

Suspended Sediment Forecast of Khlong Bang Yai, Phuket

Apichat Heednacram and Thammaratt Samitalampa

Abstract—This paper studies the accumulation of sediments in Bang Yai Canal, Phuket. The increase in sediment level decreases the water storage capacity of the canal which leads to shallow flooding problem. The study clustered the sediment into 3 levels using *k*-means algorithm, and then utilized an Artificial Neural Network to forecast the level of the sediment for the next three years using multilayer perceptron as a predictor. An efficient schedule for dredging the sediment in Bang Yai Canal can be constructed from the results of the study. Suspended sediments in tons per day from January 01, 2007 until December 31, 2011 were used as an input to our experiment. The result shows that the deposit of suspended sediments is peaked in October and is low in January of every year for the past five years. The level of sediments also increases every year for the next three years. We recommend that the municipality may alleviate the problem by dredging the sediments prior October, so that the canal drainage capacity is maintained.

Index Terms—Data Mining, *k*-means algorithm, multilayer perceptron, sediment forecast.

I. INTRODUCTION

Flooding is one of the major problems occurring around the world. The geography of Thailand is mainly lowland and also often has flooding [1]. The flood in 2011 was a severe one causing a wide range of damages. Many of canals in Thailand are bottle-like allowing water to drain slowly into the sea [2]. Phuket Island in Thailand frequently has flooding problem. It is very crucial that Phuket acquire related information in order to better manage this problem.

The Bang Yai Canal (“Khlong Bang Yai”) in Phuket connects Kathu Waterfall to the Andaman Sea, flowing through Kathu Municipality. The increasing sediment depositions in the canal from rainfall, from river bank slumping, and from construction in the area around the canal, are the major causes of soil erosion in the watershed area. Sediment carried in the canal water, increasingly accumulates in the canal over time. This decreases the water storage capacity of the canal, which leads to flooding.

Artificial Neural Networks (ANNs) or Neural Networks (NN) have been used widely in forecasting, particularly for the hydrological modeling problem. Chaowanawatee and Heednacram [3], [4] proposed Radial Basis Function (RBF) Artificial Neural Networks to calculate and forecast flood water levels. Chantasut, Charoenjit, and Tenprasert [5]

utilized ANNs for monthly rainfall predictions along the Chao Phraya River. The results showed that it is possible to predict annual rainfall one year ahead with acceptably accuracy.

The study of sediment in Bang Yai Canal for water resource management provided essential information to plan for sediment dredging in the canal. This study clustered the sediment into 3 levels, and then utilized an Artificial Neural Network to forecast the level of the sediment for the next 36 months. An efficient schedule for dredging the sediment in Bang Yai Canal can be constructed from the results of the study.

The paper is organized as follows. First, we describe the technical background necessary for understanding the experiments. Then, we explain the settings of our two experiments (clustering the sediments, and forecasting the sediments) following by a discussion on the experimental results. Finally, we summarize our work, and conclude with future research problems.

II. TECHNICAL BACKGROUND

We explain the details of Bang Yai Canal and its topography, then discuss methodologies, algorithms, and evaluation techniques for performance measures.



Fig. 1. Bang Yai canal at Ban Ket Ho, Amphur Kathu, Phuket [6].

A. Details of Bang Yai Canal

Sediment data of *Bang Yai Canal* came from Station X.190A which is on the right bank at the bridge, Ban Ket Ho, Amphur Kathu, Phuket. This sediment station is located at Latitude $8^{\circ} 46' 25''$ North, Longitude $98^{\circ} 23' 36''$ East. The drainage area is 28.92 sq.km.

The source of the water is around Kuan Wah Mountain, near Kamala. The watercourse runs steeply down through the Kathu waterfall at Kathu Village, and down to the provincial water department’s storage lake next to Ket Ho School (another branch runs from Bang Wad dam down to the school). The canal has a more level run downtown from the school, through the center of Old Phuket Town, and out into the Phuket Bay next to the Saphan Hin stadium. Fig. 1 shows Bang Yai Canal. The river has a length of about 16 kilometers. Along the river is a catchment area of approximately 63

Manuscript received November 6, 2013; revised January 8, 2014. This work was supported in part by Department of Computer Engineering, Faculty of Engineering, and Prince of Songkla University.

The authors are with the Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Kathu, Phuket, 83120, Thailand (e-mail: apichat@coe.psu.ac.th, thammaratt@coe.phuket.psu.ac.th).

square kilometers. The physical features of Bang Yai Canal are listed as follows.

- 1) Average water level, height above sea level at the sediment station is 15.00 meters.
- 2) Average water level, height above sea level at source near Kuan Wah Mountain is 40.00 meters.
- 3) Approximate width of the canal is 8.00 meters.
- 4) Nature of the canal bottom soil is clay.

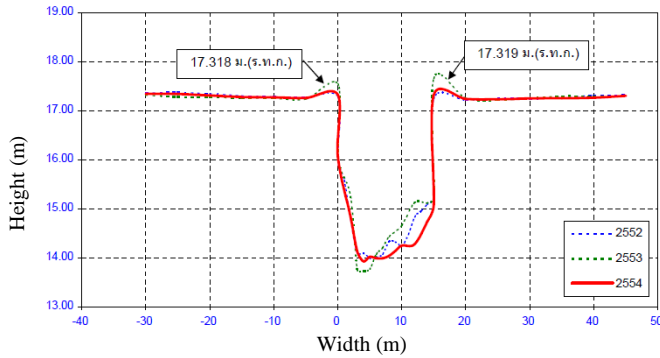


Fig. 2. Cross section of bang Yai canal (station X.109A) [6].

