: the Nash–Sutcliffe efficiency.



T :temperature ; H: humidity; W: wind speed; S: sunshine hour ; R :radiation

The normalized values shown in the ordinate

Fig. 3. The topology map of SOM (6 × 6).

The modified Penman equation:

$$ET_0 = \frac{\Delta \times 86,400 \times \frac{(R_n - S)}{\lambda} + \gamma \frac{900}{(T+273)} u (e_a - e_d)}{\Delta + \gamma(1 + 0.337u)}$$

$$ET_0 = \frac{\Delta}{\Delta + \gamma} \times 86,400 \times \frac{(R_n - S)}{\lambda} + \frac{\gamma}{\Delta + \gamma} \times 2.7 \times f(u) \times (e_a - e_d)$$

The Penman–Monteith equation:

where ET_0 is the reference evapotranspiration (mm/day), R_n is the net radiation (W/m^2), S is the soil heat flux density (W/m^2), λ is

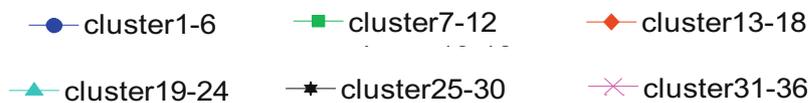
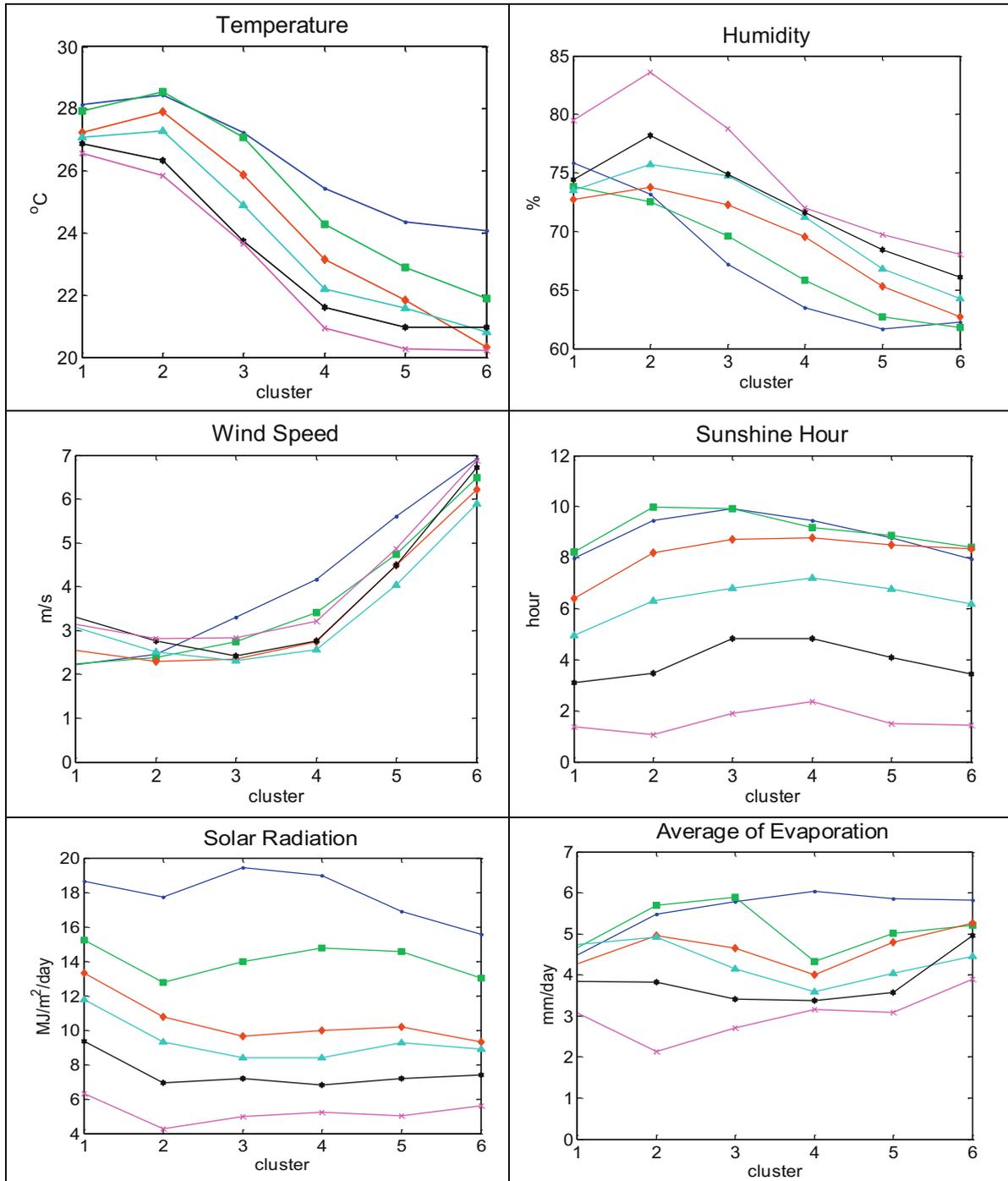


Fig. 4. The clustering results of the meteorological variables in the map.

