行政院國家科學委員會專題研究計畫 成果報告

應用遙測技術與進化演算法於水庫水質即時監測之研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 95-2221-E-216-013-
執 行 期 間 ：95 年 08 月 01 日至 96 年 07 月 31 日
執 行 單 位 ：中華大學土木與工程資訊學系

計 畫 主 持 人：陳莉

計畫參與人員：博士班研究生-兼任助理：王泰盛
碩士班研究生-兼任助理：梁惠儀、吳明銘

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 96 年 10 月 30 日

行政院國家科學委員會補助專題研究計畫■ 成 果 報 告 □期中進度報告

應用遙測技術與進化演算法於水庫水質即時監測之研究

計畫類別：■ 個別型計畫　　□ 整合型計畫
計畫編號：NSC 95 － 2221 － E － 216 － 013 －
執行期間： 95 年 08 月 01 日至 96 年 07 月 31 日

計畫主持人：陳莉
共同主持人：
計畫參與人員：王泰盛、梁惠儀、吳明銘

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：中華大學土木與工程資訊學系

中　華　民　國 九十六 年 十 月 二十九 日

摘要

平行式 GEGA 是結合 GE 和平行的 GA 所建構而成，目的是為了改善應用遙測影像進行水庫水質的監測。翡翠水庫是台北主要供給水來源，我們沿著翡翠水庫縱向航行，並進行 Chl-a 濃度之地面真值採樣。經驗式方程式的建構是來自多光譜的 Landsat 7 ETM+之資料。GE 是一個自動進化演算程式，它會在 Chl-a 觀測數據與影像數據之間自動找到複雜非線性的關係式。在搭配使用 GA 和 GE 之後，可找到最佳的合適方程式類型，而多個平行式 GA 在最佳化程序期間可以增進蒐尋的效率，經由 GEGA 和 LMR 之比較，可以發現前者為較好的模式且估算誤差值較低。

關鍵詞：進化式語法、平行遺傳演算法、水質監測、葉綠素 a、遙測影像

Abstract

Parallel GEGA was constructed by incorporating grammatical evolution (GE) into the parallel genetic algorithm (GA) to improve reservoir water quality monitoring based on remote sensing images. A cruise was conducted to ground-truth chlorophyll-a (Chl-a) concentration longitudinally along the Feitsui Reservoir, the primary water supply for Taipei City in Taiwan. Empirical functions with multiple spectral parameters from the Landsat 7 Enhanced Thematic Mapper (ETM+) data were constructed. The GE, an evolutionary automatic programming type system, automatically discovers complex nonlinear mathematical relationships among observed Chl-a concentrations and remot sensed imageries. A GA was used afterward with GE to optimize the appropriate function type. Various parallel subpopulations were processed to enhance search efficiency during the optimization procedure with GA. Compared with a traditional linear multiple regression (LMR), the performance of parallel GEGA was found to be better than that of the traditional LMR model with lower estimating errors.

Keywords：Grammatical evolution, Parallel genetic algorithm, Water quality monitoring, Chlorophyll-a, Remote-sensed imagery.

1. 前言

台灣地處亞熱帶降雨強度大，加上水庫集水區地形陡峭、及人爲活動頻繁等因素，常使我們水庫面臨濁度升高及優養化的問題。

因為翡翠水庫集水區位處台灣東北部，此為大台北地區重要的民生用水，所以中央研究院－環境變遷中心，對於翡翠水庫之採樣點高達 9 站，採樣頻率約為兩星期一次，使得數據更為豐富客觀，可加深分析探討的準確性，因此本研究以翡翠水庫來當一個研究案例，可將成果推廣至台灣地區其他重要水庫。

水質監測都大多是以現地採樣的傳統方式，所以會受到許多限制，現地採樣所花費的大量人力及物力就是一個很明顯的例子；而且許多水庫由於人力不足，無法進行有效現地採樣水質，因此監測頻率也無法達到密集的資訊，且所選取的採樣位置也受限於人員所能到達之地區；反之，衛星影像可大範圍的反應出集水區的現況（時間）與實況（空間），無需到達現地勘查，可節省可觀的人力及資源，也可突破現地採樣的一些限制，更可成為一個快速且大範圍的水質監測方法。

2. 研究目的

This sandy is to improve the monitoring techniques of using remote sensed data to estimate the water quality parameters. Because of the complex nonlinear relationship between several bands of satellite and chlorophyll-a concentration in reservoir, a new system identified method called parallel GEGA is presented at the first time. We use the parallel structure to GEGA to increase the diversity of solutions by GA for obtaining the optimal equation between one remote sensed imagery and water quality parameter efficiently. Parallel GEGA was constructed by incorporating grammatical evolution (GE) into the parallel genetic algorithm (GA) to improve reservoir water quality monitoring based on remote sensing images. A cruise was conducted to ground-truth chlorophyll-a (Chl-a) concentration longitudinally along the Feitsui Reservoir, the primary water supply for Taipei City in Taiwan.

3. 文獻回顧

Chlorophyll-a is used as an indication of the intensity of algae growth and is one of the major factors affecting water quality which can produce visible changes in the surface of waters (Ritchie et al., 1990). Observed Chlorophyll-a concentrations can help to characterize the trophic state of an aquatic ecosystem through various numerical schemes (Carlson, 1977). Its concentrations in surface waters have traditionally been measured spectrophotometrically after samples were collected, preserved and transported to the laboratory (APHA, 1992; Lind, 1985). Though this approach is well accepted and widely utilized, the labor-intensive and time-consuming field works do not allow researchers to construct contemporaneous Chlorophyll-a maps at full spatial scale (Allee & Johnson, 1999). Accordingly, to characterize the spatial variation of trophic state within vast water body by using a limited number of field data is often argued.

