A Study of Applying Genetic Algorithm to Predict Reservoir Water Quality

L. Chen, M. Jamal, C. Tan, and B. Alabbadi

Abstract—**This paper is aimed at demonstrating a genetic algorithm method and applying it to predict the water quality of reservoir in Taiwan island using remote sensing data. Genetic algorithms will be combined with operation tree (GAOT) to find the relationships between input and output data. A fittest function type will be obtained automatically from this method. The advantages of GA are global optimization, nonlinearity, flexibility and parallelism. In the current case study, GA is used to construct the relationship between algae concentration and Landsat sensor data. The results show that the model has better performance than the traditional LN transform of linear regression method, and similar performance compared with back-propagation neural network (BPNN) method.**

Index Terms—**Genetic algorithm, Landsat, LN transform linear regression, back-propagation neural network, operation tree.**

I. Introduction

In reservoir quality assessment, traditional automatic methods have been used to construct the relationship between water quality of a reservoir and the satellite images. Unfortunately, there are many hidden and useful relationships between the input and output data, which may not be recognized by the analyst. Thus many kinds of data mining techniques have been developed, such as statistics, artificial neural networks, and genetic algorithms.

Evolutionary computation techniques, which are based on a powerful principle of evolution: survival of the fittest, are very efficient optimization methods. Well-known algorithms in this domain include genetic algorithms, evolutionary programming, evolution strategies, and genetic programming. Among these methods genetic programming (GP) is one of the most popular search methods. GP has been widely used for the automatic generation of programs or equations between the inputs and outputs. It has an advantage over traditional statistical methods because it is distribution free, i.e., no prior knowledge is needed about the statistical distribution of the data [1] and its ability to discover the L. Chen, M. Jamal, C. Tan and B. Alabbadi underlying data relationships and express them mathematically. According to Chen [2], [3], constructing the data structure of a dynamic tree of the GP could be a difficult task when GP was used to create the mathematical equations. For example, it is hard to choose the proper size of a tree that can express a meaningful equation in advance.

On the other hand genetic algorithm (GA) can be used to solve a discrete optimization problem. GA represents a paradigm of evolution computation based on natural evolution and derived from the ideas of ‘‘the survival of the fittest’’ [4], such as inheritance, selection, crossover and mutation. The advantages of GA are global optimization, nonlinearity, flexibility and parallelism [5].

In recent years, Yeh and Lien [6] proposed a newly developed programming system, genetic algorithm operation tree (GAOT). GAOT has two parts, operation tree (OT) and GA. OT is a tree structure which represents a mathematical formula and can be optimized to generate a self-organized regression formula, but optimizing the tree structure is a discrete optimization which cannot be solved using mathematical programming [7]. GA can be used to optimize the OT to fit experimental data in the GAOT process.

Recently, some researches such as concrete researches and typhoon researches have been applied GAOT to predict accurate models [7]-[12]. Chen *et al.* [13] used GAOT to estimate the compressive strength of high-performance concrete. Substantial experimental data were used to compare the accuracy of the results obtained using the model-building technique. The results indicated that this model, GAOT, can be used to formulate highly nonlinear mathematical equations involving few estimation errors to predict the compressive strength of high-performance concrete. Chen *et al.* [14] applied GAOT to improve the radar-based rainfall estimation. A case study for typhoon rainfall in shih-men raingauge station. The 10 most torrential typhoon events between 2000 and 2010 are the input variables. Two GAOT models, including five layers and six layers OT, are proposed and the estimations were compared with the empirical rainfall estimation formula ($Z = aR^b$). The results showed that the GAOT with six layers is better than those of the Z-R equation by traditional regressive method but similar to those of GAOT with five layers.

Taiwan is located in a transition zone between the tropical and subtropical climates. It has several reservoirs that provide water for drinking, living and industrial use. Eutrophication is one of the major water quality problems in the reservoir [15]. Phosphorus and nitrogen nutrients applied to many fruit and vegetable gardens located on the upper stream of the reservoir are often washed into the reservoir during heavy rains. These nutrients are essential for algal growth and algal bloom, and they have been recognized as limiting nutrients in the reservoir [16].

Reservoir water quality is traditionally monitored and

Manuscript received March 3, 2017; revised April 13, 2017.

Li Chen, Mohammad Jamal Mtlak Abualghanam, and Basmah Alabaddi are with the Department of Civil Engineering, Chung Hua University, Hsinchu 707, Taiwan, ROC (e-mail: lichen@chu.edu.tw, chtan@aerc.org.tw, basmaalabadi@yahoo.com).

