27th Australasian Transport Research Forum, Adelaide, 29 September – 1 October 2004

Paper title: **Location modelling: grouping genetic algorithm using a Geographic Information Systems platform**

Author(s) name(s): Ladda Pitaksringkarn and Michael A P Taylor

Organisation(s): Transport Systems Centre, University of South Australia

Contact details:
Postal address: GPO Box 2471, Adelaide SA 5001

Telephone: (08) 8302 1771
Facsimile: (08) 8302 1880
email: Ladda.Pitaksringkarn@postgrads.unisa.edu.au

Abstract (200 words):
Business firms aim to achieve optimal performance in logistics systems for best efficiency and efficacy. Good performance in the operation of logistics systems can be initiated by determining optimal locations of facilities or, in other words, solving 'the facility location problem'. In this paper location problem analysis is considered in the context of the logistics network. The Grouping Genetic algorithm (GGA) method is employed to solve the location problem. The area of application of the models is in the planning of postharvest logistics systems for a grain crop, in this case rice. In this system, facility locations include growers, silos and markets, which all form part of a logistics systems. Partitioning production nodes into groups as markets is used with the overall aim to increase grower profits by improvement of the postharvest transportation of their crops. The study considers the location of markets and their optimum number and size by balancing transportation costs and product prices in the logistics system. GGA partitions production points and generates centroids of point groups as market locations. Fundamental procedures in GGA are presented in this paper. Current research seeks to complete the algorithm by determining control parameters for the full model using new field data. The method is being applied to a case study of the logistics system for the rice harvest from a major production area in northeast Thailand.

Introduction

Business firms aim to operate their logistics systems optimally for the best efficiency and efficacy. Good performance in operating a logistics system can be achieved in large part by determining optimal locations of facilities used in the system. How to determine facility location is one of the major problems in logistics of any product supply chain. Location modelling is used to determine the best number, sizes, and locations of facilities in a logistics network. Logistics is utilised in every sector and part of organisation including in agricultural sector. Like other products, agricultural logistics has an important role in improving entire system efficiency corresponding to the one of postharvest aims. A planner can use location analysis as a support-decision tool in looking through an entire logistics system, including capability and resources, in order to configure logistic systems optimally. The Grouping Genetic Algorithm (GGA) method is used here as a tool in solving the location problem as a partitioning tool. Although there are some differences from classical Genetic Algorithm (GA), the GGA method follows the basic GA approach to obtain a good partitioning of objects, which is a key objective in a grouping problem. In this problem, facility location includes growers, silos, and markets, which all form part of the postharvest logistics system. Finding the best partitioning of production nodes into groups as market groups is the key objective of the research, for which the underlying goal is to improve grower profits by redesigning postharvest transportation, storage and marketing systems so that growers can achieve better returns. GGA is used as a tool to partition production nodes into groups with consequences of assigning production nodes to a centroid of group as a sales market. The study considers the location of markets and their optimum number and size in a given region, by balancing transportation costs, other costs and product prices. It employs Geographic Information Systems (GIS) software as a basic platform for the model, as GIS can integrate and connect both graphical and non-graphical databases. GIS provides very useful decision support for analysis of transportion and marketing management in logistics planning and management (Valchopoulou, Silleos and Manthou, 2001). Fundamental procedures in GGA are described in detail in the paper, and operational steps and preliminary results are presented. Further research is needed to refine the control parameters used in the model algorithms.

Logistics and location problem

Demand forecasting, resource provision, production, delivery, and after-sale service are the basic processes in a life-circle production of one product termed as a supply chain. Logistics integrates the procedures to facilitate the positioning of the right products into the right places at the right time with enhancing of information technology. Therefore, logistics management is potentially used to improve the supply chain performance because logistics is a key driver moving products from one place to another in a supply chain. A diagram represents a supply chain and logistic activities shown in Figure 1. Determining facility locations is a crucial planning task in the early stage of logistics management. Well performing logistics systems may be established by properly and optimally locating facilities and organisational infrastructure. The focal point in location modelling is to determine the optimal structures in a supply chain, which comprise the number, size and geographic locations of the facilities. Optimisation techniques and heuristic algorithms have been used for many years to solve the location problem but they provide different approaches and may generate different solutions. A near optimal solution is generally obtained from the heuristics technique whereas a mathematical optimisation technique can generally generate an exact solution. However, use of an optimisation method in location modelling to obtain the 'exact' solution may prove time-consuming because of the complexity of the problem. Therefore, heuristic methods are commonly used in logistics and transportation planning including location modelling (Kasilingam, 1998).