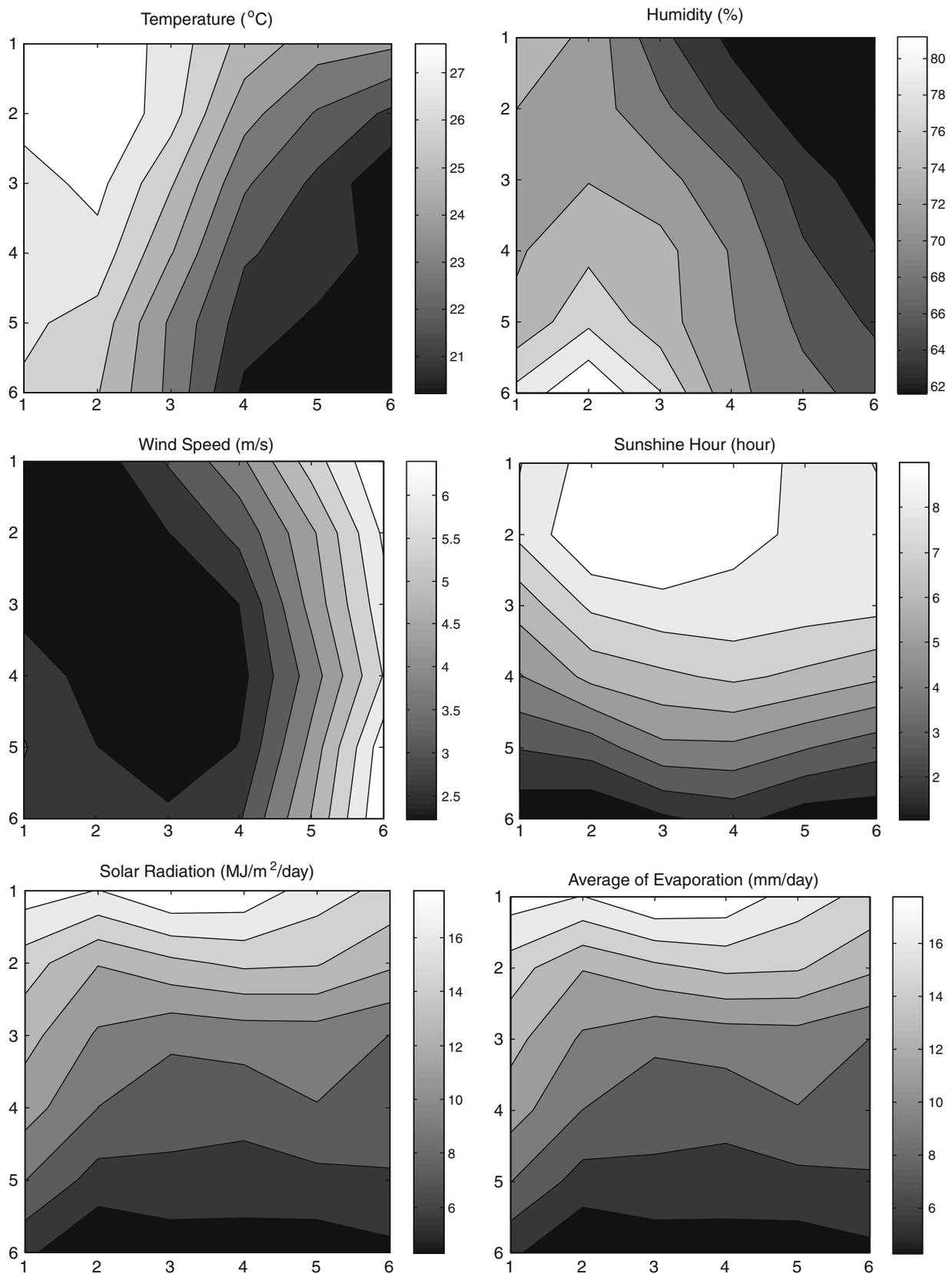


Fig. 5. The clustering results of the meteorological variables in the map.

the latent heat of vaporization (J/kg), T is the mean daily air temperature at 2 m height ($^{\circ}\text{C}$), u is the wind speed at 2 m height (m/s), e_d is the saturation vapour pressure (kPa), e_a is the actual vapour pres-

sure (kPa), $e_s - e_d$ is the saturation vapour pressure deficit (kPa), Δ is the slope vapour pressure curve ($\text{kPa } ^{\circ}\text{C}^{-1}$), and γ is the psychrometric constant ($\text{kPa } ^{\circ}\text{C}^{-1}$).

Results and discussion

Description of data

The daily climatic data of the Hengchun weather station (latitude 22°00' 19.56", longitude 120°44' 16.99") operated by the Central Weather Bureau (CWB), Taiwan, were used in the study. It is located in southeast Taiwan at the elevation of 22.1 m. The measured daily climatic data for the station were downloaded from the CWB web server. The data sample consisted of 6 years'

(2001–2006) daily records of air temperature (T), solar radiation (SR), wind speed (W), sunshine hours (S), humidity (H) and pan evaporation (E). Because the solar radiation data were recorded as zero during 2004/8/15–9/21, we deleted the data in those 2 months. A total of 34-month data in the years of 2002–2004 were used, among which, 26-month data (792 daily data sets) for training case, 4-month data (121 data sets) for validation case, and 4-month data (122 data sets) for testing case. The data sets in the years of 2001, 2005 and 2006 were used further for model evaluation.