Chih-Hung Tan is with the Department of Agricultural Engineering Research Center, No.196-1, Zhongyuan Rd., Zhongli Dist. Taoyuan City 32061, Taiwan.

DOI: 10.7763/IJMO.2017.V7.566

evaluated based on field data. Collecting and analyzing field data are expensive and time consuming. To improve traditional data collection method, utilization of remote sensing data for water quality assessment has been investigated. Therefore, objectives of this study include estimation of the water quality for a reservoir in Taiwan using Landsat 8 data and introducing a GAOT to compare with the regression method and the back-propagation neural network (BPNN) method.

Wong *et al*. [17] used Aqua/ MODIS data for estimating suspended solids and salinity in marine Hong Kong, where monitoring stations are few, and determined significant correlations between MODIS data and in situ data.

Danbara [18] used Landsat data to retrieve of water quality variables of Tana lake. The variable are suspended particulate matter, colored dissolved organic matter, chlorophyll-a and turbidity. The results showed that the most dominant water quality variable which predominantly affects the inherent optical properties of the lake is suspended particulate matter. Laili *et al*. [19] used Landsat data to develop more accurate total suspended solid and chlorophyll-a concentration retrieval algorithms at Poteran island water of Indonesia. The results showed that the developed algorithm for estimating total suspended solid and chlorophyll-a concentration produced acceptable accuracy, thus extracting water information from satellite data using these algorithms are applicable.

II. Genetic Algorithm Operation Tree

A. Genetic Algorithm

The GAs are a class of stochastic algorithms that has been successfully used as optimization techniques based on the principles of natural selection and genetics for solving a wide range of problems [20]. The GA approach differs from other optimization and search procedures in four ways: (1) GAs typically use a coding of the decision variable set rather than the decision variables themselves; (2) GAs search for optimal solutions in a population of decision variable set, rather than a single decision variable set; (3) GAs use the objective function itself, rather than derivative information on the objective function and constraints; and (4) GAs use probabilistic transition rules rather than deterministic rules [5]. The idea of GAs was first introduced by John Holland back in the 1970s and was later popularized by David Goldberg. A GA generates a population of possible solutions encoded as chromosomes, evaluates their fitness and creates a new population by applying genetic operators which are crossover and mutation. Each chromosome consists of "genes" (e.g., bits), each gene being an instance of a particular "allele" (e.g., 0 or 1). By repeating this process over many generations, the GA has five basic components which are: a genetic representation of solutions to the problem; a way to create an initial population of solutions; an evaluation function rating solutions in terms of their fitness; genetic operators that alter the genetic composition during reproduction; values for the parameters of GAs.

GAs use fitness values of individual chromosomes to carry out reproduction. As reproduction takes place, the crossover operator exchanges parts of two single chromosomes and the mutation operator changes the gene value in some randomly chosen location of the chromosome. After a number of successive reproductions, the less fit chromosomes become extinct, although those best fit gradually come to dominate the population.

The reproduction process copies parent chromosomes into a tentative new population. The probability of selected chromosomes for the next generation is directly proportional to its fitness value. A great number of selection algorithms have been presented in the literature [21] among them; Roulette Wheel selection is perhaps the most common method.

The crossover recombines two parent chromosomes to produce offspring new chromosomes for the next generation, which includes three steps. First, chromosomes from the mating pool are randomly paired. Then, it is determined whether these pairs should go for crossover or not, based on a preset crossover probability. Third, chromosome segments between mating pairs are interchanged. The operator can be one-point crossover or multi-point crossover as shown in Fig. 1 (a) or (b).

To sustain genetic diversity into the population, mutation is also made occasionally with small probability. A random position of a random string is selected and is replaced by another character from the alphabet; e.g., in the binary coding, this simply means changing a 1 to a 0 and vice versa as shown in Fig. 1 (c).