<table>
<tr><td>PLAN</td><td>SOURCE</td><td>MAKE</td><td>DELIVER</td><td>RETURN</td></tr>
<tr><td colspan="2">Demand forecasting</td><td colspan="2">Packaging</td><td rowspan="2">Reverse Logistics</td></tr>
<tr><td rowspan="3">Plant & warehouse selection</td><td>Order</td><td colspan="2">Inventory Management</td></tr>
<tr><td colspan="2">Material handling</td><td rowspan="2">Traffic & Transport</td><td rowspan="2">Part & Service Support</td></tr>
<tr><td colspan="2">Warehouse & Storage</td></tr>
<tr><td></td><td>Procurement</td><td colspan="3">Customer Service</td></tr>
</table>

Figure 1 Relationship between stages of supply chain and logistics activities

Genetic algorithm and grouping genetic algorithm

The Genetic Algorithm (GA) is a heuristic approach, which tries to mimic the natural behaviour in searching processes associated with rules of reproducing populations. It relies on an intelligent search of a large but finite solution space using statistical methods (Haupt and Haupt, 1998). GA has thus been used extensively in solving location problems. However, a customised version of GA named Grouping Genetic Algorithm (GGA) is used to solve more specific problem i.e. a grouping problem. The grouping problem is a problem of finding a good partition in a set of objects (Brown and Sumichrast, 2003, Falkenauer, 1998, Falkenauer and Delchambre, 1992). GGA makes use of a principle algorithm from classical genetic algorithms in solving grouping problem (Brown and Sumichrast, 2003). In a standard GA process, there is maintenance within a good solution population. Second, in the next generation, new solutions (offspring) are generated from recombining parts of top-selective solutions. And third, operations including selection and recombination are performed in a manner designed to potentially transmit good characteristics from current solutions to the future ones. There are also a number of differences between GGA and GA. Special encoding is required in GGA.A group of objects so called a gene is used as a unit in GGA encoding rather than an individual object in GA (Falkenauer, 1998). GGA is an 'incomplete' algorithm in that problem-specific information has to be used in designing a given model. Grouping criteria are varied on a feature of problem consequently making GGA different characteristic in solving different problems.

Encoding

The most vital difference between GGA and GA is the encoding manner. Figure 2 shows coding using binary characters (i.e. 0 and 1) that have been popular in GA when representing objects in one denotation. A group of objects is the unit used in GGA encoding. A chromosome is divided into two levels including a gene and an allele. An allele is a smallest object using in represent natural entity, as shown in Figure 3.

01100011

Figure 2 A chromosome in GA

Allele

Gene

3,4,15	**2,10,8,7**	**11,9,1**	**5,16**	**6,13,14**

Figure 3 A chromosome in GGA

Initial population

The first set of chromosomes is called the initial population and one population at one time is called a generation. The first population is randomly generated using a random number generator that returns either one or a zero, which is similar to nature in selecting from scratch. Initial population is a start set of chromosome used for reproducing new generation until the best solution is achieved. In GGA, randomly dividing a set of objects into groups is method to generate one chromosome in the first population.

Fitness function

An objective function value for each chromosome is used for evaluate chromosomes in a population at one generation. The better the value of the objective function, the higher the probability that the chromosome will survive into the next generation. The actual form of the fitness function depends on the specific problem specification.

