Feasibility of bootstrap aggregating to enhance extreme learning machine for reference evapotranspiration estimation

Min Yan Chia¹, Yuk Feng Huang¹*, Chai Hoon Koo¹ and Lik Sheng Lai¹

¹Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Malaysia.

Abstract. Estimation of evapotranspiration (ET) is a challenging, yet important task as the ET value can be used to predict many other natural phenomena. In this work, the reference evapotranspiration (ET₀) was estimated using the extreme learning machine (ELM) at two meteorological stations located in the northern region of the Straits of Malacca. Optimum designs of the ELM were first determined and it was found that the different number of hidden neurons and activation functions were favourable to various input combinations. In order to enhance the performance of the ELM, the bootstrap aggregating algorithm was integrated to resample the original dataset. However, the performance of bagged-ELM was found to be poorer than the base ELM. This could be attributed to the high stability of the base ELM model whereby the training size already overwhelmed the dimensionality of the problem itself. The bootstrap aggregation data fusion technique produced a “backfire” effect that degraded the accuracy and generalisability of the base ELM model.

1 Introduction

Evapotranspiration (ET) is an essential process where water is lost to the atmosphere through the combined effects of evaporation (from the surface of Earth) and transpiration (from the surface of vegetation). This process is vital in the regulation of the energy and water budgets, which subsequently will be affecting many other natural phenomena [1]. As such, accurate measurement, and even prediction of the ET become an important topic to researchers all around the world, especially in regions that are vulnerable to drought and are highly dependent on the crop yield. The lysimetric measurement is a common way of obtaining real time ET value, but however, this method is highly constrained by the lack of spatial coverage, tedious to carry and not to mention the accompanying financial burden and ecological footprint. Therefore, numerous empirical models have been developed over the years, in an attempt to replace the lysimetric measurement, such as the Penman-Monteith (PM) equation endorsed by the Food and Agriculture Organisation of the United Nations, Hargreaves-Samani equation, Turc equation and so forth [2].

*Corresponding author: huangyf@utar.edu.my
Nevertheless, the utilisation of such empirical models is problematic, as many meteorological parameters have to be collected upfront. Hence, alternative tools are being explored continuously. In recent years, machine learning appears in the spotlight due to their efficiency in performing prediction whilst having highly flexible internal structures [3]. Machine learning models, including the artificial neural network (ANN), support vector machine (SVM) and fuzzy based models have been developed and showed promising performance in estimating/predicting the ET related metrics, such as the reference evapotranspiration ($ET_0$) [4-6].

Huang, Zhu and Siew [7] developed a novel variant of the ANN known as the extreme learning machine (ELM) in 2006. In general, the ELM has a structure of single-layered neural network, whereby the input weights and biases are generated randomly at the initialisation. Thereafter, the output weights are solved analytically via simple generalised inverse operation. The algorithm of the ELM is simple enough to render fast computation, alongside with good generalisation. Several studies also had proven the efficacy of the ELM in estimating the $ET_0$ [8, 9]. However, the lack of stochastic training of the ELM model would lead to the possible problem of convergence to the local optima. Owing to this issue, integrating the ELM with data fusion techniques can be a possible solution to this problem.

One of the most commonly applied data fusion techniques is the bootstrap aggregating, also known as the bagging method [10]. By using this technique, the original dataset is resampled into several bags of equal size, in order to increase the number of samples to train the machine learning model. Estimations by models trained using different bags are aggregated to obtain the average estimation. The intention to perform bootstrap aggregating is to change the structure of the original dataset, so that the resultant estimation or prediction is less bias and more accurate. In other words, the bootstrap aggregating is a data centric approach where the performance of the model is enhanced by modifying the structure of available data.

In this study, the ELM is first optimised in terms of number of hidden neurons as well as the most suitable kernel function. Then, the bootstrapped data are used to train the ELM to observe the feasibility of the data fusion technique in improving the performance of the ELM. The effects of various input combinations of meteorological parameters are also evaluated in this work.

2 Methods

2.1 Study Area and Data

Two meteorological stations, located at two separate islands of Malaysia, in the northern region of the Straits of Malacca are included for this case study. The two stations are the Station 48600 (Pulau Langkawi) and Station 48601 (Bayan Lepas) and they are located in a region with typical tropical climate. Table 1 shows the geographical details of the selected stations. The exact locations of the two meteorological stations are presented in Figure 1. Daily meteorological data dated from January 2014 to December 2018 were acquired from the Malaysia Meteorological Department (MMD). These meteorological parameters include the maximum, minimum and mean temperature ($T_{max}$, $T_{min}$ and $T_{mean}$), relative humidity ($RH$), wind speed ($u$) and global radiation ($R_s$).
2.2 Penman-Monteith Equation

The PM equation is being regarded as the standard for ET$_0$ calculation. Therefore, the meteorological data were utilised to calculate the ET$_0$ using the PM equation, as shown in Equation 1.

$$ET_0 = \frac{0.408\Delta (R_n - G) + \gamma \frac{900}{\text{temp}} U_2 (e_s - e_a)}{\Delta + \gamma (1 + 0.34 U_2)}$$  \hspace{1cm} (1)

where $R_n$ is net radiation (MJ m$^{-2}$ day$^{-1}$), $G$ is soil heat flux (MJ m$^{-2}$ 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), $\Delta$ is slope of vapour pressure curve (kPa°C) and $\gamma$ is psychrometric constant.