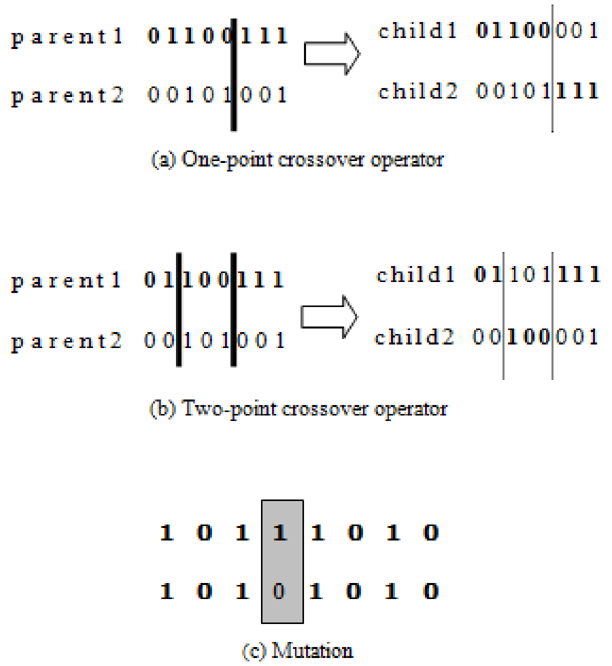

Fig. 1. Binary code.

B. Operation Tree

The OT is a tree structure which represents a mathematical formula. A five-layered OT model is shown in Fig. 2. N_1 is the root node denoted a mathematical operation (+, –, ×, ÷, ln or exp). N_2–N_{15} are interior nodes denoted a variable, a constant, or a mathematical operation. N_{16}–N_{31} are leaf nodes denoted a variable or a constant [22]. Fig. 3 shows an example of operation tree model. The model used mathematical operations –, ÷, and ln, variables A, and C, and constant 15.

The operation tree can express the following equation (1).

$$y = \frac{A}{B} + [C - D] \tag{1}$$

Establishing a regression model using OT needs to determine appropriate mathematical operations, variables and constants in the root, branches, and leaves of the OT. When the tree-style structure is set up to represent a specific mathematical formula, operation tree can generate predicted output value for each data by substituting the input values of data into the variables on branches or leafs of the tree-style structure. OT performance can be evaluated with root mean squared error (RMSE) between predicted and actual output values. The minimum the RMSE of OT, the best the OT fits the dataset. The conventional regression analysis requires predetermined formula structure and is only allowed to adjust the regression coefficients in the predetermined structure. The disadvantage can be overcome by OT. OT is a tree-style data structure which represents a flexible mathematical formula and optimizing the structure to fit the data best is a discrete optimization problem. Therefore, the optimization of OT cannot be solved with conventional mathematical programming. GA, that can solve discrete optimization problem, is adopted in this study to optimize the OT to fit the data best.

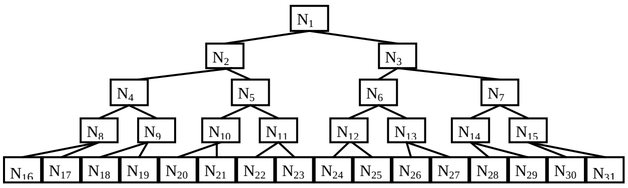

Fig. 2. five-layered operation tree.

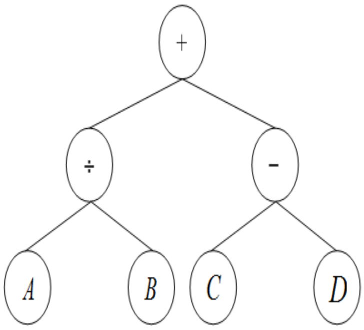

Fig. 3. The parse tree of equation 1.

C. Genetic Algorithm Operation Tree

If we have a set of values for the variables on the OT, the OT (formula) can deduce a function value [6]. How to produce an OT whose formula can fit the data best is an optimization problem when we own a set of data of the algae. Because it is a discrete and continues hybrid optimization problem, it's very difficult to solve it. The GA is one of optimization paradigms about mimicking the natural evolution mechanism. The framework of the GA naturally corresponds to a discrete optimization problem. Therefore, this study employed OT to express a regression formula which can fit the data of algae best, and GAs to optimize the OT to fit the data set to produce a self-organized regression formula. OT plays the architecture to represent an explicit formula and GA plays the optimization mechanism to optimize the OT to fit experimental data in the GAOT process, Fig. 4. In the framework, the database is simply shuffled using a random sampling, and divided into training and testing data. The training data and testing data are used to evaluate the fitness of each individual (OT) in the population [6]. However, the fitness of training data is used in the reproduction process but the fitness of testing data is not. So, it is used to evaluate the generalization of OT.