Selection

Two parents are selected from current generation in order to reproduce one or more new chromosomes called ‘offspring’. New offspring will compete with chromosomes in that generation to survive into the next generation. Trying to emulate natural selection, those chromosomes that have better fitness function values will have more chance to be selected in producing new offspring. There are four selection techniques (Haupt and Haupt, 1998): simple, random, weighted random and tournament selection.

Crossover operator

Mating is a procedure used in generating one or two children from two selected parents. A crossover operator is one of the operators used as a device in the mating procedure. A simple crossover operator starts off by selecting one or more crossing points and swapping cut sections to another parent. Two offspring are then produced as shown in Figure4.

Parent 1: 11|000100101
Parent 2: 10|110010111

Offspring 1: 10|110010101
Offspring 2: 11|000100111

Figure 4 mating of chromosome in GA

Because of the special methods adopted in GGA encoding, the classical operator in GA may not truly harmonise within GGA. The operator has to work with a group of objects instead of a single object as in the typical GA. Steps in mating using crossover operator in GGA are illustrated in Figure 5. However, problem-specific information is needed in designing a model during the execution of an operation especially in the steps of reinsertion and elimination of duplicate objects. To solve a given problem, the GGA application therefore has to be modified to suit that problem.

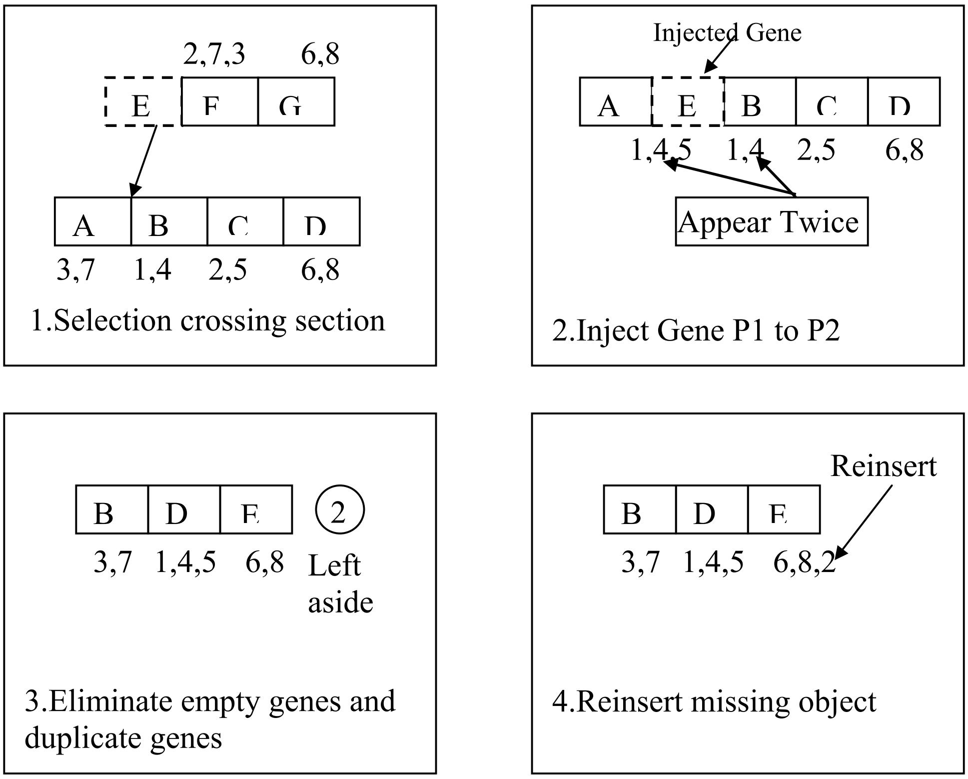

Figure 5 Mating using crossover in GGA [source: Falkenauer (1998)]

Terminate condition

Three possible stopping criteria may be used in heuristics methods (Lake, 2001)

- an optimal solution is found
- a maximum number of iterations is reached
- the number of iterations when changes of fitness function are greater than specification value.

Which condition(s) to use in the GGA facility location model has (have) yet to be determined, these will depend on model and problem characteristics.