- 5) Cross sectional form of the canal is U-shaped (see Fig. 2).
- 6) Top of the left canal bank is 17.134 meters.
- 7) Top of the right canal bank is 17.231 meters.
- 8) Nature of the left canal embankment is flat ground with weeds.
- 9) Nature of the right canal embankment is level with the neighboring residential area.

B. Data Mining

Data mining is the process of discovering the hidden patterns in data. The process should be automatic or semiautomatic. The patterns discovered should lead to a meaningful economic advantage. Several data mining techniques [7] have been developed and are in common use, including association, classification, clustering, prediction, decision trees, etc. In our research, we focused on the clustering and forecasting data mining techniques.

C. Clustering Algorithm

Clustering is a technique for finding similarity groups in data, called clusters. It groups data instances that are similar to (near) each other into one cluster, and data instances that are very different (far away) from each other into a different cluster. Clustering is one of the most utilized data mining techniques [8]. Clustering algorithms can be divided into two groups: hierarchical and partitional. Hierarchical clustering algorithms recursively find nested clusters either in agglomerative mode (starting with each data point in its own cluster and merging the most similar pair of clusters successively to form a cluster hierarchy) or in divisive (top-down) mode (starting with all the data points in one cluster and recursively dividing each cluster into smaller clusters). Compared to hierarchical clustering algorithms, partitional clustering algorithms find all the clusters simultaneously as a partition of the data, and do not impose a hierarchical structure. The most well-known hierarchical algorithms are single-link and complete-link; the most

popular and the simplest partitional algorithm is *k*-means [9].

The *k*-means algorithm partitions the given data into *k* clusters. Each cluster has a cluster center, called a *centroid*. The user provides the *k* value. The *k*-means algorithm is given in Fig. 3.

```

Algorithm k-means (k,D)
1 Choose k data points as the initial centroids
2 repeat
3   for each data point x ∈ D do
4     compute the distance from x to each centroid
5     assign x to the closest centroid
6   end for
7 re-compute the centroids using the current cluster
  memberships
8 until the stopping criterion is met
    
```

Fig. 3. The *k*-means algorithm [8].

To calculate the mean of a cluster, we need a distance function. The Euclidean and Manhattan (city block) functions are the most commonly used ones [8]. In Euclidean space, the mean of a cluster is computed with:

$$m_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \quad (1)$$

where $|C_j|$ is the number of data points in cluster C_j . The distance from one data point x_i to a centroid m_j is computed with:

$$dist(x_i, m_j) = \sqrt{(x_{i1} - m_{j1})^2 + (x_{i2} - m_{j2})^2 + \dots + (x_{ir} - m_{jr})^2} \quad (2)$$

The equation for a Manhattan function is:

$$dist(x_i, m_j) = |x_{i1} - m_{j1}| + |x_{i2} - m_{j2}| + \dots + |x_{ir} - m_{jr}| \quad (3)$$

D. Multilayer Perceptron

An Artificial Neural Network (ANN) is a type of biologically inspired computational model which has been loosely based on the functioning of the human brain. An ANN performs an input-output mapping using a set of simple processing nodes or neurons where the inputs are drivers to the process, and the output in the case of this research is the sediment level. Each individual neuron integrates information from the model input or from other neurons, and outputs this value using a transfer function. An ANN consists of a series of these neurons arranged in a set of weighted, interconnected layers. Data enter the network through the input units arranged in an input layer. These data are then fed forward through successive layers, including the hidden layer in the middle, and emerge from the output layer.

ANN development involves two main stages: training and testing. In the training or learning stage, the weights between the neurons are adjusted until the network is capable of predicting the desired output [10].

Multilayer perceptron is a neural network that trains using back propagation. Although the back propagation algorithm can be used very generally to train neural networks, it is the most famous for applications to layer feed forward networks,

or multilayer perceptrons. A network with a single layer can approximate any function, if the hidden layer is large enough. This has been proved by a number of people, generally using the Stone-Weierstrass theorem. Thus, multilayer perceptrons are representationally powerful [11]. The back propagation neural network is a multilayered, feed forward neural network and is by far the most extensively used. It is also considered one of the simplest and most general methods used for supervised training of multilayered neural networks. Back propagation works by approximating the non-linear relationship between the input and the output by adjusting the weight values internally. It can further be generalized for the input that is not included in the training patterns [12].

The back propagation network has two stages, training and testing. Fig. 4 shows the topology of the back propagation neural network that includes an input layer, one hidden layer and an output layer. It should be noted that back propagation neural networks can have more than one hidden layer. The detail of back propagation algorithm is given in Fig. 5.

E. Performance Measures

Evaluation is very important to indicate real progress in data mining.

TABLE I: PERFORMANCE MEASURES FOR NUMERIC PREDICTION [7].