There are some parameters that may affect the performance of GA. Reference [22] suggested the following parameters: (1) crossover rate = 0.4–0.99; (2) mutation rate = 0.0001–0.1; (3) population size = 10–1,000. Therefore, the following parameters as show in Table I were used in this study. In addition, a five-layered OT was employed in this study as shown in Fig. 2. Table II lists the gene codes of mathematical operations, variables, and constants. GAOT must comply with the following rules:

- When a node is assigned with code 6 (natural logarithm operation), Table II, its code of right branch will be neglected.
- When a node (gene) is assigned with code 8, 9…15 (variable), it will be an end node (leaf), i.e., its offspring nodes (right- and left- offspring) will be neglected.
- When a node (gene) is assigned with code 7 (constant), it would be an end node (leaf).

The main steps of GAOT follows below:

- Initialization and parameters setting: This step sets the parameters of GAOT and randomly generates the initial population.
- Training dataset: The data set is divided into training data and testing data.
- Operation tree training model: operation tree is a learning tool able to build an explicit formula for use as a prediction model.
- Fitness evaluation: used to evaluate each chromosome.
- Genetic algorithm procedure: genetic algorithm begins with the selection process, crossover, and mutation. This procedure will repeat until termination criterion is satisfied.
- Optimal solution.

The correlation coefficient between the predicted and the actual values is adopted as the fitness function of GAOT. GAOT is able to achieve both a "high linear correlation" and a "small estimating error" simultaneously in most cases, so we chose the former as the objective function, equation (2).

$$y = \alpha + \beta * f \tag{2}$$

where f is the predicted value of the operation tree; y is the modified predicted value; α and β are the regression coefficients.

$$\alpha = \overline{y} - \beta * \overline{f} \tag{3}$$

$$\beta = \frac{\sum_{i=1}^{n}(f_i - \overline{f})*(y_i - \overline{y})}{\sum_{i=1}^{n}(f_i - \overline{f})^2} \tag{4}$$

where $\overline{y} =$ the mean of actual value of dataset, $\overline{f} =$ the mean of predicted value of dataset, $y_i =$ the actual value of *i* data of dataset, and $f_i =$ the predicted value of i data of dataset (*i*=1,…, n).

TABLE I: SETS PARAMETERS DURING GAOT APPLICATION

Set Parameters	Value
Population Size	300
Chromosome Length	8 bit
Crossover Rate	0.9
Mutation Rate	0.01
Generations	190
Generation Gap	0.98
Constant Range (K)	-100~100
Elitist	Yes
Run	7

TABLE II: GENETIC CODE OF MATHEMATICAL OPERATION, VARIABLE AND CONSTANT

Code	1	2	3	4	5
Meaning	+	-	×	÷	x^y
Code	6	7	8	9	10
Meaning	Ln	k	Band1	Band2	Band3
Code	11	12	13	14	
Meaning	Band4	Band5	Band6	Band7	

III. ALGAE ESTIMATION BY GENETIC ALGORITHM OPERATION TREE

The purpose of this application was to utilize Landsat 8 data to evaluate algae concentrations in a reservoir. Whenever the relationship was made, the algae concentration in the reservoir may be computed in time. Whereas the relationship between algae concentration in the reservoir and corresponding image data was constructed through the genetic algorithm operation tree. This system identification problem may be viewed as a search for a function type, which maps input values onto an output value. The linear correlation coefficient is used as the objectives in this study. This system identification problem may be viewed as a search for a function type, which maps input values of Landsat 8 onto an output value of algae concentration. The CCs value is used as the objectives in this study.

A. Study Area Deji Reservoir

The Deji Reservoir (Fig. 5), formed by Techi Dam, is thin, hyperbola-shaped reservoir. The reservoir is located in the middle of Taiwan on the upstream catchment of the Dajia stream. It has a total area of 60, 160 hectares with a height of 180 to 290 meters. In blocking the waters of the Dajia River, it has created a reservoir measuring 14 kilometers in length and 592 hectares in area. It provides municipal water, generates hydroelectric power, is used for recreation and prevents flooding. The water quality of the reservoir has been routinely monitored at a frequency of once a season since 1998. Monitoring includes different parameters such as algae concentration.