Methodology

Grouping Genetic Algorithm (GGA) in Geographic Information Systems (GIS)

Because of the potentially large number of participants in a supply chain, typical facilities location analysis problems can be characterised as very complex and data intensive (Valchopoulou *et al*, 2001). Techniques using in modelling location analysis must therefore solve the problem efficiently and effectively to be able to identify the best solution within reasonable computational limits. GIS is a powerful tool because it integrates both spatial and non-spatial databases for analysis and display. In location analysis, hidden relationships between producers, customers, competitors and providers can be displayed in the form of GIS computerised maps based on demographics and costs (Valchopoulou *et al*, 2001). In our study therefore, ArcGIS, GIS Software is used to develop a GGA using its built-in Visual Basic programming language as a customisation tool.

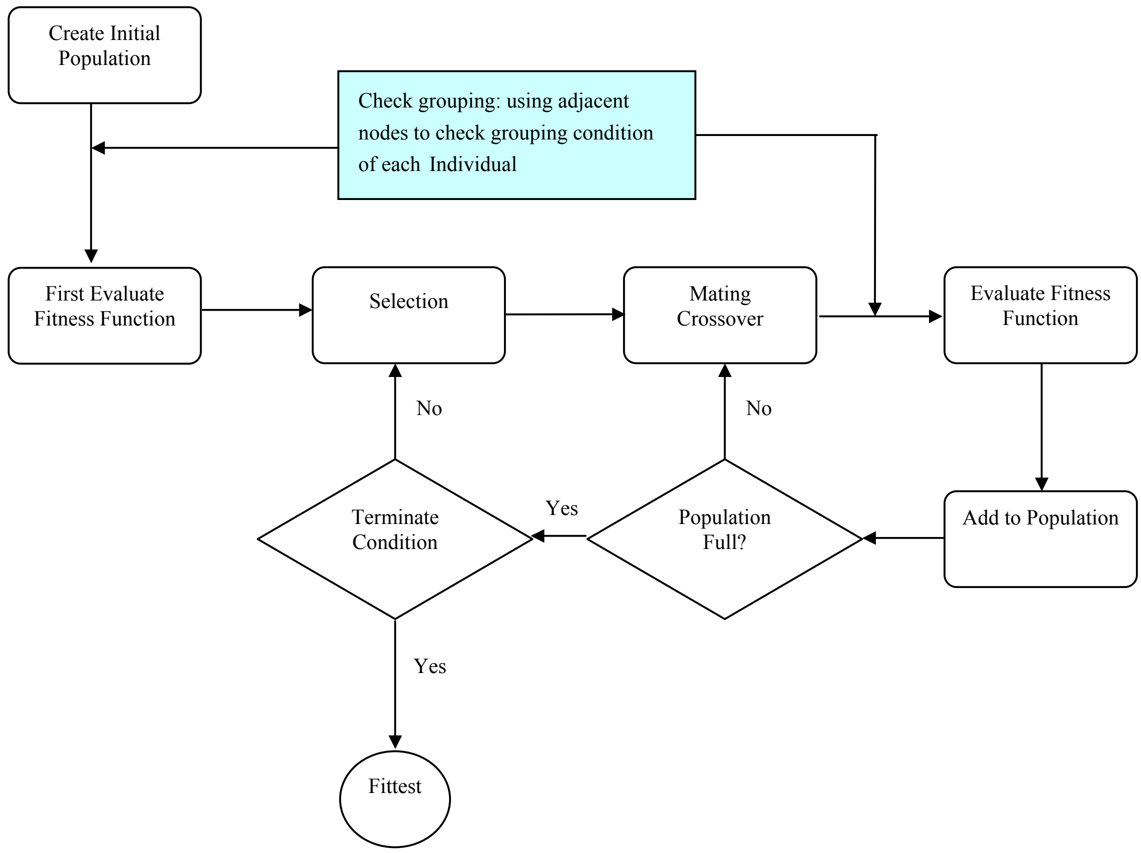

Figure 6 Flowchart of GGA Procedures

Group1	Group2	Group3
1,5,6,7,8,20,21,22,23,24,25, 38,39,40,41,42,43,44,45,55, 56,57,58,59,60,61,62,63,64, 65,66	2,39,10,11,12,13,16,17,18, 19,46,47,48,49,50,51,52,53, 54,70,72,72,73	4,14,15,26,27,28,29,30 31,32,33,34,35,36,37,67,68, 69