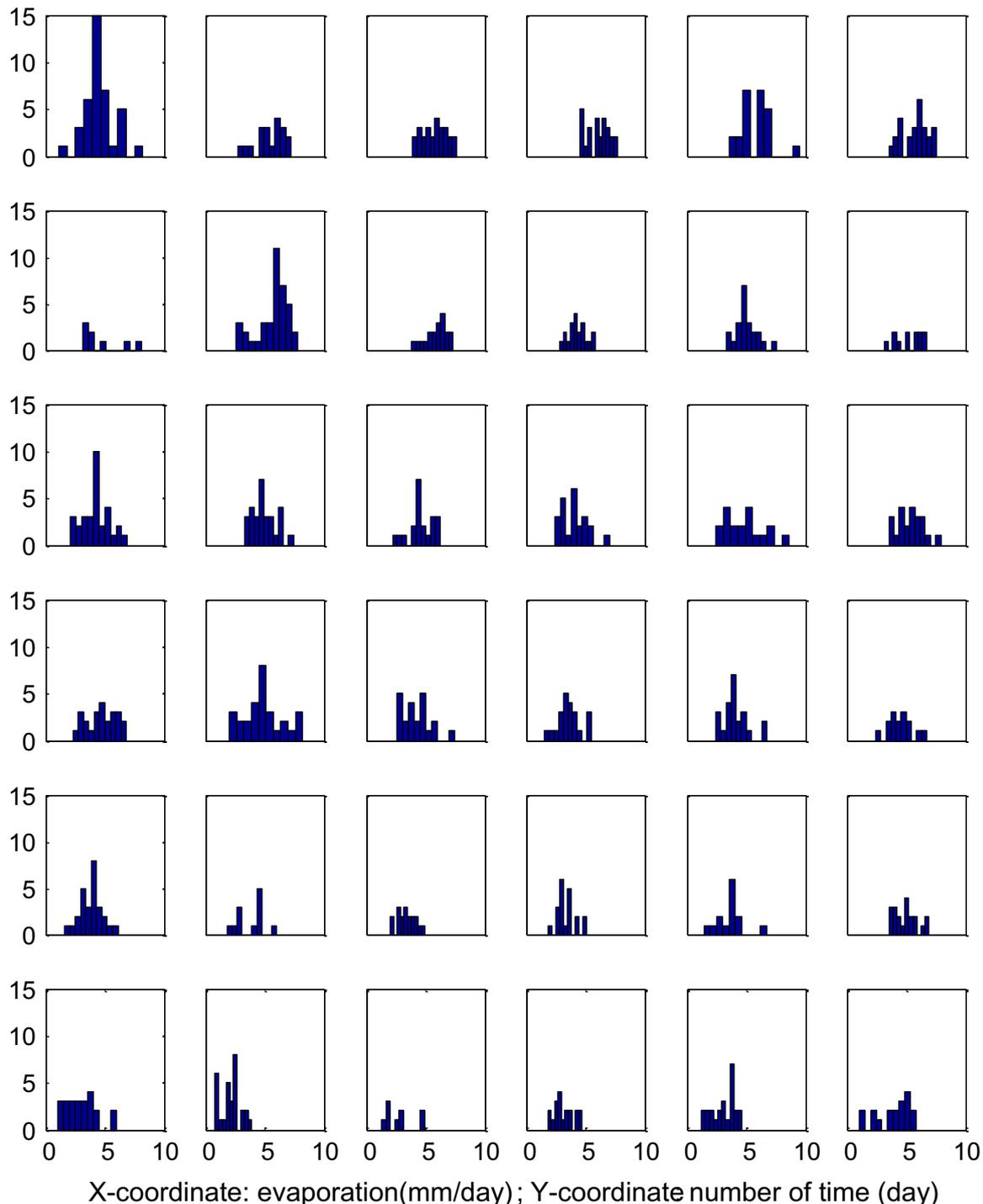


Fig. 6. The evaporation distributions in the clustered nodes of the 6 × 6 topology map.