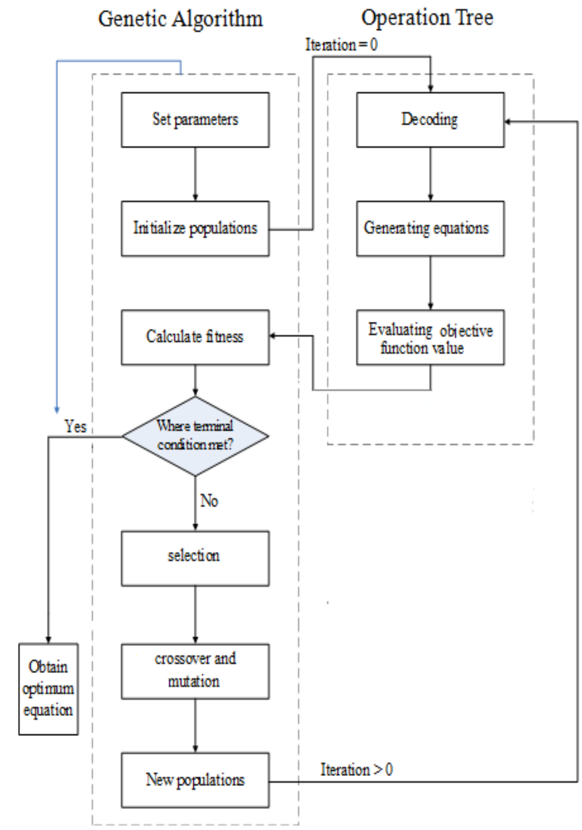

Fig. 4. Flow chart of genetic algorithm operation tree.

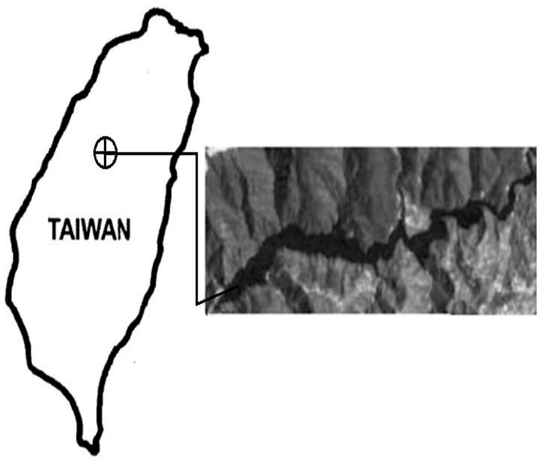

Fig. 5. Deji reservoir location in Taiwan.

B. Algae Data Set

In this study, the actual data of algae (Dinophyta) concentration for 4, 6, 8, 10 /2014, are used and it is expected to have a high correlation with Landsat 8 data. These data were obtained from Agricultural Engineering Research Center, Taiwan R.O.C., and are used to validate the algae distributions which derived from Landsat 8 data. Actual data are selected at same time of Landsat Satellite data overpass.

The Landsat 8 images were acquired from the United States

Geological Survey (USGS) explorer for. This product is atmospherically corrected, and contains the entire Earth every 16 days in an 8-day offset from Landsat 7.

Image data was imported and processed by using ERDAS Imagine 2010. Geometric corrections of images data were performed in order to compare the images data with algae monitoring locations. The geometric correction was applied, and data band for each image was obtained by using ERDAS. Statistical properties of the seven bands are shown in Table III. Forty three entries are used as training data and six as predictive data; the total number of data entries is 49. In order to compare the predicting ability of GAOT and Linear Regression, Correlation coefficients (CCs) was used.

TABLE III: STATISTICAL PROPERTIES OF SEVEN BANDS

Bands	Minimum	Maximum	Mean
$Band_1$	8640.6	27243.2	17941.9
$Band_2$	7772.6	27158.1	17465.4
$Band_3$	6782.4	25110.6	15946.5
$Band_4$	5927.1	25405.3	15666.2
$Band_5$	5488.6	28340.5	16914.6
$Band_6$	5126.8	23100.0	14113.4
$Band_7$	5071.9	19516.6	12294.3

C. Estimation the Algae of Reservoir

Using LN (LR).To estimate the spatial variation of algae concentration in the reservoir using data of satellite images, empirical relationship between digital numbers of the pre-processed image bands and algae concentration was established using LN transform of linear regression (LR) method initially. This model utilized Landsat 8 bands 1 to 7 was given by the following equation (5).

$$LN(a\lg ae)_{LR} = -8.6 + 0.003X_1 - 0.001X_2 + 0.001X_3 - 0.003X_4 + 0.0001X_5 - 0.0002X_6 + 0.0003X_7 \quad (5)$$

where X_1= band$_1$, X_2 = band$_2$, X_3= band$_3$, X_4= band$_4$, X_5= band$_5$, X_6 = band$_6$ and X_7= band$_7$. In equation (5), the weight of X_1 (0.003) is similar with those of X_2 (-0.001), X_3 (0.001) and X_4 (-0.003), which are all higher than those of X_5 (0.0001), X_6 (-0.0002) and X_7 (0.0003). The model was applied on Landsat 8 image and the results show on Fig. 6.