2.3 Extreme Learning Machine

The ELM has a structure of a feedforward neural network with only one hidden layer. An activation function is used to transform the input into the random feature space of the ELM. The estimation of the ELM can be calculated using Equation 2.

$$f_L(x) = \sum_{i=1}^{L} h_i \beta_i$$  \hspace{1cm} (2)

where $L$ is the number of hidden neurons, $h$ is the hidden neuron output weight and $\beta$ is the output weight. The output weight matrix is obtained by solving Equation 3.
\[ H = \beta T \]  

where \( H \) is the vector for the output of hidden layer, \( \beta \) is their corresponding weight vector and \( T \) is the target vector. Several activation functions were evaluated to determine their suitability for \( ET_0 \) estimation under different situations. The activation functions used in this study include the sigmoid function, sine function and radial basis function (RBF), as shown in Equation 4 to Equation 6, respectively.

\[ f(x) = \frac{1}{1+e^{-x}} \]  
\[ f(x) = \sin(x) \]  
\[ f(x) = g(b_i||x - a_i||) \]

where \( x \) is the input, \( a \) is the centre of RBF node, and \( b \) is impact factor of RBF node.

### 2.4 Bootstrap Aggregating

The algorithm of bootstrap aggregating (or bagging) involves the resampling of data from the original data set. In this study, 10 bags were resampled from the daily meteorological data, and 10 corresponding ELMs were trained using the data in each bag. The estimation of the 10 ELMs were aggregated via simple averaging. The overall mathematical expression of bootstrap aggregating is shown in Equation 7.

\[ f_{bagging}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x) \]

where \( B \) is the number of bags (10 in this study).

### 2.5 Data Pre-Processing

Before performing the training, the data have to be normalised in order to remove the absolute scale effect, using Equation 8.

\[ x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \]

where \( x \), \( x_{max} \) and \( x_{min} \) are the input, maximum value of input dataset and minimum value of input dataset, respectively. Apart from that, the k-fold cross validation was also performed to validate the estimation of the ELM, whereby a 10-fold validation technique was used in this study.

### 2.6 Performance Evaluation

The performance of the ELM and the bagged ELM models were evaluated from different aspects. The mean absolute error (MAE) was used to assess the accuracy of the model, together with the root mean square error (RMSE) which can detect large errors. The goodness-of-fit of the model was measured by the Nash-Sutcliffe efficiency (NSE) while the generalisation of the ELM was evaluated using the mean bias error (MBE). The equations of calculating the MAE, RMSE, NSE and MBE are shown in Equation 9 to Equation 12.
where $N$ is the number of observations, $y_{\text{predicted}}$ is the predicted ET$_0$, $y_{\text{actual}}$ is the actual ET$_0$ and $\bar{y}_{\text{actual}}$ is the mean of the actual ET$_0$.

### 3 Results and Discussion

#### 3.1 Optimum Design of ELM and Preferred Input Combinations

The six meteorological parameters were used to formulate 63 resulting input combinations, which were used separately to train the ELM of different activation functions. At the same time, by using the grid search method, the number of neurons in the hidden layer of the ELM was determined. The assessment of the performance of the ELM was done by comparing the values of the performance evaluation metrics. The optimum design of the base ELM, as well as the preferred input combination (with different numbers of meteorological parameters input) are summarised in Table 2.

<table>
<thead>
<tr>
<th>Station</th>
<th>Number of Meteorological Parameters</th>
<th>Combination of Meteorological Parameters</th>
<th>Number of Hidden Neurons</th>
<th>Activation Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station 48600 (Pulau Langkawi)</td>
<td>6</td>
<td>$T_{\text{max}}, T_{\text{min}}, T_{\text{mean}}, RH, u, R_s$</td>
<td>22</td>
<td>Sigmoid</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$T_{\text{max}}, T_{\text{min}}, RH, u, R_s$</td>
<td>32</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$T_{\text{max}}, RH, u, R_s$</td>
<td>32</td>
<td>RBF</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$T_{\text{max}}, u, R_s$</td>
<td>40</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$u, R_s$</td>
<td>42</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>RH</td>
<td>11</td>
<td>RBF</td>
</tr>
<tr>
<td>Station 48601 (Bayan Lepas)</td>
<td>6</td>
<td>$T_{\text{max}}, T_{\text{min}}, T_{\text{mean}}, RH, u, R_s$</td>
<td>42</td>
<td>Sigmoid</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>$T_{\text{max}}, T_{\text{mean}}, RH, u, R_s$</td>
<td>35</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>$T_{\text{max}}, RH, u, R_s$</td>
<td>24</td>
<td>Sigmoid</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>$RH, u, R_s$</td>
<td>13</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>$RH, R_s$</td>
<td>50</td>
<td>Sine</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>$R_s$</td>
<td>7</td>
<td>Sigmoid</td>
</tr>
</tbody>
</table>
As suggested from observations in Table 2, the ELM trained at the Station 48600 (Pulau Langkawi) and Station 48601 (Bayan Lepas) behaved dissimilarly even though their geographical characteristics are almost identical. At Station 48600 (Pulau Langkawi) $T_{mean}$, $T_{min}$, $RH$ and $T_{max}$ were discounted in sequence when the number of meteorological parameters reduced from six to two. However, for the input combination with one meteorological parameter, the ELM selected $RH$ once more while disregarding $R_s$ and $T_{max}$. The $RH$ appeared to be the most important meteorological parameter at Station 48600 (Pulau Langkawi), despite its temporary absence for the input combination with two and three meteorological parameters. As for the preferred activation functions, the sigmoid function was favoured for the six-parameter input combinations. The sine function was favoured for input combinations with five, three and two meteorological parameters whereas the RBF function was chosen for the remaining input combinations.