Performance Measure	Formula
MAE	$\frac{ p_1 - a_1 + \dots + p_n - a_n }{n}$
MSE	$\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}$
RMSE	$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}$
RAE	$\frac{ p_1 - a_1 + \dots + p_n - a_n }{ a_1 - \bar{a} + \dots + a_n - \bar{a} }$
RRSE	$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}}$

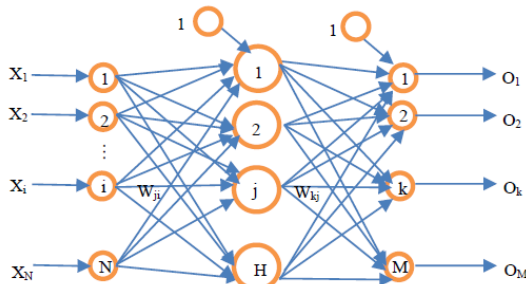


Fig. 4. Back propagation neural network with one hidden layer [12].

Table I shows alternative measures that used to evaluate the success of forecasting and their formula. The predicted values are p_1, p_2, \dots, p_n and the actual values are a_1, a_2, \dots, a_n . The symbol p_i is the numeric value of the prediction for the i -th test instance.

Mean Squared Error (MSE) is the most commonly used measure. MSE measures the average of the square of the error (predict value – actual value). Mean Absolute Error (MAE) is the average of error without taking account of their signs. Root Mean Square Error (RMSE) is the square root of MSE.

RMSE measures the quality of the fit between the actual data and the predicted model. Sometimes relative errors are more importance than absolute error. The square root of the relative squared error is taken to give the same dimensions as the predicted values themselves. Relative absolute error (RAE) is the total absolute error made relative to what the error would have been if the prediction simply had been the average of the actual values. Relative squared error (RSE) is the total squared error and normalizes it by dividing by the total of squared error of the default [7].

```

Assign all network inputs and output
Initialize all weights with small random numbers,
typically between -1 and 1
repeat
  for every pattern in the training set
    Present the pattern to the network
    //Propagate the input forward through the network
    for each layer in the network
      for every node in the layer
        Calculate the weight sum of the inputs
        to the node
        Add the threshold to the sum
        Calculate the activation for the node
      end
    end
    //Propagate the errors backward through the network
    for every node in the output layer
      calculate the error signal
    end
    for all hidden layers
      for every node in the layer
        Calculate the node's signal error
        Update each node's weight in the
        network
      end
    end
  end
  //Calculate Global Error
  Calculate the Error Function
end
while ((maximum number of iterations < than specified)
AND (Error Function is > than specified))

```

Fig. 5. Backpropagation algorithm [12].

III. EXPERIMENTAL RESULTS

In this section, we explain the settings of two experiments which consist of the clustering of the sediments and forecasting of the sediments. Inputs used in the experiment are described. Results are presented and the evaluation of the system is discussed, respectively.

We conduct both experiments on WEKA version 3.7.9 [13]. All experiments are performed on a computer with Intel Core 2 Duo CPU, at 2.20GHz and 2.19GHz with 1.86GB of RAM. We read the input data from an ARFF file which is the file format of WEKA and can be easily imported from EXCEL [14].

Suspended sediments in ton/day [15] for 1826 days from January 01, 2007 until December, 31 2011 were populated into the input file. The example data of year 2007 is given in the Appendix A. Note that data of year 2012 is unavailable for the study. Therefore, based on the training of this five-year data we will forecast the suspended sediment for year 2012, 2013, and 2014. The general characteristics of 1826 instances hold the minimum value of 0 ton/day, the maximum value of 1537.84 ton/day, the average of 25.627 ton/day, and the standard deviation of 69.996. Note that the high standard deviation indicates a wide range of data, in our case the sediments are varied from 0 to 1537.84 ton/day making them

more difficult to cluster.

A. Clustering the Sediments

The objective of this experiment is to cluster the sediments into 3 levels (low, medium, and high). The question raised here is in what month the level of sediment is high (and low vice versa). By acknowledging this information it can help alleviate the shallow flooding problem because the sediments may be removed prior the month that normally causes the problem.

We cluster data using the k -means algorithm. The k -means is the most popular clustering algorithm because it is easy to understand and to implement [8]. Moreover, its time complexity is only $O(tkn)$ where n is the number of data points, k is the number of clusters, and t is the number of iterations. The algorithm randomly chooses k data points (seeds) to be the initial cluster centers (centroids). Then it assigns each data point to the closest centroid and re-calculates the centroids using the current cluster memberships. The process is repeated until a convergence criterion is met.

For this experiment, we specify $k = 3$ and seed = 70 which is chosen as close as to the standard deviation of the data. We set maximum number of iterations to 5,000. The distance function to use for instances comparison is set to Euclidean distance and also Manhattan distance. Then, a separate experiment is held for checking which of the two functions yields clusters that fit better to our criteria.

Fig. 6 shows two plots of sediment (ton/day) for 1826 instances (days) where Cluster 0 is blue indicating medium level of sediments, Cluster 1 is red indicating low level of sediments, and Cluster 2 is green indicating high level of sediments. Although within cluster sum of squared errors of Fig. 6 (a) is only 0.8365, only 6 out of 1826 instances (0.3%) belong to Cluster 2. The result in Fig. 7 confirms that Euclidean distance function appears to be sensitive to outliers (data points that are very far away from other data points).