To improve traditional data collection method, utilization of remote sensing data for water quality assessment has been investigated. The application of remote sensing to assess freshwater production has escalated recently due to its capability of scanning wide water body within a short time period (Harrington et al., 1992). Water quality assessment using satellite data has been carried out since the first remote sensing satellite Landsat- Multi-Spectral Scanner (MSS) became operational (Thiemann & Kaufmann, 2000).

Although the Landsat Thematic Mapper (TM) sensor, which provides the longest continuous dataset of high-spatial-resolution imagery of Earth, is able to present a synoptic monitoring of water quality problems, its quantitative use is still a difficult task (Dekker et al., 2002). In addition, chlorophyll-a is one of the most important variables of great ecological significance examined by many researchers in this discipline, yet, it is probably the most complicated one to map correctly, at least if a universal type of relation is desired (Zhang et al., 2003).

Statistical applications are used traditionally to establish algorithms for predictions of various water quality variables. Regression analysis is popular in formulation of predictive

models (Allee & Johnson, 1999). However, nonlinear transfer functions are often found when relate water quality variables to the satellite imagery observations (Zhang et al., 2003). Generally, sophisticated regression models must go through time-consuming trail and error procedures thus the correct regression type can be obtained. Evolutionary algorithms, such as genetic programming (GP), have been used with much success 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 like the back-propagation network (BPN) (Kishore et al., 2000). Nevertheless, it is well known that the BPN is considered as a non-linear black-box model, and it is not unusual for it to be criticized as not enhancing our understanding of the physical mechanisms because of its complex weighting coefficients and numerous other parameters.

Chen (2003) pointed out that constructing the data structure of dynamic tree of the genetic programming (GP) could be a difficult task while applying GP to estimate the reservoir trophic state using remote sensed data. For example, it is hard to choose the proper size of a tree which can express a meaningful equation in advance. On the other hand, the newly developed grammatical evolution (GE) performs the evolutionary process on a variable-length binary string which is more flexible than a dynamic tree structure. A mapping process is employed to generate programs in any language by using the binary strings to select production rules (O'Neill & Ryan, 2001).

4. 研究方法

4.1 GA

Genetic algorithms (GAs) originated in the mid 1970s (Holland, 1975) have been developed into powerful optimization approach. A principle difference between optimization using a GA versus more traditional methods is that the decision space is searched from an entire population of potential designs. According to a couple of our previous works, real-coded GAs have advantages over binary coded GAs (Chang and Chen, 1998; Chang et al., 2005). Hence, in this study, we only considered real-coded GAs.

4.1.1 GA with parallel structure

The conventional GA is likely to be trapped in a region that does not contain the global optimum. This problem, called premature convergence, has been recognized as a serious failure mode for GAs (Eshelman & Schaffer, 1991). A multi-population GA divides a single population into smaller subpopulations. Each subpopulation evolves by genetic operations in parallel with the other, while maintaining a limited but powerful interaction between all subpopulations. The most important advantage of subpopulations is the enhanced diversity among the subpopulations (Herrera and Lozano, 2000).

4.1.2 Topology of the parallel GA

A hypercube topology distributed GA, called HDGA, with three dimensions was presented in Chen and Chang (2006), as shown in Fig. 1. Each subpopulation uses three genetic operators including linear ranking selection, blend crossover and Gaussian mutation. The topology includes two important different sides. The front side is devoted to exploration. There are four subpopulations E1,…, E4, to which exploratory crossover operators are applied. The exploratory degree increases clockwise, starting at the lowest E1, and ending at the highest E4. The other side (the rear side) is for exploitation. There are four subpopulations e1,…, e4, and exploitative crossover operators are used. The exploitation degree increases clockwise, starting at the lowest e1, and finishing at the highest e4. With this structure, a parallel multi-resolution is obtained by using the crossover operation, which allows a diversified search (reliability) and an effective local tuning (accuracy) to be achieved simultaneously (Fig. 1a).

4.1.3 Refinement and expansion

In order to improve the reliability and accuracy, two effects are introduced. The first effect is "refinement" (Fig. 1b), which makes migrations form an exploratory subpopulation toward

exploitative one, i.e., from Ei to ei, or between two exploratory subpopulations from a higher degree to a lower one, i.e., from Ei+1 to Ei, or between two exploitative subpopulations from a lower degree to a higher one, i.e., from ei to ei+1. The second effect is "expansion" (Fig. 1c), which makes migrations in the opposite direction.

4.1.4. Migration

An emigration model is one in which migrants are sent only toward immediate neighbors along a dimension of the hypercube, and each subsequent migration takes place along a different dimension of the hypercube. Particularly, the best individual of each subpopulation is sent toward the corresponding subpopulation every five generations, as shown in Fig. 2. The sequence of application is from left to right, i.e., first, the refinement migrations, second the refinement/expansion migrations, third the expansion migrations, and then, the sequence starts again.

4.1.5 Parallel GA incorporated with GE

Fig. 3 shows a combination of GE and GA, called GEGA, to generate the optimal relationship among inputs and outputs automatically. First, a GE was employed to transfer the real-coded string through BNF grammars to mathematical function. The data from several bands of remotely sensed imagery were used in the GE as inputs to estimate the water quality in the reservoir. Further, a GA was incorporated with this GE to optimize the objective value of those functions. In other words, the GA was used as a search strategy to determine the most proper relationship among the remotely sensed data and water quality. Finally, this GEGA was implemented as a parallel structure to improve the searching efficiency and prevent the premature convergence during the optimization.

4.2 BNF

BNF is a notation for expressing the grammar of a language in the form of production rules (Naur, 1963). BNF grammars consist of so called terminals, which are items that can appear in the language, e.g., +, -, etc., and nonterminals, which can be expanded into one or more terminals and nonterminals. A grammar can be represented by the tuple {N、T、P、S}, in which N is the set of nonterminals, T the set of terminals, P a set of production rules that maps the elements of N to T, and S is a start symbol that is a member of N. When there are a number of productions that can be applied to one particular N, the choices are delimited with the '|' symbol.