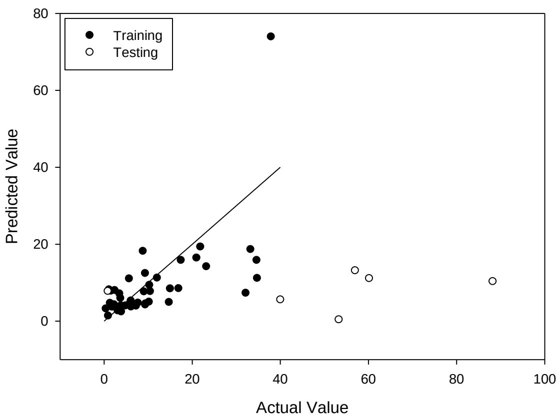

Fig. 6. Taiwan reservoir algae concentration-LN(LR).

The CCs of equation (5) are 0. 63 and 0. 25 for training set and testing set, respectively as shown in Table 4. Since nonlinear relationships may exist between the inputs and outputs, it is necessary to use a more advanced automatic programming and optimization model, such as GAOT to fit the complex nonlinear transfer function between the Landsat 8 bands and algae concentration parameter.

TABLE IV: THE CRITERIA OF LN (LR), BPNN AND GAOT ON THE TRAINING AND TESTING DATA

Models	LN (LR)	BPNN	GAOT
	CCs	CCs	CCs
Training	0.63	0.85	0.83
Testing	0.25	0.81	0.82

Using BPNN. Back propagation neural networks (BPNNs) have quickly become the most widely encountered artificial neural networks. BPNNs were trained using various parameters, initial conditions and numbers of hidden units decided by testing data set to avoiding over fitting at training stage. Then, the BPNN used one hidden layers with four nodes and terminated after 1,000 iterations for the learning procedure. The CC of the BPNN are 0.85 and 0.81 for the training set and testing set, respectively as shown in Table IV, which were similar to those of the GAOT. Fig. 7 demonstrates the scatter diagram of predicted values versus actual values of algae for the BPNN.

Using GAOT. The same data were used to compare with the LN (LR) method described above. The GAOT model is implemented in C++ Language. Fig. 8 shows the result of the parse tree of GAOT. And the final optimal equation obtained from GAOT is shown as equation (6).

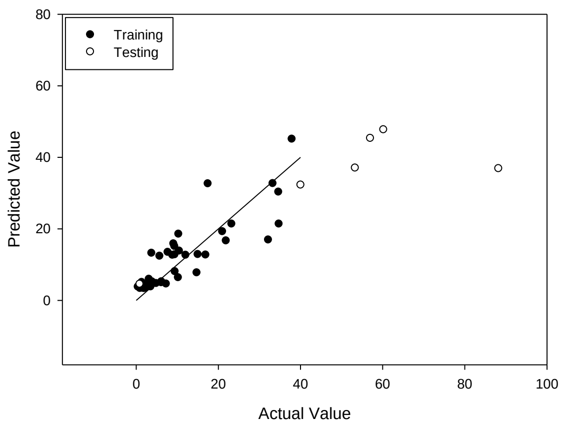

Fig. 7. Taiwan reservoir algae concentration- BPNN.

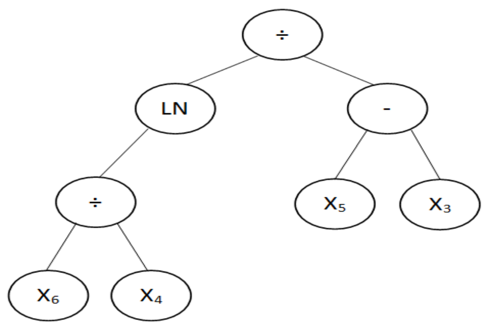

Fig. 8. The parse tree of GAOT.

$$A\lg ae_{GAOT} = 22.88 - 123451.79\left[LN\left(\frac{X_6}{X_4}\right) \Big/ X_5 - X_3 \right] \quad (6)$$

The result shows that only four input variables X3, X4, X5 and X6 were chosen automatically from total seven input variables by GAOT to form equation (6) through a lot of generations' evolutions and competitions. It shows the four input variables have most strong effects on the predicted algae concentration. Fig. 9 shows the algae map using GAOT model.