The centroids of Cluster 0, 1, and 2 are 129.13, 9.74, and 884.14, respectively (see Fig. 8). The centroid of Cluster 2 is too high and its members mostly are outliers. Manhattan distance function, on the other hand, deals better with data points that are much further away from the centroids than other data points. When the Manhattan distance is used, then centroids are computed as the component-wise median rather than mean [16]. Fig. 9 shows that the centroids of Cluster 0, 1, and 2 are 34.43, 2.04, and 107.38, respectively. The proportion of clustered instances belongs to Cluster 0, 1, and 2 are 16%, 73%, and 11%, respectively. The Manhattan distance function fits better to our criteria of having medium, low and high level of clustered sediments.

Fig. 9 also reveals that most instances in October have relatively high level of sediment and most of instances in January have relatively low level of sediment. The biggest portion of Cluster 2 (high level cluster) is made up of 50 instances which are from October. On the opposite, the biggest portion of Cluster 1 (low level cluster) is made up of 152 instances which are from January. In addition, most of instances in November have relatively moderate level of sediment.

B. Forecasting the Sediments

We specify 1095 days (three years) as the number of time steps to forecast beyond the end of the supplied data. We forecast the sediment using the most common neural network model called multilayer perceptron. Since the increase of sediments is not linear, using linear regression will fail to forecast the data. On the other hand, more complex network topologies have more weights that are adjusted in the training process of the network. The reason for the overwhelming use of multilayer perceptron is because such network architecture is akin to multivariate non-linear regression model [17].

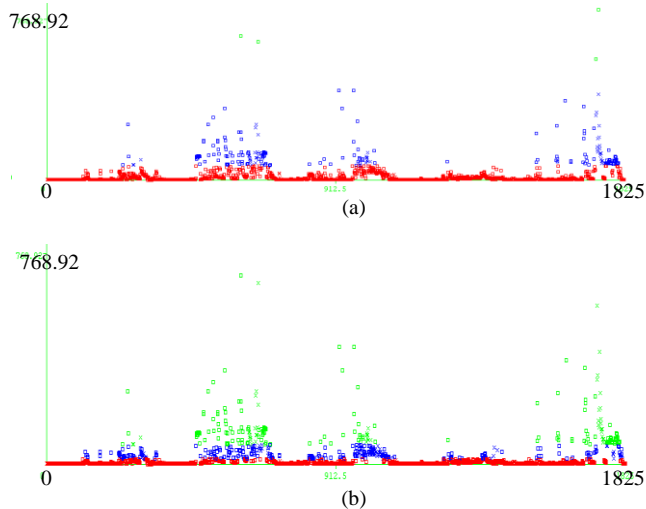


Fig. 6. Instance number versus sediment (ton/day) where cluster 0 is blue, cluster 1 is red, and cluster 2 is green. (a) Euclidean distance (b) Manhattan distance.

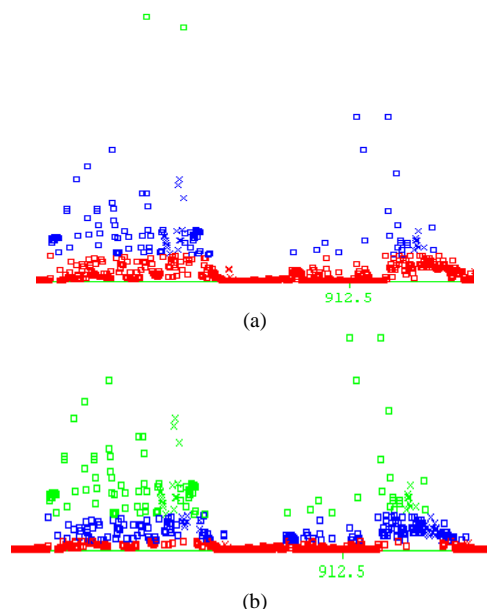


Fig. 7. An extended view of Fig. 6 zooming in the three clusters (a) Euclidean Distance (b) Manhattan Distance.

For this experiment, we specify the number of hidden layers of neural network to be 3. The amount the weights updated (learning rate) is set to 0.1. Momentum applied to the weights during updating is 0.05. Seed used to initialize the random number generator for setting the initial weights of the

connections between nodes, and also for shuffling the training data is set to 0. The number of epochs to train through is limited to 5,000 epochs. We also apply a decay parameter enabling the decrease of the learning rate. It divides the starting learning rate by the epoch number to determine the current rate. This may stop the network from diverging from the target output, as well as improve general performance.

```

=== Run information ===
Scheme: weka.clusterers.SimpleKMeans -V -N 3 -A
"weka.core.EuclideanDistance -R first-last" -I 5000 -num-slots
1 -S 70
Relation: data_mining
Instances: 1826
Attributes: 4
          sediment_numbers
Ignored:
          Date
          Month
          Level
Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===
kMeans
=====
Number of iterations: 16
Within cluster sum of squared errors: 0.8365518955818525
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute          Full Data      Cluster#
                   (1826)        0          1          2
=====
sediment_numbers   25.6273        129.1314          9.743
884.1433
                   +/-69.9965      +/-71.4016      +/-14.4345
+/-355.367

Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===
Clustered Instances
0      199 ( 11%)
1      1621 ( 89%)
2         6 (  0%)

Class attribute: Month

Classes to Clusters:
  0  1  2 <-- assigned to cluster
  0 155  0 | Jan
  0 141  0 | Feb
  2 153  0 | Mar
 14 136  0 | Apr
  7 148  0 | May
 16 134  0 | June
 15 140  0 | July
 17 137  1 | Aug
 28 121  1 | Sep
 46 105  4 | Oct
 41 109  0 | Nov
 13 142  0 | Dec
Cluster 0 <-- Oct
Cluster 1 <-- Jan
Cluster 2 <-- Aug
    
```