Below is an example BNF, where

N = {expr, op, pre_op}

T = {Sin, Cos, Log, +, -, *, /, Variable X, Constant 1.0}

S = <expr>

And P can be represented as

(1) <expr> :: = <expr><op><expr> ……..rule 0

| (<expr><op><expr>)…….rule 1

|<pre-op> (<expr>)……...rule 2

| <var>……...rule 3

(2) <op> :: = + ……..…..rule 0

| - ……..….rule 1

| / ……..….rule 2

|* ……..….rule 3

(3) <pre-op> :: = Sin ……..…rule 0

| Cos ……rule 1

| Log ……rule 2

（4）<var> :: = X ……rule 0

| 1.0 ……rule 1

4.2.1 Mapping Process(映射處理)

The genotype is used to map the start symbol onto terminals by reading codons of 8 bits to generate a corresponding integer value from which an appropriate production rule is selected by using the following mapping function:

Rule = (codon integer value) MOD (number of rules for the current nonterminal) (1)

Considering the following rule, i.e., given the nonterminal op, there are four production rules to select from:

<op> :: = + …… rule 0

| - …… rule 1

| / …… rule 2

|* …… rule 3

If we assume the codon being read produces the integer 6, then

6 MOD 4 = 2

would select <op> as rule 2: /. Each time a production rule has to be selected to map from a nonterminal, another codon is read. In this way, the system traverses the genome.

For example, consider the individual in Table 1. There are fourteen 8-bit binary condons in one string. The decoding process is described as follows.

(1) First, concentrating on the start symbol <expr>, we can see that there are four productions to choose from. To make this choice, we read the first codon from the chromosome “11001000” and use it to generate a number “200”. Because the standard decode of the binary 11001000 is , $1\times 2^7+1\times 2^6+0\times 2^5+0\times 2^4+1\times 2^3+0\times 2^2+0\times 2^1+0\times 2^0$ which equals to 200. This number will then be used to decide which production rule to use according to Eq. (1) in BNF. Thus, we have 200 MOD 4 = 0, meaning we must take the zeroth production, rule (0), so that <expr> is now replace with

<expr><op><expr>.

(2)Continuing with the first <expr>, i.e., always starting from the leftmost nonterminal, a similar choice must be made by reading the next codon value 160 and again using the given formula we get 160 MOD 4 = 0，i.e., rule 0. The leftmost <expr> will now be replaced with <expr><op><expr> to give

<expr><op><expr><op><expr>

(3)Again, we have the same choice for the first <expr> by reading the next codon value 206, the result being the application of rule 2 to give

<pre-op>(<expr>)<op><expr><op><expr>.

(4)Now, the leftmost <pre-op> will be determined by the codon value 96 that gives us rule 0, which is <pre-op> becomes Sin. We have the following:

Sin(<expr>)<op><expr><op><expr>

.

(14)The mapping continues until eventually we are left with the following expression:

Sin(X)*Cos(X)+1.0

Notice that if there had been any extra codons, they would have been simply ignored during the genotype-to-phenotype mapping process. It is possible for individuals to run out of codons and in this case, we wrap the individual and reuse the codons. This technique of wrapping the individual draws inspiration from the gene-overlapping phenomenon, which has been observed in many organisms (Elseth & Baumgardner, 1995). It is possible that an incomplete mapping could occur even after several wrapping events and in this case, the individual in question is given the lowest possible fitness value.

Because there is a problem that only integers can be presented by using the binary coding scheme, we revised it as a real-coded representation. The real numbers which imply that, each chromosome is a real-valued vector, as opposed to binary-coded GA, where chromosomes are 0-1 vectors. It is very useful and efficient to generate the real-number constants and coefficients shown in these output equations. Whereas mapping chromosomes to the BNF rules, just need turn out them to be integers. This revision makes GE be combined with a real-coded GA very easily described as follows.

5. A case study in an oligotrophic/mesotrophic reservoir

As mentioned earlier, the Landsat 7 Enhanced Thematic Mapper (ETM+) data was used to map the spatial distribution of surface chlorophyll-a concentrations (Chl-a) in an oligotrophic/mesotrophic reservoir, Feitsui, in Taiwan. The relationship between observed Chl-a and corresponding image data was constructed by using a parallel GEGA presented above. This system identification problem may be viewed as a search for an optimal function type, which maps input values of ETM+ bands onto an output value of water quality such as chlorophyll-a. The traditional method of linear multi-regression (LMR) was also analyzed for comparison.

5.1 The study area- Feitsui Reservoir

The Feitsui Reservoir at 25o27' N and 121o33' E is the most important reservoir of northern Taiwan, supplying drinking water for more than four million people in the Taipei City (Fig. 4). There are eight regular sampling stations in the reservoir and gathering water quality data once a month. Although water quality in Feitusi Reservoir is the best in Taiwan, the reservoir is still receives much attention because of significant watershed nutrient load (Kuo et al., 2006). The ground-truthing was conducted on April 18, 2005 since April is usually the first peak of Chl-a occurred every year.

5.2 Ground-truth observationGround-truth observation

Total 24 surface water (0-0.8m) samples longitudinally along the Feitsui Reservoir (Fig. 4) were analyzed for Chl-a (μg/L) before the noon time when the Landsat satellite overpass. Each sampling site was geographically located using a Global Positioning System (Garmin, GPS 100 SURVEY Ⅱ) with 2 m of precision. The Chl-a was measured by fluorometery method (Turner Design 10-AU-005) after acetone extraction. The minimum and maximum values of the 24 sampling are 0.48 and 4.02 (μ g/L), respectively, with a detection limit of 0.1 (μg/L).

5.3 Concurrent remote sensed data

To quantitatively measure water quality in the study area, the Landsat 7 ETM+ image of 18 April 2005 was selected to match the simultaneous mission of water quality sampling at 24 points in the Feitsui Reservoir. The spatial resolution of ETM+ data is 30 m (except band 6 of thermal infrared channel with 120 m). In addition, ortho-rectified air photos of the Feitsui Reservoir area with spatial resolution of 25 cm, which have been rectified to the 2 degree Transverse Mercator (TM2) coordinate system commonly adopted in Taiwan, were also used for geographic rectification of the Landsat 7 ETM+ image and for a detailed land use/cover of the surroundings

of the reservoir.