On the other hand, at Station 48601 (Bayan Lepas), the preferred input combinations were slightly different from Station 48600 (Pulau Langkawi). At Station 48601 (Bayan Lepas), the $T_{min}$, $T_{mean}$, $T_{max}$, $u$ and $RH$ were removed sequentially as the number of input meteorological parameters was reduced. The sigmoid function was selected to deal with input combinations with six, four and one meteorological parameters at this round, and the remaining input combinations performed better when the sine function was used. It is interesting to note that the preferred activation function and number of hidden neurons did not have any noticeable relationship with any other variables. Therefore, when estimating the $ET_0$ using the ELM, it is important to perform a preliminary screening to determine the optimum number of neurons in the hidden layer.

### 3.2 Effect of Bootstrap Aggregating Integration

The base ELMs with optimum designs and preferred input combinations were trained with bootstrapped data. The original dataset was resampled into 10 bags of equal size, which meant that 10 ELMs would be produced for a given input combinations and aggregated via simple averaging to form the bagged-ELM. The performance of the bagged-ELM is contrasted with that of base ELM, as shown in Table 3.

<table>
<thead>
<tr>
<th>Station</th>
<th>Number of Meteorological Parameters</th>
<th>ELM</th>
<th>Bagged-ELM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MA E</td>
<td>RMS E</td>
</tr>
<tr>
<td>Station 48600 (Pulau Langkawi)</td>
<td>6</td>
<td>0.18</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.19</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.25</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.29</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.65</td>
<td>0.81</td>
</tr>
<tr>
<td>Station 48601</td>
<td>6</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.14</td>
<td>0.16</td>
</tr>
</tbody>
</table>
It is clear that as the number of input meteorological parameters decreased, the performance of the base ELM and the bagged-ELM deteriorated, as portrayed by the increment in MAE, RMSE and MBE along with the reduction in NSE. This finding is reasonable as the lesser number of inputs would mean that the information that can be extracted by the base ELM and bagged ELM were lesser. The base ELM in Station 48600 (Pulau Langkawi) was more susceptible to the removal of meteorological parameters from the input, as shown by the drastic decline in the performance evaluation metrics, especially when only one meteorological parameter was fed into the models.

Most importantly, it can be seen that actually the integration of the bootstrap aggregating did not enhance the performance of the base ELM. In fact, the bagged-ELM had poorer performance as compared to the base ELM. The bagged-ELM registered higher MAE, RMSE and MBE as well as lower NSE as compared to the base ELM at both stations, which put paid to the initial intention of performing bootstrap aggregating. According to Breiman [10], the effectiveness of the bootstrap aggregating is dependent on the stability of the model. Performing bootstrap aggregating on a stable model could lead to a backfire effect. A model is deemed unstable if the size of training data is comparable to the dimensionality of the problem [11]. In this study, the size of the original dataset (1826 data points) overwhelmed the dimensionality of the problem (at most six). Hence, in our case the base ELM is already stable and the data centric bootstrap aggregating is therefore not further helpful in this case.

4 Conclusion

The base ELM was developed at Station 48600 (Pulau Langkawi) and Station 48601 (Bayan Lepas) to estimate \( \text{ET}_0 \) with different number of input meteorological parameters. The optimum designs of the base ELM varied with the location and input combinations. These include the number of hidden neurons as well as the activation functions chosen. In was observed that integration of the bootstrap aggregating (bagging) onto the base ELM to form the bagged-ELM, however, deteriorated the performance. This is due to the fact that the available training data in this research work is very large as compared to the dimensionality of the problem. Hence, the data centric bootstrap aggregating is not required to resample the original dataset. Since modification of the structure of original dataset could not improve the performance of the ELM, hence, future studies should focus on the tuning of the internal parameters as well as training algorithm of the ELM to enhance its accuracy and generalisability.

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

References