Fig. 8. Simulation result of clustering based on *k*-means algorithm (using euclidean distance).

```

=== Run information ===
Scheme: weka.clusterers.SimpleKMeans -V -N 3 -A
"weka.core.ManhattanDistance -R first-last" -I 5000 -num-slots
1 -S 70
Relation: data_mining
Instances: 1826
Attributes: 4
          sediment_numbers
Ignored:
          Date
          Month
          Level
Test mode: Classes to clusters evaluation on training data

=== Clustering model (full training set) ===
kMeans
=====
Number of iterations: 13
Sum of within cluster distances: 13.365493159236326
Missing values globally replaced with mean/mode
Cluster centroids:
Attribute          Full Data      Cluster#
                   (1826)        0          1          2
=====
sediment_numbers   4.09           34.43           2.04           107.38
                   +/-69.9965      +/-14.0929      +/-4.3328
+/-157.0717

Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===
Clustered Instances
0      296 ( 16%)
1     1329 ( 73%)
2       201 ( 11%)

Class attribute: Month

Classes to Clusters:
  0  1  2 <-- assigned to cluster
  3 152  0 | Jan
  0 141  0 | Feb
  2 151  2 | Mar
 15 121 14 | Apr
 14 134  7 | May
 21 113 16 | June
 22 118 15 | July
 19 118 18 | Aug
 51  72 27 | Sep
 56  49 50 | Oct
 55  55 40 | Nov
 38 105 12 | Dec

Cluster 0 <-- Nov
Cluster 1 <-- Jan
Cluster 2 <-- Oct
    
```

Fig. 9. Simulation result of clustering based on *k*-means algorithm (using Manhattan Distance).

Three hidden layers of the neural network are shown in Fig. 10. The red nodes are hidden layers. The orange node is the output node with a label on the right showing the class the output node represents. The inputs are provided from the green labels on the left.

Fig. 11 displays the sediment (ton/day) from January 01,

2007 to December 31, 2011 (1826 day-period) and forecasted sediment until the end of 2014 (1095 day-period). The three years forecasted sediment is closely monitored in Fig. 12 where the up-down pattern is emerged. We can see the peaks in every October and the valleys in every January. At the same point in time sediment level appears to be higher and higher every year. In our opinion this may be a result of floods that are more frequent and more severe.

We performed an evaluation of performance on the test data. Note that the last 45 days are used as the test data while the first 1781 days are used as the training data. The performance measures are:

- 1) Mean absolute error (MAE)
- 2) Mean squared error (MSE)
- 3) Root mean squared error (RMSE)
- 4) Relative absolute error (RAE)
- 5) Root relative squared error (RRSE)

The result of this evaluation is presented in Table II. To verify our previous choice of a number of hidden layers, several errors are listed for 1, 3, 5, 7, and 9 hidden layers, respectively.