The Landsat 7 ETM+ image was acquired from the USGS National Center for Earth Resources Observation and Science (EROS). Due to an instrument malfunction occurred onboard Landsat 7, all image data acquired by the Landsat 7 ETM+ from July 14, 2003 has been collected in Scan Line Corrector turned off (SLC-off) mode (USGS, 2007). We acquired the image in level 1G SLC-off mode, that is radiometrically corrected, USGS systematically replace the duplicated image data caused by the SLC failure. The Feitsui Reservoir was located in the center of the scene that SLC-off had very little affect to the image quality since the SLC-off effects are most pronounced along the edge of the scene and gradually diminish toward the center (USGS, 2007). Since analysis was made for a single image with a quite small angular range, the atmospheric correction has little effect on correlation analysis (Zhang et al., 2003). Thus, the atmospheric correction was ignored for this clear image. The image was geometrically rectified to the TM2 coordinate system with ortho-rectified air photos. In order to extract the TM data at the water sampling locations, the mean value in a 3 by 3 window of the image was used to represent the ground point.

Landsat 7 ETM+ bands 1, 2, 3, 4, 5, 7 are optical bands, recording electro-magnetic radiation from 0.45 to 2.35 micrometers, providing visible, near infrared, to mid-infrared responses of the land surface. The properties of these six bands were shown in Table 2. Landsat 7 ETM+ band 6 is unique among the various bands in that it shows the emitted radiation from a surface in the thermal region of the spectrum, it is not suitable to perform classification with other bands. Allee and Johnson (1999) used band 1 to 5 and 7 of Landsat 7 ETM+ to estimate surface chlorophyll a for each sampling site in Bull Shoals Reservoir, USA, which is also classified as an oligo-mesotrophic level of productivity. Those are the same input bands used in this study based on similar trophic states on both reservoirs. Besides, the correlation coefficient, R, between the different ETM+ bands and water quality parameters of Chl-a ranged from 0.487 to 0.740 are all obviously significant, especially for band 4 and 7. Therefore, all these six bands 1~5 and 7 except band 6 were used as input variables to estimate Chl-a.

The image classification was performed by the unsupervised classification algorithm named Iterative Self-Organizing Data Analysis Technique (ISODATA) using software ERDAS Image v8.6. The classification result classified the Chl-a of water body as six levels in this reservoir was shown in Fig. 5. Most of these pixels were classified into the sixth level (the highest Chl-a values) located on the middle areas of reservoir.

5.4 Estimation the chlorophyll-a of reservoir

5.4.1 Using Linear Multi-regression Method

To map the spatial variation of water quality parameters in the reservoir with remotely sensed images, empirical relationship between digital numbers of the preprocessed image bands and Chl-a was established using the simplest type of linear multi-regression (LR) method initially.

This model utilized ETM+ bands 1 to 5, and 7 was given by:

$$Chl-a = 4.483 + 0.022(B1) + 0.031(B2) - 0.041(B3) - 0.13(B4) + 0.108(B5) - 0.235(B7) \quad (2)$$

Then, the result of using ETM+ bands through logarithmic transformation was also presented as follows:

$$LNChl-a = 17.239 - 2.593LN(B1) + 0.111LN(B2) + 0.191\ln(B3) - 0.961LN(B4) + 0.347LN(B5) - 1.473LN(B7) \quad (3)$$

The correlation coefficients, R, of Equ. (2) and Equ. (3) are 0.823 and 0.765 respectively. The other statistical parameters including the sum of square errors (SSE), and root mean square errors (RMSE) of the above two equations are shown in Table 3. Obviously, the transformation did not improve the retrieval accuracy than that of simple regression analysis significantly. 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 GEGA to fit the complex

nonlinear transfer function between the ETM+ bands and water quality parameters.

5.4.2 Using Parallel GEGA

The same data were used to compare with the traditional regression method described above. We add "EXP" operator to the tuple of terminals of BNF grammars in GE to generate the equations. In parallel GA, there are total eight different subpopulations, including four exploratory subpopulation E1,…, E4, and four exploitative subpopulation e1,…, e4 to construct the topology as a three dimensional hypercube. Each subpopulation size was set to be 50 and use different degree of exploratory/ exploitative blend crossover operators to optimize the equations of six bands and Chl-a generating by GE separately. The objective function can be written as

$$Minimize \quad RMSE = \left[\sum_{i=1}^{n} \left(o_i - e_i \right)^2 / n \right]^{1/2} \tag{4}$$

where oi is the actual value of Chl-a, ei is the estimated value by GEGA, and n is the total number of the ground sampling data (=24).

The optimal (best) fitness was found at the end of each generation and these on-line behaviors (the best objective value at each generation) of eight subpopulations are demonstrated in Table 4. All subpopulations are converging to the optimal solution 2.1673 within 300 generations. It also shows that the optimal solutions of exploration subpopulations E1…E4 are more diverse than the exploitation subpopulations e1…e4 at the early stages. At the later stages, most subpopulations are converged because of the effect of migrations.

The optimal relationship among remotely sensed imageries and chlorophyll-a concentration (Chl-a) was acquired through parallel GEGA which can be represented as Equ. (5).

$$Chl-a = \left\{ LN\left[e^{\left(15.765706 \times \left(\frac{60.304886}{e^{\frac{B7}{B1}}} \right) \right)} + 31.6022906 - B_4 \right] + \left[LN\left(\frac{B_1}{95.110605} \right) \right] \times \left(9.884309 - \left(\frac{B5 \times B7}{26.271711} \right) \right) \right\} \tag{5}$$

There are only four bands B1, B4, B5 and B7 automatically combined in this nonlinear equation through a lot of generations' evolutions and competitions. These coefficients and forms in the above equation are found to be the optimal solutions based on the balance between the complexity of equation and the number of input bands. The final results are also shown in Table 3.