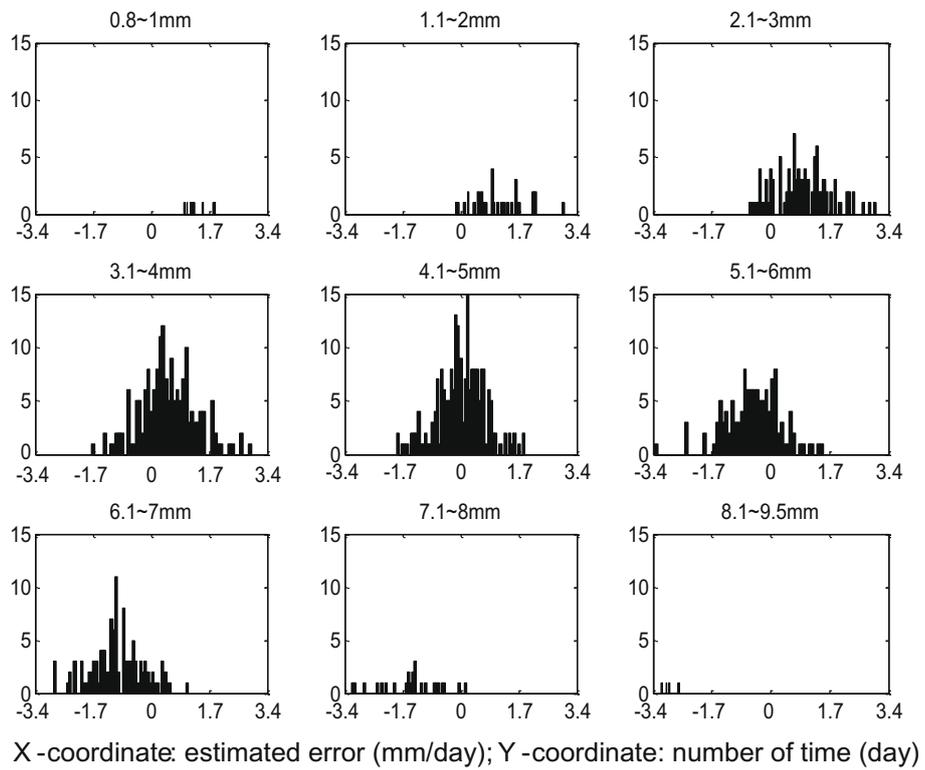


Fig. 7. The estimated error distributions of evaporation in different ranges.

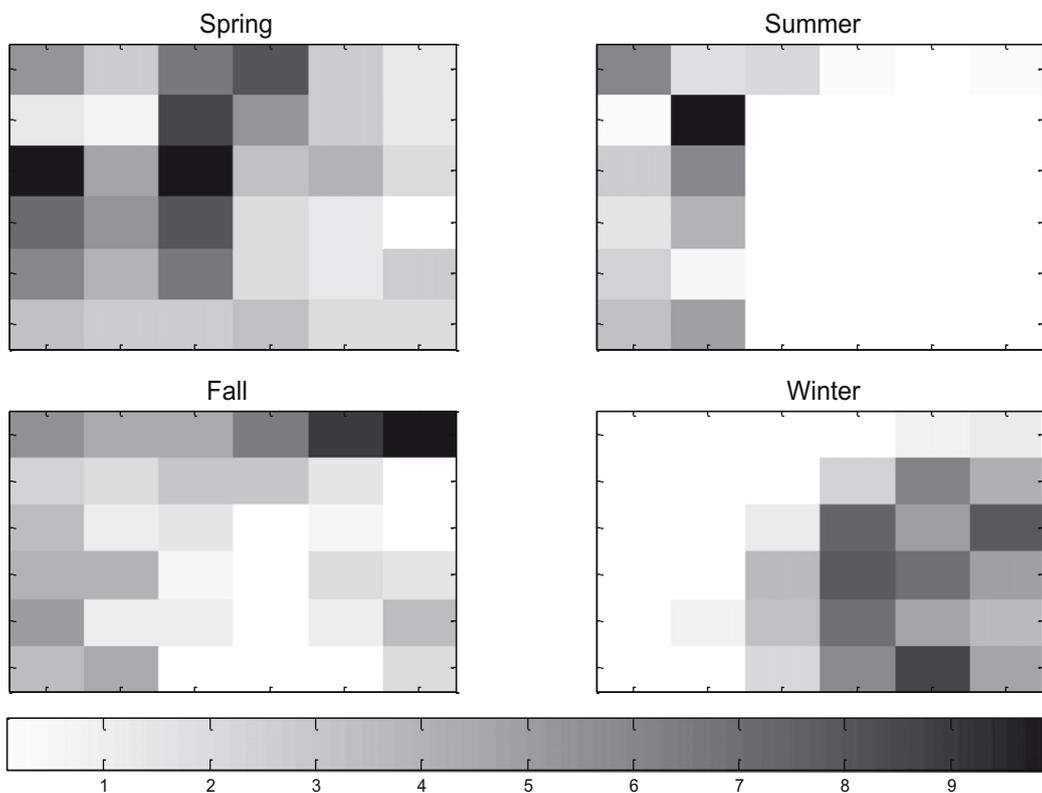


Fig. 8. The frequency of the input patterns in the 6 × 6 topology map.

Table 3
The monthly cumulative errors (mm/month).

mm/month	SOMN	BPNN	Modified Penman	Penman–Monteith
2004/2 (winter)	−1.95	−6.21	13.13	−21.29
2004/5 (spring)	−6.20	−9.10	15.02	−14.84
2004/7 (summer)	8.89	7.71	35.98	11.00
2002/10 (fall)	−5.95	−9.02	2.15	−30.99
MAE	5.78	8.01	16.6	19.5

Table 4
The daily, cumulative monthly and yearly estimating errors (mm; in years of 2001, 2005, and 2006).

Model	SOMN	BPNN	Modified Penman	Penman–Monteith
<i>Daily</i>				
RMSE	1.34	1.33	1.44	1.41
MAE	1.01	1.02	1.12	1.09
<i>Monthly</i>				
RMSE	14.4	13.9	20.3	20.6
MAE	11.4	10.8	17.4	16.7
<i>Yearly</i>				
RMSE	105	108	209	184
MAE	88	92	192	160

The daily statistical parameters of the climatic data are given in Table 1. Being located in a coastal area, the weather station is of moderately to highly humid condition with relative humidity reaches more than 90% in some days and pressure shows significantly high variation. The evaporation losses in site are moderately high due to high temperature. The annual mean temperature is 25.3 °C, where July has the highest temperature (mean 28.6 °C) and January has the lowest temperature (20.7 °C). The relative humidity is also high in summer, about 80%, and low in winter, approximately 65%. The average wind velocity is 3.52 m/s with a high value during winter. The average sunshine hours are 6.51 h. The yearly evaporation rate is approximately 1600 mm.