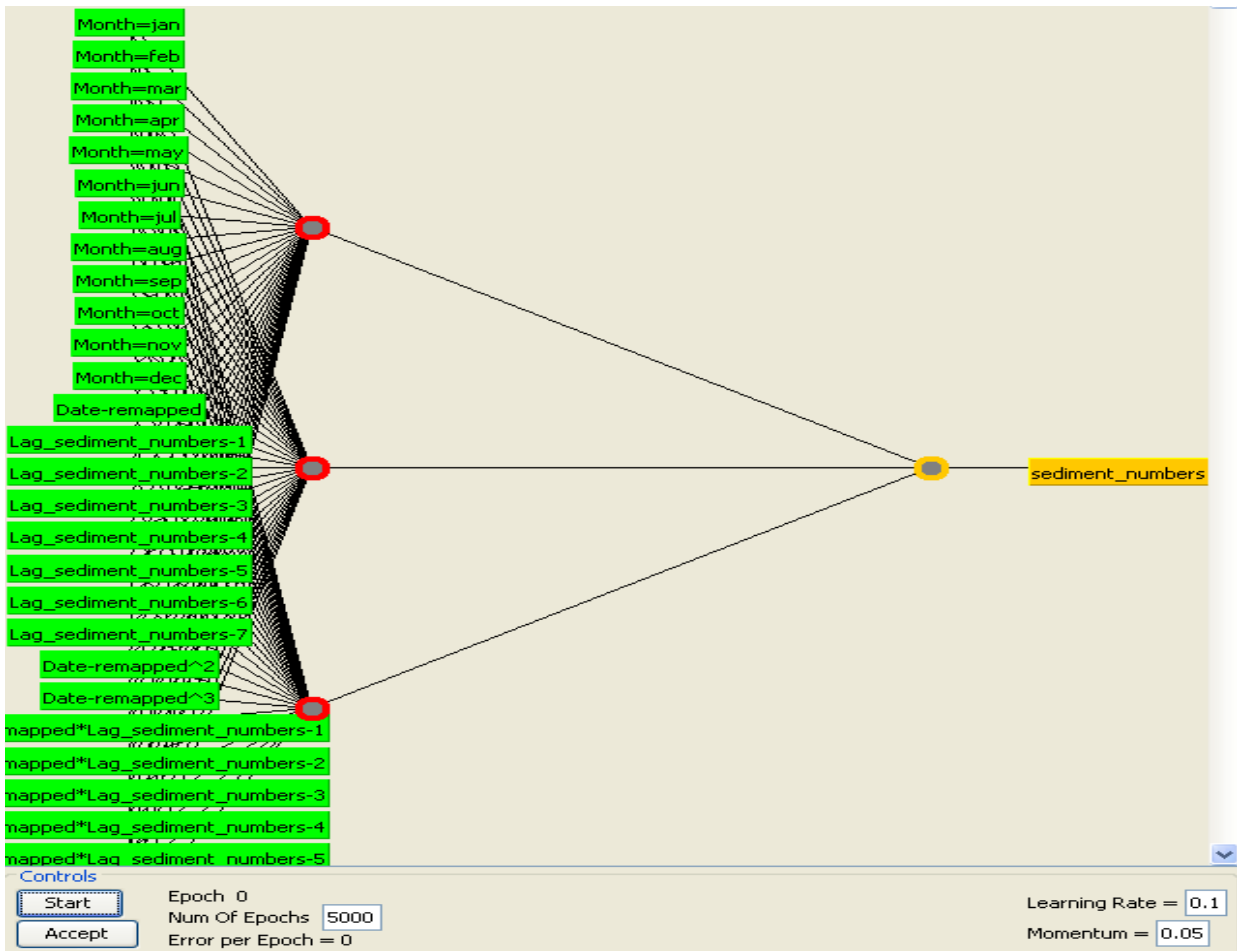


Fig. 10. WEKA's neural network GUI showing 3 hidden layers.

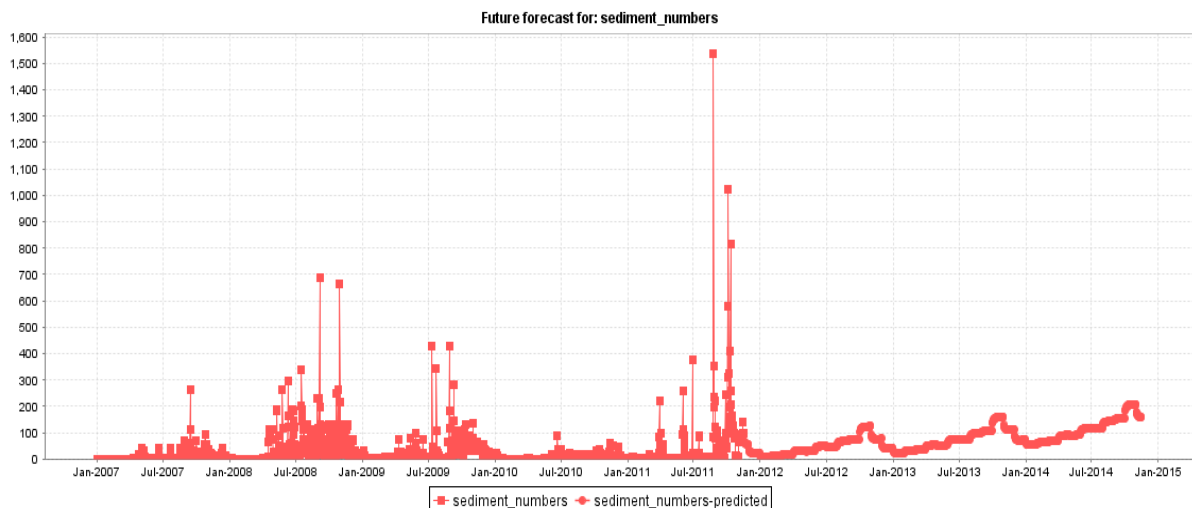


Fig. 11. Date versus sediment (ton/day) from 01-Jan-2007 to 31-Dec-2011 plus the forecasted sediment level until the end of 2014

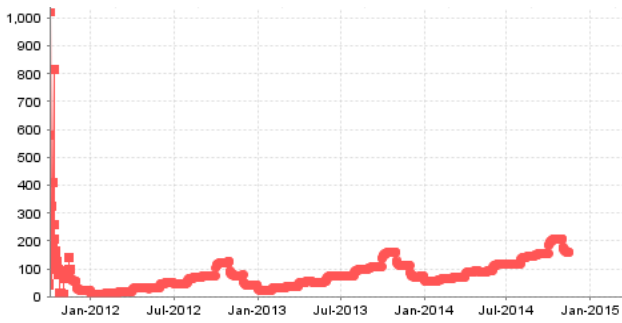


Fig. 12. Date versus forecasted sediment (ton/day) from 01-Jan-2014 to 31-Dec-2014

TABLE II: EVALUATION RESULTS OF NEURAL NETWORK FOR FIVE DIFFERENT NUMBERS OF HIDDEN LAYERS.