It indicates that the correlation coefficient R=0.89 and the determination coefficient R2 =0.79 of parallel GEGA is better than those of the two types of linear multiple regression. Moreover, the sum of square errors SSE=2.17 and root mean square errors RMSE=0.30 of parallel GEGA is lower than those of the other methods. Parallel GEGA was found better than the traditional LMR model for chlorophyll-a concentration estimation as indicated by the higher correlation coefficient and lower error between the observed data and the estimated value of models.

We checked the domain of these four input variables in Equ. (5) by generating the map of whole water body in reservoir, the output values of Chl-a were reasonable which was described in the following section.

5.5 Spatial distribution of chlorophyll-a

In order to realize the chlorophyll-a concentrations of broad area of reservoir, the 6677 pixels of the Landsat image covering the whole water body in reservoir will be efficient tools. In this study, based on the LMR and parallel GEGA, spatial distribution of Chl-a of the Feitusi Reservoir are demonstrated in Fig.s 6 and 7. It is shown that the reservoir was mesotrophic at the central and lower areas while oligotropic in the upper places; this phenomenon has been observed

by both methods. Besides, the tendency of distributions of Chl-a were coincident with the classification imagery (Fig.5) described in section 4.3. However, the maximum and minimum values of Chl-a generated by the parallel GEGA are closer to the ground observations than those of the LMR. The statistical properties of the 6677 estimated Chl-a values by these two methods are shown in Table 5. The trophic state of the water body in reservoir had a wider representation by the parallel GEGA than LMR.

6. 結論與建議

The main contribution of this paper is to provide a new parallel GEGA algorithm, which creates potentials to monitor chlorophyll-a level in specific time frame and reservoir. It can deal easily with nonlinear transfer problems between remote sensed imagery and water quality in reservoir and is shown to be very efficient and robust optimization tool. However, variability in species structure of phytoplankton assemblage may generate different optical spectrum, even in Feitsui Reservoir, there are slight seasonal changes in phytoplankton community. Therefore, through all the procedures described above in the text including the field data collection and remote sensed imagery process, this newly developed method is flexibly applied to other reservoirs.

7. 參考文獻

1. Allee, R.J. and Johnson, J.E. (1999) Use of satellite imagery to estimate surface chlorophyll a and secchi disc depth of bull shoals reservoir, Arkansas, USA. International Journal of Remote Sensing 20 (6), 1057-1072.
2. APHA (American Public Health Association) (1992) Standard methods for the examination of water and wastewater. 16th edition (Washington, DC: American Public Health Association).
3. Carlson, R.E. (1977) A trophic state index for lakes. Limnology and Oceanography 22 (2), 361-369.
4. Chang, F.J. and Chen, L. (1998) Real-coded genetic algorithm for rule-based flood control reservoir management. Water Resource Management 12, 185-198.
5. Chang, F.J. Chen, L. and Chang, L.C. (2005) Optimizing the reservoir operation rule curves by genetic algorithms. Hydrological Processes 18, 2277-2289.
6. Chen, L. (2003) A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data. International Journal of Remote Sensing 24 (11), 2265-2275.
7. Chen, L. and Chang, F.J. (2006) Applying real-coded multi-population genetic algorithm to multi-reservoir operation. Hydrological Processes 21, 688-698.
8. Dekker, A.G. Vos, R.J. and Peters, S.W.M. (2002) Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. International Journal of Remote Sensing 23, 15-35.
9. Elseth, G.D. and Baumgardner, K.D. (1995) Principles of Modern Genetics. St. Paul, MN: West.
10. Eshelman, L.J. and Schaffer, J.D. (1991) Preventing premature convergence in genetic algorithms by preventing incest. in proc 4th International Conference Genetic Algorithms. R. Belew and L. B. Booker. Fads. San Marco, CA, Morgan Kaufmann.115-122.
11. Harrington, J.A. Schifebe, F.R. and Nix, W.E. (1992) Determination of phytoplankton chlorophyll concentrations in the Chesapeake Bay with aircraft remote sensing. Remote Sensing of Environment 40, 79-100.
12. Herrera, F. and Lozano, M. (2000) Gradual distributed real-coded genetic algorithms. IEEE Transactions on Evolutionary Computation 4 (1), 43-63.
13. Holland, J.H. (1975) Adaptation in natural and artificial systems, Ann Arbor, MI, The University of Michigan Press.
14. Kishore, J.K. Patnaik, L.M. Mani, V. and Agrawal, V.K. (2000) Application of genetic programming for multicategory pattern classification. IEEE Transactions on Evolutionary Computation 4 (3), 242-257.

15. Kuo, J.T. Wang, Y.Y. and Lung, W.S. (2006) A hybrid neural-genetic algorithm for reservoir water quality management. Water Research 40, 1367-1376.
16. Naur, P., 1963. Revised Report on the Algorithmic Language ALGOL 60. Commun. ACM. 6 (1), 1-17.
17. O'Neill, M. and Ryan, C. (2001) Grammatical evolution. IEEE Transactions on Evolutionary Computation 5 (4), 349-357.
18. Ritchie, J.C. Cooper, C.M. and Schiebe, F.R. (1990) The relationship of MSS and TM digital data with suspended sediments, chlorophyll, and temperature in Moon Lake, Mississippi. Remote Sensing of Environment 33, 137-178.
19. Thiemann, S. and Kaufmann, H. (2000) Determination of chlorophyll content and trophic state of lakes using field spectrometer and IRS-1C satellite data in the Mecklenburg Lake District, Germany. Remote Sensing of Environment 73, 227-235.
20. Zhang, Y. Pulliainen, J.T. Koponen, S. S. and Hallikainen, M.T. (2003) Water quality retrieval from combined Landsat TM data and ERS-2 SAR data in the gulf Finland. IEEE Transactions on Geoscience and Remote Sensing 41 (3), 622-629