In Table IV, the result indicates that the CC = 0.83 for training set and CC = 0. 82 for testing set of GAOT is better than of LN (LR) and similar to BPNN. GAOT was found better than the traditional model for algae concentration estimation for both training and testing sets as indicated by the higher CC.

In order to realize the performances of LN (LR), BPNN and GAOT, their diagrams are depicted and compared with each other. The horizontal axis is the actual value of algae, and the vertical axis is the predicted value of algae. Fig. 9 shows that the predicted values for GAOT are closer to ideal line (45°) than LN (LR) model, Fig. 6.

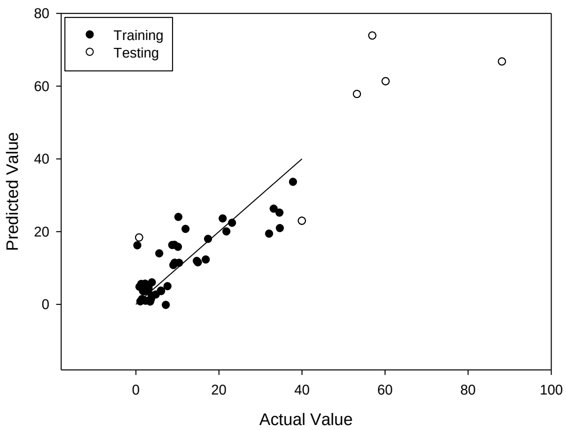

Fig. 9. Taiwan reservoir algae concentration-GAOT.

IV. Conclusion

This paper demonstrate the possibilities of adopting GAOT method coupled with the GA and OT to predict the algae concentration of Taiwan reservoir by Landsat sensor data data. GAOT deals with data to generate a fittest mathematical equation. Few significant variables can be chosen from all input variables automatically. The result shows that GAOT used real number coding as an efficient and robust model. It shows although the use of the GAOT was not simple as LN transform linear regression, it provided an appropriate model to predict reservoir algae using the four input variables. The response figure demonstrates that the relationship between predicted algae and actual algae generated by GAOT was reasonable. The result of this case study indicates that the CC = 0.83 for training data, and CC =0. 82 for testing data of GAOT are better than those of LN (LR) (CC = 0.63 for training data, and CC = 0.25 for testing data), and similar to BPNN (CC = 0.85 for training data, and CC = 0.81 for testing data) as shown in Table 4. The results confirms that GAOT would be the better option than LN (LR) because it models algae concentration without the limitation of linear property which LN (LR), and another method cannot conquer. The current study shows a successful application of GAOT on algae monitoring. Because this method is flexible and possibly applied to other data sets, a more complex equation including more available variables can be achieved. Further studies can be expanded to analyze the other water quality parameters as additional input variables for lakes, rivers, seas, reservoirs and oceans.

References

[1] J. K. Kishore, L. M. Patnaik, V. Mani, and V. K. Agrawal, "Application of genetic programming for multicategory pattern classification," *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 3, pp. 242-257, 2000.

[2] L. Chen, "A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data," *International Journal of Remote Sensing*, vol. 24, no. 11, pp. 2265-2275, 2003.

[3] L. Chen, "A study of applying macro-evolutionary genetic programming to concrete strength estimation," *Journal of Computing in Civil Engineering*, vol. 17, no. 4, pp. 290-294, 2003 b.

[4] L. Davis, *Handbook of Genetic Algorithms*, New York: Van Nostrand Reinhold, 1991.

[5] D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Massachusetts: Addison Wesley, 1989.

[6] I. C. Yeh, and L. C. Lien, "Knowledge discovery of concrete material using genetic operation trees," *Expert Systems with Application*, vol. 36, no. 3, pp.5807-5812, 2009.

[7] C. H. Peng, I. C. Yeh, and L. C. Lien, "Building strength models for high-performance concrete at different ages using genetic operation trees, nonlinear regression, and neural networks," *Engineering with Computers*, vol. 26, pp. 61-73, 2010.