Models evaluations

The results of SOMN, BPNN, and empirical formulas are summarized in Table 2. The results show that (1) the SOMN and BPNN could adequately produce suitable estimation of evaporation with

relatively small mean error; (2) the performance of the BPNN with only two hidden nodes is slightly behind those of the SOM networks in the training and testing sets but slightly better in the validation set, and (3) the neural networks (BPNN and SOMN) have better performance than the empirical formulas do. As we investigate the model performances, the mean square error (MSE) values of SOMN are 1.048, 1.121, and 1.162 for the training, validation and testing cases, respectively, while the MSE values of Penman–Monteith formula are 1.137, 1.252, and 1.307 for the training, validation and testing cases, respectively. It appears that the SOMN is comparable to the BPNN, and the SOMN has much better performance (about 20% improvement) than the empirical formulas do.

SOM features

As mentioned above, one of the most significant characteristics and contributions of using SOM is that the results obtained by SOM can be visualized from its topology map. Once the SOM has converged, it stores the most relevant information about the process in its topology map and allows all such information to be displayed. Fig. 3 presents the SOM topology maps, which have 6 × 6 grids (neurons) with each neuron represents a cluster of similar input patterns. Each neuron includes five meteorological variables whose normalized mean value is represented as the center of each neuron. The topology map displays the general trend of meteorological factors within the 6 × 6 grid. For example, the temperature in the left side is greater than that in the right side, and the radiation on the top row is much higher than that on the bottom row. We could also easily identify the differences in various nodes, especially the differences among the four corners of the map, whereas the nearby grids are similar in each corner. We demonstrate the feature map is topologically ordered in the sense that the spatial location of a neuron in the lattice corresponds to a particular feature of input patterns.

To more clearly distinguish the 6 × 6 SOM clustering results for each variable and interpolate its response, the center (mean) values of the variables in each node of the topology map are calculated and listed. The six values of each variable in each row are then connected and shown in Fig. 4. We can easily identify that the values of the upper rows (nodes) are generally greater than those of the lower rows (nodes) for the variables of temperature (T), sunshine hours (S), solar radiation (SR), and evaporation (E), while the relative humidity (H) has the inverse phenomenon and the lines of wind speed (W) are intersected. The information

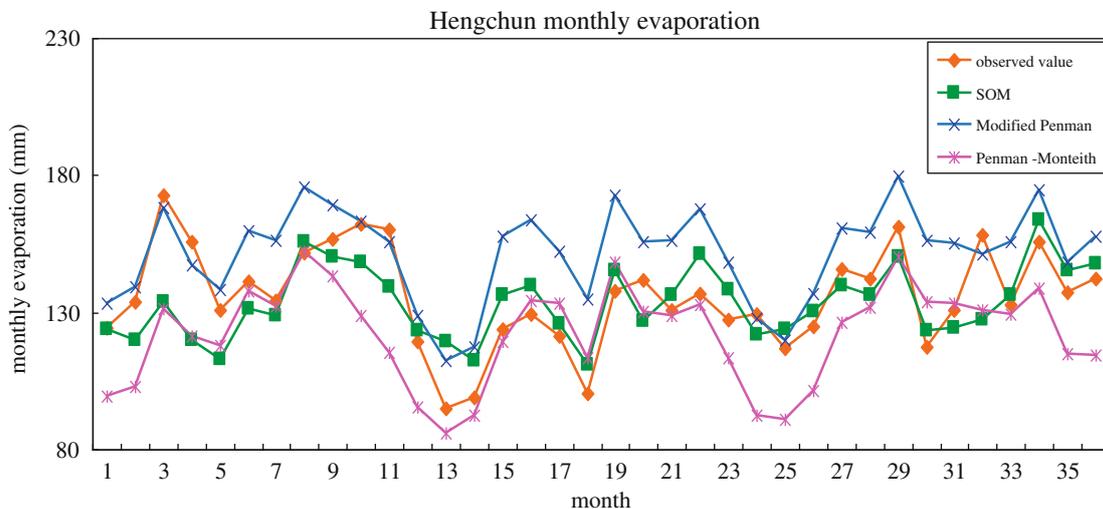


Fig. 9. The monthly evaporations in the years of 2001, 2005 and 2006.

presented in Fig. 4 is interesting and can be further interpreted. As we check the correlation coefficients (CC) between evaporation with the meteorological variables listed in Table 1, the SOM clustering results are highly consistent with the correlation coefficients, where for positive CC, such as T and S, the values of connected lines are decreased from top to bottom; and for negative CC, such as H, the values of the lines are increasing from top to bottom. For the small CC, such as wind speed, the lines are intersected. The mean evaporations of each node in the SOM are also presented in the Fig. 4. It shows that the values of the upper rows (nodes) are generally greater than those of the lower rows (nodes).