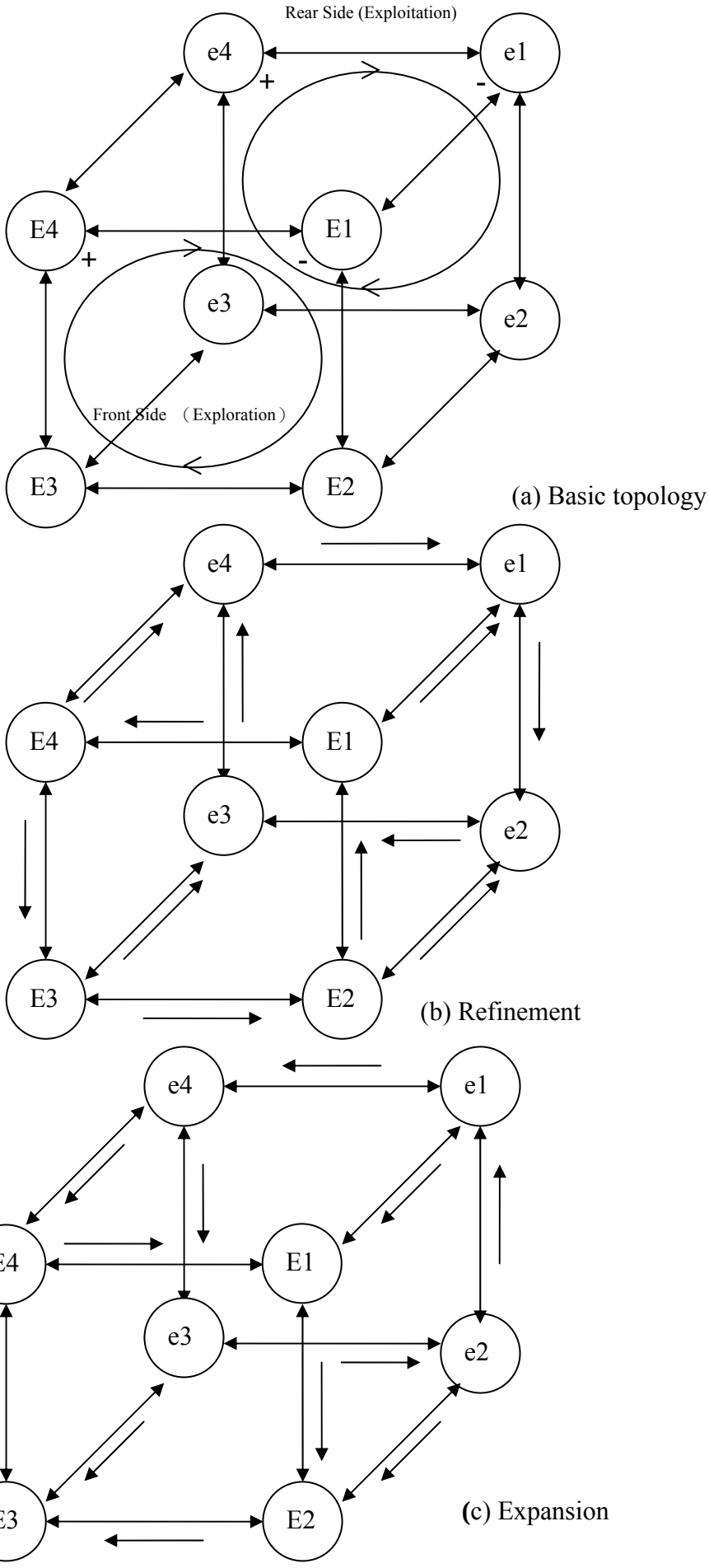

Fig. 1. Structure of an HDGA: (a) basic topology; (b) refinement; (c) expansion.