[8] I. C. Yeh, C. H. Lien, A. H. Peng, and L. C. Lien, "Modeling Concrete Strength Using Genetic Operation Trees," in *Proc. 2010 International Conference on Machine Learning and Cybernetics*, 2010, pp.1572-1576.

[9] C. H. Peng, I. C. Yeh, and L. C. Lien, "Modeling strength of high-performance concrete using genetic operation trees with pruning techniques," *Computers and Concrete*, vol. 6, no. 3, pp. 203-223, 2009.

[10] Y. C. Tsai "Artificial intelligence techniques in the prediction of rainfall during typhoons in shihmen reservoir," M.S. dissertation, Dept. Civil Engineering, Chung Hua University, Hsinchu, TW, 2010.

[11] C. H. Chiu, "The application of artificial intelligence optimization model in predicting typhoon rainfall using northward-moving typhoons as examples," M.S. dissertation, Dept. Civil Engineering, Chung Hua University, Hsinchu, TW, 2011.

[12] B. Alabbadi, L. Chen, C. Kou, and S. Ma, "Applying genetic algorithm combining operation tree (GAOT) for estimating typhoon precipitation using passive microwave satellite," in *Proc. 2012 International Conference on Theoretical and Mathematical Foundations of Computer Science*, 2012, pp. 402-407.

[13] K. T. Chen, L. Chen, Y. R. Fu, and S. H. Chen, "Estimating strength of concrete using a genetic algorithm combining operation tree (GAOT)," *Applied Mechanics and Materials*, vol. 325-326, pp. 1293-1296, 2013.

[14] L. Chen, H. C. Yeh, C. C. Wei, J. C. Huang, and Y. R. Fu, "Combining genetic algorithm and operation tree model for improving the radar-based rainfall estimation-a case study for typhoon rainfall in shih-men," *Journal of Taiwan Agricultural Engineering*, vol. 59, no. 4, pp. 67-80, 2013.

[15] J. Garnier, J. Nemery, G. Billen, and S. Thery, "Nutrient dynamics and control of eutrophication in the marne river system: modelling the role of exchangeable phosphorus," *Journal of Hydrology*, vol. 304, pp. 397-412, 2005.

[16] J. E. G. *Raymont, Plankton and productivity in the oceans, Phytoplankton*, , New York: Pergamon Press, 1980, p. 496.

[17] M. S. Wong, K. H. Lee, Y. J. Kim, J. E. Nichol, Z. Li, and N. Emerson, "Modeling of suspended solids and sea surface salinity in hong kong using aqua/MODIS satellite images," *Korean Journal of Remote Sensing*, vol. 23, no. 3, pp. 161-169, 2007.

[18] T. T. Danbara, "Deriving water quality indicators of lake tana, ethiopia, from landsat-8," M.S. dissertation, Dept. Science in Geo-information Science and Earth Observation, University of Twente, Netherlands, 2014.

[19] N. Laili, F. Arafah, L. M. Jaelani, L. Subehi, A. Pamungkas, E. S. Koenhardono, and A. Sulisetyono, "Development of water quality

parameter retrieval algorithms for estimating total suspended solids and chlorophyll-a concentration using landsat-8 imagery at poteran island water," in *Proc. 2015 International Geoinformation Conference*, 2015.
[20] M. Mitchell, *An Introduction to Genetic Algorithms*, Massachusetts: MIT Press, 1996.
[21] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, New York: Springer-Verlag, 1995.
[22] L. Lien, I. Yeh and M. Cheng, "Modeling strength of high performance concrete using genetic algorithms and operation tree," *Journal of technology*, vol. 21, no.1, pp .41–54, 2006.

Li Chen is a professor in the Department of Civil Engineering Department of Chung Hua University in Taiwan. Her current research interests include environmental impact assessment, climate change, water resources planning, soil and water conservation, remote sensing and artificial intelligence.

Mohammad Jamal Mtlak Abualghanam is currently a PhD student at Civil Engineering Department in Chung Hua University under the supervision of Prof. Li Chen. His research algae estimation in a reservoir using remote sensing data.

Chih-Hung Tan is currently a courtesy assistant professor and a head of Department of Agricultural Engineering Research Center in Taiwan. He has a PhD from Agricultural and Biological Engineering Department of University of Florida. His expertise in Environmental Science, Waste Management, Irrigation and Water resources.

Basmah Alabbadi did her PhD in estimation sea surface salinity by MODIS data in the Department of Civil Engineering at Chung Hua University under the supervision of Prof. Li Chen.