Synoptic clustered meteorological patterns

To provide a clear image regarding the SOM synoptic clustered meteorological patterns, six 6×6 SOM of daily meteorological images are shown in Fig. 5. We can easily find that there is a continuum of change across the SOM array for each of the six meteorological variables, for instance, with warm patterns on the upper-left and cool patterns on the lower right; with more sunshine hour and solar radiation patterns on the upper and less sunshine hour and solar radiation patterns on the bottom. Superimposed on these SOM patterns of input meteorological variables, i.e. temperature, humidity, wind speed, sunshine hour, and solar radiation with the output pattern (evaporation), we could capture their relations and multi-variability. This representation not only highlights the topology maps of each input meteorological factor, but reveals their co-variability with the evaporation. Thus, as depicted in Fig. 5, the clustering results of the meteorological variables in the map provide a good approximation and display important and meaningful relationships of the input–output patterns.

Fig. 6 presents the correspondent outputs (evaporations) to the set of input variables in each hidden node, where the longitudinal coordinate represents the number of times (days) and the latitudinal coordinate represents the values of evaporation. We could find that most of the outputs (evaporations) of the nodes in the upper-right corner are greater than 5 (mm/day), while large portions of outputs of the nodes in the low-left corner are less than 5 (mm/day). The widespread distribution of the evaporation in each node

might be a reason why the estimation error (MAE) could not be small (around 0.85–1.1 mm/day). This is mainly because similar input condition (same clustered node) could have very different output value.

Fig. 7 gives the estimated error distributions in six different evaporation ranges. As the evaporation is less than 3 mm/day (lower evaporation), the network would constantly provide higher evaporation and its estimation errors are constantly greater than zero; while the evaporation is larger than 6.1 mm/day (higher evaporation), the network would provide lower evaporation and its estimation errors are tended to be negative values. When the evaporation is falling within the range of 3–6 mm/day, the estimated error distributions are likely normal-distributed within the range of -1.7 to 1.7 mm/day. For a number of abnormal events whose evaporations are either very high (e.g., 8.1–9.5 mm/day) or very low (e.g., 0.8–1 mm/day), the model could provide notable estimating surplus or shortage to adjust the abnormal conditions. The results suggest that the model would adjust the evaporation when it is much higher, or lower, than the average. The adjustment is solely based on the relative meteorological variables used in building the SOMN.

It is interesting to learn whether the SOM is capable to identify the seasonal meteorology. To visualize the seasonal effect on the evaporation, we count the number of times (days) clustered in each node, which could display the lumped effect of input meteorological variables, from four different seasons. The degree of darkness of a node in the map represents the relative frequency of the events (days) clustered in that category. The frequency of occurrence of each season is shown in Fig. 8 to highlight seasonal differences. Some individual patterns are noteworthy: (1) most of the summer days are clustered on the left side; (2) in contrast, most of the winter days are sorted on the right side; and (3) the tendencies of the clustering maps of spring and fall days are relatively fuzzy. These patterns represent given data from input meteorological conditions, the SOM is able to select a set of best features for approximating the underlying distribution.

To learn the accuracy of the constructed models in different seasons and years, the cumulated errors of monthly evaporation rates were made for 4 months in testing cases as shown in Table 3. The results show that (1) the SOMN provides the smallest estimation errors of cumulated monthly evaporation in all the months (four

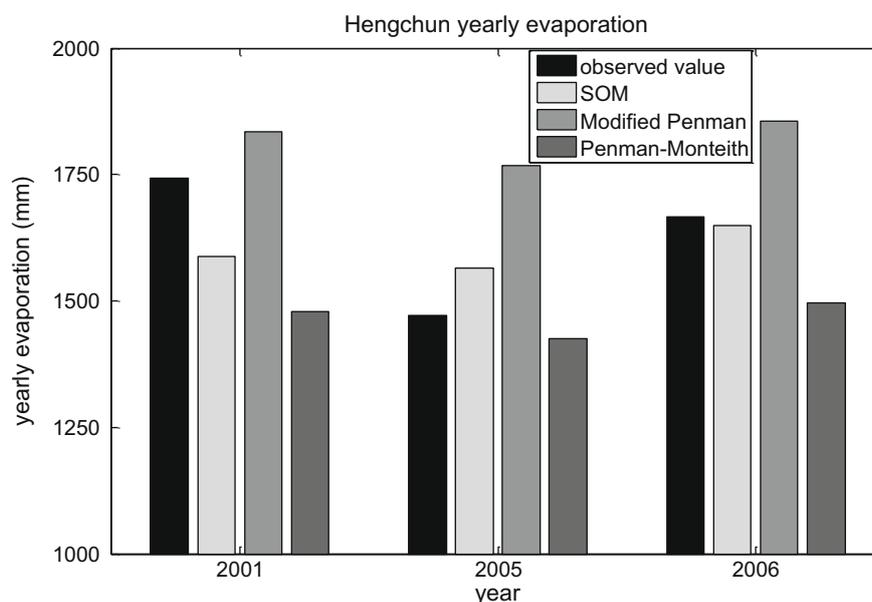


Fig. 10. The observed and estimated evaporations in 3 years.