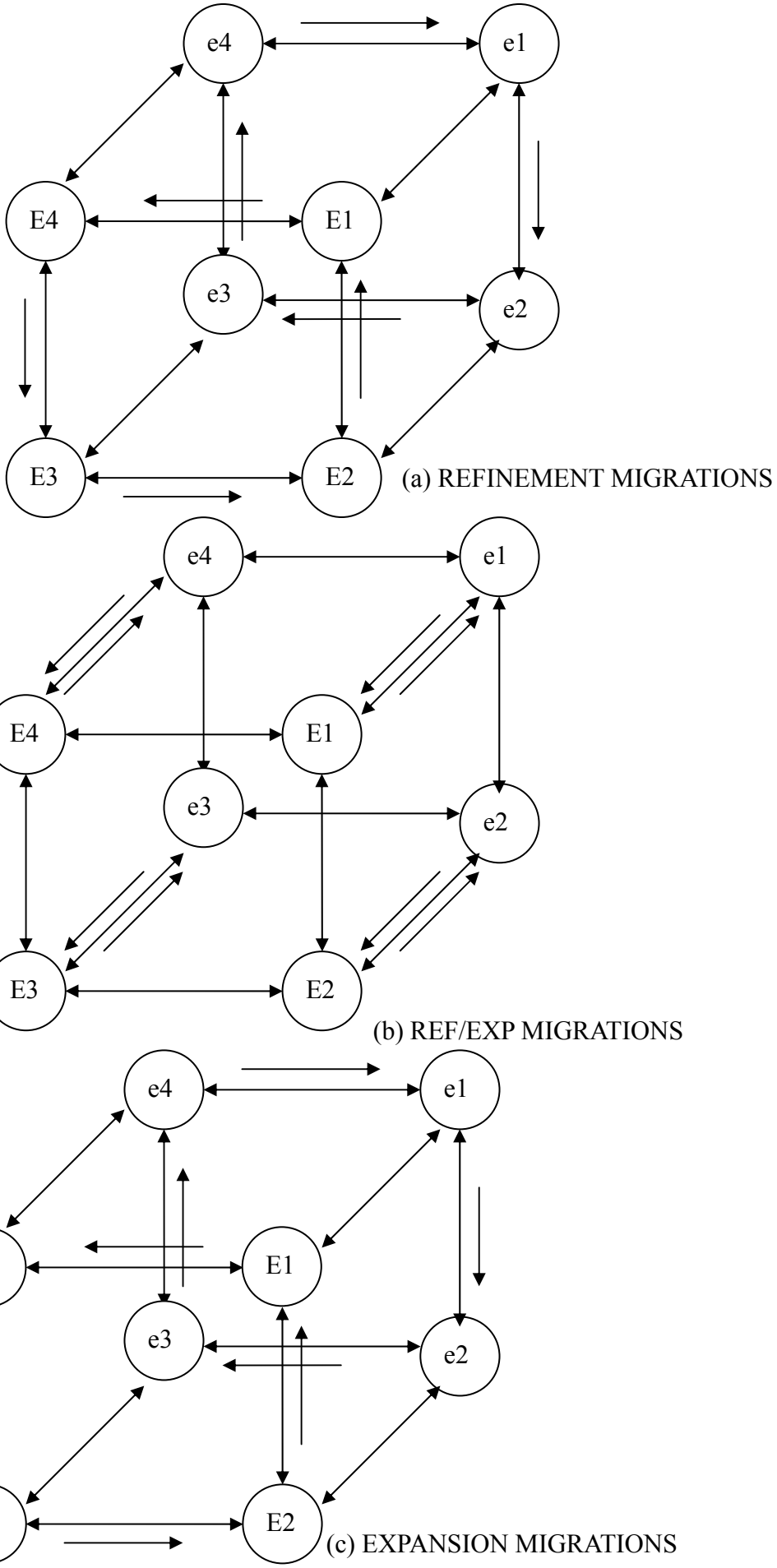

Fig. 2. Three types of migration in an HDGA: (a) refinement migrations;(b) ref/exp migrations; (c) expansion migrations

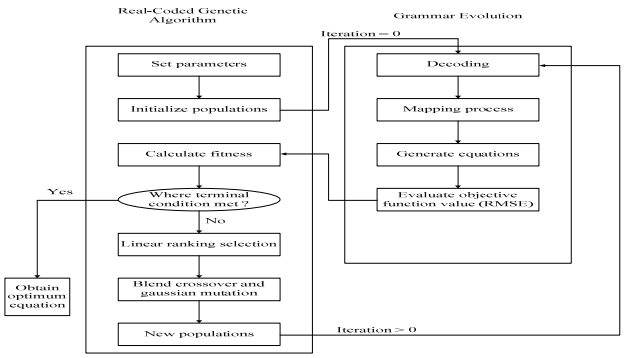

Fig. 3. The flowchart of GE combined with GA

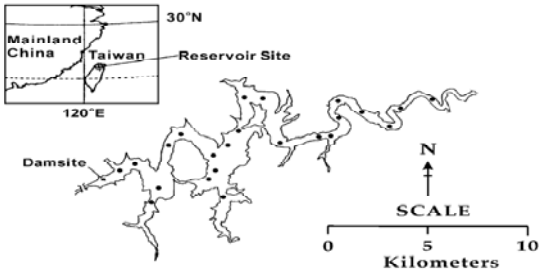

Fig. 4. The 24 sampling sites of Feitsui Reservoir in Taiwan

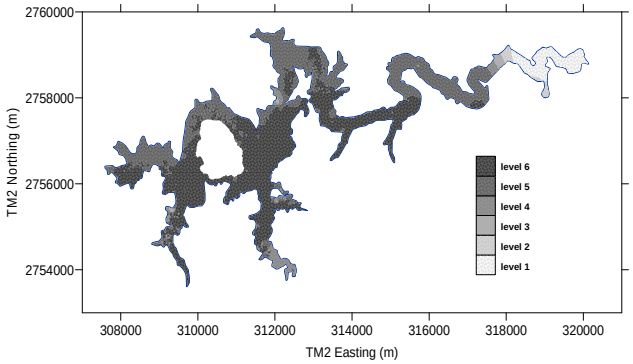

Fig. 5. Six levels of chlorophyll-a concentration by ISODATA classification

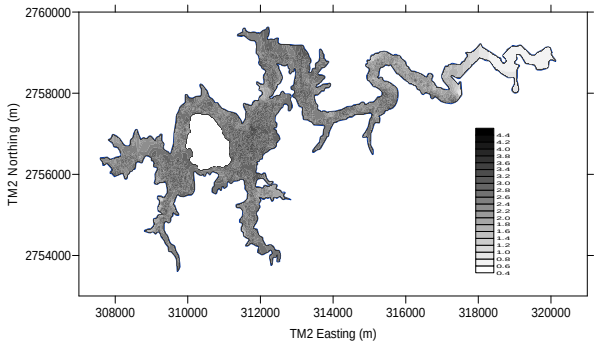

Fig. 6. Chlorophyll-a concentration distribution of the reservoir by LMR (Equ.(2))

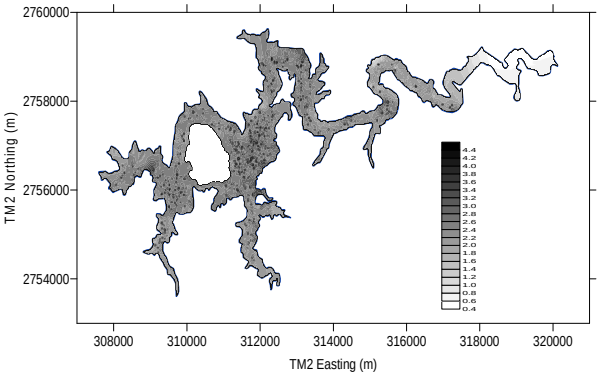

Fig. 7. Chlorophyll-a concentration distribution of the reservoir by parallel GEGA (Equ.(5))