Errors	Number of Hidden Layers				
	1	3	5	7	9
MAE	25.7419	24.528	24.957	25.1828	26.2532
RMSE	30.8989	29.3553	29.7573	30.1275	31.3255
RRSE	199.7097	189.7605	192.142	194.5755	202.1151
RAE	234.7813	223.786	227.2345	229.3673	238.7277
MSE	954.7428	861.731	885.4982	907.6672	981.2847

WATER YEAR 2007 SOUTHERN PENINSULA WEST COAST Khlong Bang Yai at Ban Ket Ho, Phuket (X.190A)														
Location Lat. 08° 46' - 25" N., Long. 98° 23' - 36" E., on right bank at the bridge, Ban Ket Ho, Amphoe Ka Thu, Phuket														
Drainage Area	28.92 sq.km.													
Method of Sampling	Depth Integrating													
Instrument Used	US DH - 59													
Period of Available Records	2006 Cont'd													
Actual Measurement	2006 Cont'd													
Using Rating Curve Water Year	2006-2007													
Number of observation	45													
R - Square	0.7100													
Remarks	Continued Sediment Station													
QS = 6.2761 QW ^{1.588}														
WATER YEAR - 2007 SUSPENDED SEDIMENT IN TONS PER DAY, WATER YEAR APRIL 1, 2007 TO MARCH 31, 2008														
DATE	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	ANNUAL	
1	0.08	3.87	1.01	1.45	2.92	20.39	71.62	29.58	8.50	3.50	0.35	0.52		
2	0.05	7.06	0.90	1.22	2.79	26.42	71.62	24.81	8.04	1.45	0.27	0.52		
3	0.14	3.87	0.80	1.11	2.65	29.58	29.58	26.42	2.79	1.22	0.61	0.52		
4	0.14	43.83	1.22	1.01	2.65	23.23	16.10	23.23	1.11	1.45	0.61	0.52		
5	0.14	33.63	1.11	0.90	2.45	18.89	10.89	37.82	1.45	1.11	0.61	0.52		
6	0.14	10.40	1.11	0.90	2.18	10.40	13.43	21.80	2.79	0.61	0.52	0.52		
7	0.14	6.79	1.22	0.90	1.81	9.44	23.23	18.89	5.03	3.14	0.52	0.52		
8	0.14	4.25	1.11	0.90	1.81	6.28	17.53	16.10	0.14	0.80	0.52	0.52		
9	0.14	12.29	1.11	0.90	1.69	1.93	13.43	18.89	21.44	0.43	0.70	0.43		
10	0.14	32.84	1.01	2.45	1.57	0.35	9.91	18.89	14.80	0.43	0.61	0.43		
11	1.69	26.92	1.11	2.18	1.57	1.69	8.96	24.81	1.22	1.45	0.61	0.43		
12	2.18	10.11	1.01	2.06	1.57	0.52	8.96	20.39	42.13	1.33	0.61	0.43		
13	0.61	3.28	1.01	1.81	1.33	10.89	7.59	16.10	34.57	1.22	0.52	0.43		
14	0.27	3.14	1.01	1.81	1.81	27.93	10.40	13.43	31.92	1.01	0.52	0.43		
15	0.20	1.81	1.22	1.69	2.06	264.65	8.50	3.14	2.18	0.90	0.43	0.43		
16	0.17	1.93	1.57	1.69	0.52	62.38	6.71	0.00	2.79	1.01	0.35	0.35		
17	0.17	0.90	1.93	1.45	10.89	110.95	5.44	13.43	2.79	1.11	0.35	0.35		
18	0.14	1.11	3.87	1.22	39.61	32.71	6.71	13.43	2.79	1.01	0.35	0.35		
19	0.14	1.11	43.83	1.11	41.28	24.81	16.10	9.44	2.18	0.90	0.35	0.27		
20	0.14	1.01	30.23	1.11	31.13	14.80	10.89	1.33	2.18	0.80	0.35	0.27		
21	0.11	0.80	4.63	1.22	36.05	10.40	9.44	1.45	2.06	0.80	0.35	0.24		
22	0.11	0.70	2.45	1.22	16.10	8.50	8.04	2.18	3.50	0.70	0.35	0.24		
23	0.11	0.52	2.06	42.13	5.44	7.14	4.63	0.61	13.53	0.61	0.35	0.24		
24	0.08	0.43	1.93	7.06	0.05	6.28	24.81	3.14	0.20	0.61	0.43	0.24		
25	0.08	0.43	1.93	4.09	24.81	12.19	58.04	0.70	0.90	0.52	0.52	0.24		
26	16.10	0.35	1.93	5.28	7.14	32.71	95.59	0.00	0.80	0.52	0.52	0.20		
27	0.80	0.27	1.81	5.03	1.81	14.80	36.05	0.90	3.28	0.43	0.52	0.11		
28	0.20	0.43	1.81	3.72	12.19	1.45	41.28	0.43	1.93	0.43	0.52	0.11		
29	0.80	0.80	1.69	3.87	53.64	0.61	53.64	0.00	1.57	0.43	0.52	0.11		
30	1.01	1.45	1.57	4.25	71.62	29.58	39.61	5.44	2.18	0.35	0.08	0.08		
31	1.57			2.92	23.23		32.71		2.65	0.35		1.45		
TOTAL	26.36	217.90	119.20	108.66	406.37	821.90	771.44	366.78	223.44	30.63	13.84	12.02	3118.54	Ton/day
MEAN	0.88	7.03	3.97	3.51	13.11	27.40	24.89	12.23	7.21	0.99	0.48	0.39	8.51	Ton/day
MAX	16.10	43.83	43.83	42.13	71.62	264.65	95.59	37.82	42.13	3.50	1.45	264.65		Ton/day
MIN	0.05	0.27	0.80	0.90	0.05	0.35	4.63	0.00	0.14	0.35	0.27	0.08	0.00	Ton/day
Stop - Program terminated.														