seasons); (2) the modified Penman tends to over-estimate the evaporation rate, while the Penman–Monteith is likely to under-estimate the evaporation. The mean absolute errors (MAE) of the SOMN, BPNN, the modified Penman and Penman–Monteith methods are 5.78 mm/month, 8.01 mm/month, 16.6 mm/month and 19.5 mm/month, respectively, where the SOMN provides much smaller error than those two empirical methods do. The reliabilities of these models are further evaluated by using the data sets in the years of 2001, 2005, and 2006. These data sets are never involved in a model's construction. The root mean square error (RMSE) and mean absolute error (MAE) of daily, monthly and yearly cumulative evaporation by the SOMN, BPNN, and two empirical methods are presented in Table 4. The results of daily evaporation show that the models' performances in these three evaluated years are quite similar to those in constructed years (2002, 2003, and 2004), and the reliability and consistency of the constructed models are confirmed. We can also easily find that the constructed SOMN does provide much better performance, in terms of small values of RMSE and MAE in daily, cumulated monthly and yearly evaporation, than the two empirical methods do. In 2001, 2005, and 2006, their cumulated monthly evaporation by the constructed methods is shown as a continuous monthly time series in Fig. 9, and their values of cumulated yearly evaporation are shown in Fig. 10. Again, the results demonstrate that the SOMN well simulated the observed values, the values of modified Penman seem appear on the upper bound, and the values of Penman–Monteith are likely in the lower bound.

Conclusions

The study presented a self-organizing map network (SOMN) for assessing the variability of daily evaporation based on meteorological variables. The daily climatic data sets (air temperature, solar radiation, wind speed, sunshine hour and humidity) of a weather station in the southeast part of Taiwan from 2001 to 2006 were collected for model construction. To assess the reliability and stability of SOMN, BPNN and two commonly used evaporation-estimating equations, i.e., the modified Penman and Penman–Monteith methods, comparisons were made according to the various statistic measures and data sets. The results show that the constructed SOMN is slightly better than the commonly used BPNN and performs much better than the modified Penman and Penman–Monteith equations do. The daily mean absolute error in the testing periods obtained by SOMN is 0.881 mm/day and by BPNN is 0.891 mm/day, while those by using the FAO-modified Penman and Penman–Monteith methods are 1.035 mm/day and 1.014 mm/day, respectively. A further testing case was conducted by using the constructed models with the data sets, which are not used in model constructing processes, in three different years (2001, 2005, and 2006). The results give clear evidence that the SOMN through training with input–output examples provides an easy and effective method to new sets of climatic conditions and yields more accurate estimation of evaporation than traditional estimating equations do. The study demonstrated that modeling of daily evaporation is possible through the use of network data driven techniques, such as SOMN and BPNN, from easily available meteorological variables. We also want to note that the FAO-modified Penman equation, which has gained acceptance as the standard method of estimating reference crop evapotranspiration, was used without calibration for estimating pan evaporation, while the used ANN-based models were sophisticatedly trained and verified by the observed data.

SOMN can establish the relationships among input meteorological variables through representing regions of high signal density on the topological map and preserving neighborhood relations in

the input data. We could visually identify clusters on the 2D topology map and gain ideas through the behavior of the input variables and the multi-relations among variables. For example, in this study we find the temperature in the left side is greater than that in the right hand side, and the radiation on the top row is much higher than that on the bottom row. We could also easily identify the differences in various nodes, especially the differences among the four corners of the map, whereas the nearby grids are similar in each corner. The frequency of occurrence of each season highlights seasonal differences, i.e. most of the summer days are clustered on the left side, while most of the winter days are sorted on the right side. These patterns represent the SOMN can select a set of best features for approximating the underlying distribution. The SOMN fuses the SOMN for identifying the cluster centers of topology map and the least square regression technique for optimizing the connected weights between the centers and the output node. We demonstrate that the SOMN is a systematic and efficient way of extracting the multi-collinear relationship between evaporation and meteorological variables, and suitably and reliably formulates the meteorological input–output patterns for evaporation estimation.

Acknowledgements

This paper is based on partial work supported by National Science Council, Taiwan, ROC (Grant No. NSC 97-2313-B-002-013-MY3). The authors are grateful to the editors and anonymous reviewers for their valuable comments and suggestions.

References

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper, 56.
- Barhak, J., Fischer, A., 2002. Adaptive reconstruction of freeform objects with 3D SOM neural network grids. Computers & Graphics – UK 26 (5). PII S0097-8493(02)00129-2.
- Bruton, J.M., Mcclendon, R.W., Hoogenboom, G., 2000. Estimating daily pan evaporation with artificial neural networks. Transactions of the ASAE 43 (2), 491–496.
- Burman, R.D., 1977. Intercontinental comparison of evaporation estimates. Journal of the Irrigation and Drainage Division – ASCE 103 (3), 381.
- Cancelliere, A., Giuliano, G., Ancarani, A., Rossi, G., 2002. A neural networks approach for deriving irrigation reservoir operating rules. Water Resources Management 16 (1), 71–88.
- Chang, F.J., Chang, L.C., Wang, Y.S., 2007. Enforced self-organizing map neural networks for river flood forecasting. Hydrological Processes 21 (6), 741–749.
- Chang, L.C., Chang, F.J., 2001. Intelligent control for modeling of real time reservoir operation. Hydrological Processes 15 (9), 1621–1634.
- Chau, K.W., 2006. Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. Journal of Hydrology 329 (3–4), 363–367.
- Chen, Y.H., Chang, F.J., 2009. Evolutionary artificial neural networks for hydrological systems forecasting. Journal of Hydrology 367, 125–137.
- Cheng, C.T., Xie, J.X., Chau, K.W., Layeghifard, M., 2008. A new multi-step-ahead model for long-term hydrologic prediction. Journal of Hydrology 361 (1–2), 118–130.
- Chiang, Y.M., Chang, L.C., Chang, F.J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modeling. Journal of Hydrology 290, 297–311.
- Cigizoglu, H.K., Kisi, O., 2005. Flow prediction by three back propagation techniques using *k*-fold partitioning of neural network training data. Nordic Hydrology 36 (1), 49–64.
- Coulomb, C.V., Legesse, D., Gassea, F., Travic, Y., Chernetd, T., 2001. Lake evaporation estimates in tropical Africa (Lake Ziway, Ethiopia). Hydrological Processes 245 (1–4), 1–18.
- Gavin, H., Agnew, C.A., 2004. Modelling actual, reference and equilibrium evaporation from a temperate wet grassland. Hydrological Processes 18 (2), 229–246.
- Grieco, S., Thiery, F., Traore, A., Nguyen, T.P., Barreau, M., Polit, M., 2006. KSOM and MLP neural networks for on-line estimating the efficiency of an activated sludge process. Chemical Engineering Journal 116 (1), 1–11.
- Hsu, K.L., Gupta, H.V., Gao, X., Sorooshian, S., Imam, B., 2002. Self-organizing linear output map (SOLO): an artificial neural network suitable for hydrologic modeling and analysis. Water Resources Research 38 (12), 10–26.
- Keskin, M.E., Terzi, O., 2006a. Artificial neural network models of daily pan evaporation. Journal of Hydrologic Engineering 11 (1), 65–70.