Table1. Example of each codon converted into corresponding BNF grammar

No.	8-bit binary codon	Integer value	Mapping function	BNF grammars
1	11001000	200	200 MOD 4 = 0	<expr><op><expr>
2	10100000	160	160 MOD 4 = 0	<expr><op><expr><op><expr>
3	11001110	206	206 MOD 4 = 2	<pre-op>(<expr>)<op><expr><op><expr>
4	01100000	96	96 MOD 3 = 0	Sin(<expr>)<op><expr><op><expr>
5	00011011	27	27 MOD 4 = 3	Sin(<var>)<op><expr><op><expr>
6	01001000	72	72 MOD 2 = 0	Sin(X)<op><expr><op><expr>
7	01101011	107	107 MOD 4 = 3	Sin(X)*<expr><op><expr>
8	00111110	62	62 MOD 4 = 2	Sin(X)*<pre-op>(<expr>)<op><expr>
9	00010110	22	22 MOD 3 = 1	Sin(X)*Cos(<expr>)<op><expr>
10	00110111	55	55 MOD 4 = 3	Sin(X)*Cos(<var>)<op><expr>
11	01011000	88	88 MOD 2 = 0	Sin(X)*Cos(X)<op><expr>
12	01100100	100	100 MOD 4 = 0	Sin(X)*Cos(X)+<expr>
13	11001011	203	203 MOD 4 = 3	Sin(X)*Cos(X)+<var>
14	00101001	41	41 MOD 2 = 1	Sin(X)*Cos(X)+1.0

Table 2. The properties of Landsat 7 ETM+ data of the Feitsui Reservoir on April 18, 2005

Band	Spectral Wavelength (micrometers)	Range of Digital Numbers on 24 Sampling sites	Range of Digital Numbers on whole water body
B1 (Blue)	0.45~0.52	91~103	83~103
B2 (Green)	0.52~0.60	61~69	56~72
B3 (Red)	0.63~0.69	40~52	38~57
B4 (Near Infrared)	0.76~0.90	19~29	18~30
B5 (Mid-Infrared)	1.55~1.75	14~34	10~35
B7 (Mid-Infrared)	2.08~2.35	10~21	10~24

Table 3. The results of *Chl-a* estimation by using LMR and parallel GEGA

Method	Equation	Correlation coefficient	SSE	RMSE
LMR	(2)	0.823	3.309	0.371
	(3)	0.765	4.741	0.444
Parallel GEGA	(5)	0.891	2.167	0.301

Table 4. The objective values of eight subpolulations in parallel GEGA

Generation	e1	e2	e3	e4	E1	E2	E3	E4
10	4.5627	4.702	4.2355	4.7456	4.8142	4.9364	4.3214	4.9869
20	4.4527	4.648	4.1484	4.7373	4.244	4.7217	4.3214	4.4569
30	4.3748	4.4184	4.0466	4.43	4.0663	4.7029	4.3214	4.3099
40	4.3687	4.2124	3.9826	4.3182	4.0609	4.7014	4.3214	4.2365
50	4.2273	4.141	3.9816	4.2136	4.0591	4.1947	4.2476	4.0325
60	4.2273	4.0875	3.9815	4.2105	4.0529	4.1491	4.2456	3.93
70	4.2273	4.0673	3.9815	4.2068	4.0771	4.0458	3.9043	3.9138
80	4.2113	4.0394	3.9815	4.1932	4.0582	4.0446	3.7125	3.7783
90	4.1792	3.9817	3.9815	4.2063	4.0393	3.9815	3.7106	3.6936
100	4.1713	3.9817	3.9815	4.2063	4.0368	3.9815	3.7106	3.4831
110	4.0509	3.9817	3.9815	4.2026	4.0365	3.9815	3.6694	3.4575
120	4.0509	3.9817	3.9815	3.9825	4.0365	3.9815	3.5692	3.4567
130	4.0509	3.9058	3.9815	3.9825	4.0365	3.9815	3.5526	3.4278
140	4.0509	3.6733	3.9815	3.9815	4.0364	3.9815	3.5168	3.4081
150	4.0291	3.6732	3.9815	3.9815	4.0364	3.9815	3.4901	3.4044
160	3.7011	3.6721	3.9815	3.9815	4.0364	3.9809	3.4826	3.4044
170	3.4487	3.672	3.9815	3.9815	4.0287	3.9775	3.4186	3.3952
180	3.448	3.3718	3.4044	3.9815	3.8049	3.9752	3.3877	3.3938
190	3.448	2.8995	3.4044	3.919	3.8049	3.9752	3.37	3.3889
200	3.446	2.8995	3.4044	3.8447	3.8049	3.9752	3.3849	3.379
210	3.068	2.8993	3.3847	3.8444	3.4623	3.9752	3.3864	3.3773
220	3.068	2.8993	3.37	3.5228	3.1485	3.5999	3.3855	3.3679
230	3.068	2.8993	3.0679	3.3089	2.7437	3.5998	3.3854	3.3679
240	2.4459	2.3679	3.0679	2.808	2.7356	3.5998	3.3854	3.3678
250	2.4459	2.3679	2.3668	2.7952	2.7348	2.8994	3.3825	3.1665
260	2.1673	2.3678	2.3668	2.78	2.7421	2.8994	3.3814	2.8774
270	2.1673	2.1673	2.3668	2.7797	2.7358	2.4221	2.9802	2.8774
280	2.1673	2.1673	2.3659	2.1673	2.7356	2.4126	2.3776	2.3661
290	2.1673	2.1673	2.1673	2.1673	2.1673	2.4049	2.1673	2.3661
300	2.1673	2.1673	2.1673	2.1673	2.1673	2.1673	2.1673	2.1673

Table 5. The statistical properties of 6677 estimated *Chl-a* data by using LMR and parallel GEGA

	Parallel GEGA (Equ.(5))	LMR (Equ.(2))
Minimum	0.48	0.76
Average	2.60	2.76
Maximum	4.86	3.76