Fig. 13. Example of suspended sediment in tons per day for 2007 from hydrology and water management center for southern region [15]

According to Table II, the neural network model with 3 hidden layers has smaller errors than the other models. Hence, this verifies our choice of using 3 as a number of hidden layers in our experiment.

IV. CONCLUSIONS

This paper presented the real world problem in preventing the shallow flooding problem of Khlong Bang Yai at Ban Ket Ho, Phuket. We have learnt that the deposit level of suspended sediments is peaked in October and is low in January. We are aware now that the sediments may be dredged prior October, so that the canal drainage capacity is maintained. Also, we observed from the forecasting result that the level of sediments will be higher every year. The municipality should have an action plan in place to alleviate the problem. For the future research we would like to clearly highlight the associations between the level of precipitation, the level of hydrograph and the level of suspended sediments.

Another challenging task for us is to take nearby houses and construction sites into the consideration for a more accurate prediction of suspended sediments.

ACKNOWLEDGMENT

The authors wish to thank CoE lecturers for their useful discussion of the subject. This work was supported by a grant from Faculty of Engineering, Prince of Songkla University.

REFERENCES

- [1] J. Sunkpho and C. Ootamakorn, "Real-time flood monitoring and warning system," *Songklanakar Journal of Science and Technology (SJST)*, vol. 33, no. 2, pp. 227-335, April 2011.
- [2] S. Sornsakda and A. Heednacram, "Simulation of Flow Acceleration by Water Jets for Flood Relief," in *Proc. the 1st Annual PSU Phuket International Conference*, Phuket, January 2013, pp. 1-7.
- [3] K. Chaowanawatee and A. Heednacram, "Implementation of Cuckoo Search in RBF Neural Network for Flood Forecasting," in *Proc. the 4th International Conference on Computational Intelligence, Communication Systems and Networks*, Phuket, July 2012, pp. 22-26.
- [4] K. Chaowanawatee and A. Heednacram, "Improved cuckoo search in RBF neural network with gaussian distribution," in *Proc. the 8th IASTED International Conference on Advances in Computer Science*, Phuket, April 2013, pp. 379-386.
- [5] N. Chantasut, C. Charoenjit, and E. Tenprasert, "Predictive mining of rainfall predictions using artificial neural networks for chao phraya river," presented at the AFTTA/WCCA Joint congress on it in agriculture, Thailand, 2004.
- [6] Hydrology and Water Management Center for Southern Region. (2013). River Cross-sectional Data for Southern Peninsula West Coast. [Online]. Available: <http://hydro-8.com/main/station/sta55/X190a.pdf>.
- [7] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical machine learning tools and techniques.*, 3rd edition, Morgan Kaufmann, San Francisco, 2011, ch 4-6.
- [8] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, 2nd ed. Springer, Heidelberg, 2011, ch 4.
- [9] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, June 2010.
- [10] T. Chaipimonplin, L. See, and P. Kneale, "Improving neural network for flood forecasting using radar data on the Upper Ping River," in *Proc. the 19th International Congress on Modelling and Simulation*, Perth, December 2011, pp. 1070-1077.
- [11] S. Seung. (September 2002). Multilayer perceptrons and backpropagation learning. [Online]. Available: <http://hebb.mit.edu/courses/9.641/2002/lectures/lecture04.pdf>
- [12] W. Lu, "Neural network model for distortion buckling behaviour of cold-formed steel compression members," M. S. thesis, Helsinki Univ. of Technology, Laboratory of Steel Structures Publications 16, 2000.
- [13] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, 2009.
- [14] Weka Examples. (2013). A collection of example classes, as part of an ANT project, included in the WEKA snapshots. [Online]. Available: <https://svn.scms.waikato.ac.nz/svn/weka/trunk/wekaexamples/>
- [15] Hydrology and Water Management Center for Southern Region. (2013). Data of Suspended Sediment in Tons per Day for Southern

Peninsula West Coast. [Online]. Available: <http://hydro-8.com/main/TKD/X.190A.pdf>.

- [16] D. Arthur, and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in *Proc. of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027-1035.
- [17] S. G. Anantwar, and R. R. Shelke, "Simplified approach of ANN: strengths and weakness," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 1, no. 4, pp. 73-77, April 2012.



Apichat Heednacram is from Phuket, Thailand. He received B.Eng. microelectronic engineering (first class honours) from Griffith University, Australia in 2001. He pursued his Ph.D. in 2007 at Griffith University funded by Australian Endeavour Postgraduate Award. He obtained the Ph.D. (Algorithms and Theory) from Griffith University, Australia in 2010. His dissertation is the NP-Hardness

of Covering Points with Lines, Paths and Tours and their Tractability with FPT-Algorithms.

Since 2005, he has been working as a lecturer in Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University, Phuket Campus, Thailand. He was the deputy head of Department of Computer Engineering from 2011-2012. He is currently a supervisor of Master and Ph.D. students. His recent research interests are in the area of algorithms, neural networks and image processing.



Thammaratt Samitalampa is from Phuket, Thailand. He received M.S. in information systems and MBA. from Eastern Michigan University, USA, in 1997 and 2000, respectively.

He used to work at Parke-Davis, Ann Arbor, Michigan and Ajilon, Southfield, Michigan, USA. Since 2002, he has been working for Department of Computer Engineering, Prince of Songkha University, Phuket Campus, Thailand. His research interests are in the area of software engineering and data mining.