- Keskin, M.E., Terzi, O., 2006b. Evaporation estimation models for Lake Egirdir, Turkey. *Hydrological Processes* 20 (11), 2381–2391.
- Kim, S., Kim, H.S., 2008. Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modeling. *Journal of Hydrology* 351 (3–4), 299–317.
- Kisi, O., 2005. Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* 50 (4), 683–696.
- Kisi, O., 2006. Daily pan evaporation modelling using a neuro-fuzzy computing technique. *Journal of Hydrology* 329 (3–4), 636–646.
- Kisi, O., Ozturk, O., 2007. Adaptive neurofuzzy computing technique for evapotranspiration estimation. *Journal of the Irrigation and Drainage Division – ASCE* 133 (4), 368–379.
- Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43 (1), 59–69.
- Kohonen, T., 1990. The self-organizing map. *Proceedings of the IEEE* 78 (9), 1464–1480.
- Kumar, D.N., Srinivasa, R.K., Sathish, T., 2004. River flow forecasting using recurrent neural networks. *Water Resources Management* 18 (2), 143–161.
- Odhiambo, L.O., Yoder, R.E., Yoder, D.C., Hines, J.W., 2001. Optimization of fuzzy evapotranspiration model through neural training with input–output examples. *American Society of Agricultural Engineers* 44 (6), 1625–1633.
- Parasuraman, K., Elshorbagy, A., Carey, S.K., 2007. Modelling the dynamics of the evapotranspiration process using genetic programming. *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* 52 (3), 563–578.
- Regis, C., Frederic, S., Arthur, C., Sylvain, M., 2005. Using self-organizing maps to investigate spatial patterns of non-native species. *Biological Conservation* 125 (4), 459–465.
- Richardson, A.J., Risien, C., Shillington, F.A., 2003. Using self-organizing maps to identify patterns in satellite imagery. *Progress in Oceanography* 59 (2–3), 223–239.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323 (6088), 533–536.
- Supharatid, S., 2003. Tidal level forecasting and filtering by neural network model. *Coastal Engineering Journal* 45 (1), 119–138.
- Sudheer, K.P., Gosain, A.K., Ramasastri, K.S., 2003. Estimating actual evapotranspiration from limited climatic data using neural computing technique. *Journal of the Irrigation and Drainage Division – ASCE* 129 (3), 214–218.
- Sudheer, K.P., Gosain, A.K., Rangan, D.M., Saheb, S.M., 2002. Modelling evaporation using an artificial neural network algorithm. *Hydrological Processes* 16 (16), 3189–3202.
- Szilagyi, J., Jozsa, J., 2009. Analytical solution of the coupled 2-D turbulent heat and vapor transport equations and the complementary relationship of evaporation. *Journal of Hydrology* 37 (1–4), 61–67.
- Tadeusz, P., Andrzej, K., Maria, G., Małgorzata, D., 2006. Patterning of impoundment impact on chironomid assemblages and their environment with use of the self-organizing map (SOM). *Acta Oecologica* 30 (3), 312–321.
- Tayfur, G., 2002. Artificial neural networks for sheet sediment transport. *Hydrological Sciences Journal – Journal Des Sciences Hydrologiques* 47 (6), 879–892.
- Trajkovic, S., Todorovic, B., Stankovic, M., 2003. Forecasting of reference evapotranspiration by artificial neural networks. *Journal of the Irrigation and Drainage Division – ASCE* 129 (6), 454–457.
- Warnaka, K., Pochop, L., 1988. Analyses of equations for free water evaporation estimates. *Water Resource Research* 24 (7), 979–984.
- Wu, C.L., Chau, K.W., Li, Y.S., 2008. River flow prediction based on a distributed support vector regression. *Journal of Hydrology* 358 (1–2), 96–111.
- Xu, C.Y., Singh, V.P., 1998. Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrological Processes* 12, 429–442.