Figure 7 Encoding of one chromosome in GGA

GGA procedures as shown in Figure 6 include encoding, initiating chromosomes, calculating fitness function values, selecting parents, mating with crossover operator and terminating algorithm if best solution reached. Within GGA encoding, 73 production nodes will be divided into groups (Figure 7). The first set of chromosomes (the initial population) is randomly generated using rectilinear distance-based rules. In a building chromosome procedure, a number of groups (G) and a number of each group members (m) are randomly selected from unconditional circumstance in order to have mixed and varied chromosomes, as found in the nature. Assembling a group begins with single out any point by chance as a starting point then collects adjacent points for m member (s) and at last one chromosome is built. Names of both geographical layer and attribute table are added into the Population Table (Figure8). The Population table is used to maintain and manage chromosomes for reproducing the next generation.

$$P = TD - TC \qquad (1)$$

$$TD = \sum_{i=1}^{73} TradingN_i \qquad (2)$$

$$TradingN_i = PN_i SP_i$$

$$TC = \sum_{i=1}^{73} TripCost_i D_i \qquad (3)$$

$$TripCost_i = PD_i CD_i$$

P = Profit, baht
TC = Transport costs, baht
TD = Trading Value, baht
$TradingN_i$ = Trading Product Value at Node i, ton
PN_i = Production of Node i, ton
SP_i = Selling price at Sink i, baht per ton
$TripCost_i$ = Transport Cost for Node i
PD_i = Product shipped to market i
CD_i = Cost by distance i, B per ton per km

In the rice harvest production system, the profit function is the objective function of each chromosome, which is calculated from formulae (1) – (3). The profit function is derived as differences between trading values and transport costs. The fitness function and accumulated function are calculated for each chromosome and assigned to the Population table in each generation. The parent selection from all chromosomes in a population is the next move. Parent selection is performed by, random selection a number from the scale of one and the upper limit number. The upper limit is a final total fitness function in Population table and it is changeable following the accumulated fitness function of each generation. For instance in three matings, six parents are chosen by using a roulette wheel selection algorithm, providing results such as those shown in Figure 9. A crossover operator is used to produce two new chromosomes or offspring in one mating. Two parents are chosen and then two single numbers are selected from each parent that becomes an injector. One injector from one parent will be inserted to the other like position replacement. The group number of an injector will replace the previous group number of the parent member if the production nodes have the same Object ID. A number of tables are systematically built during these procedures, including Population table, Selection table, Injector tables, and Offspring tables. Consequently, an offspring table will be decoded into the GIS map form including the grouped of production nodes as well as group centroids. Offspring will be repeatedly generated and added into Population table. Fittest Chromosome from each generation is determined and the model will be stopped if the best solution is revealed or the number of iteration is reached.

Population Table: Generation # 1

No.	Chromosome	Sink	Fitness Function	Sum Fitness
1	Individual1	SinkNode1	231622	231622
2	Individaul2	SinkNode2	343986	575608
3	Individual3	SinkNode3	432678	1008286
4	Individaul4	SinkNode4	108536	1116822
5	Individual5	SinkNode5	743997	1860819
6	Individual6	SinkNode6	469004	2329823

Figure 8 Population table

Drawing results

Draw No.	Random Value	Sum Fitness Value	Selected Chromosome
1	0.08	186385	Individual 1
2	0.68	1584279	Individual 5
3	0.45	1048420	Individual 4
4	0.89	2073542	Individual 6
5	0.21	489262	Individual 2
6	0.51	1164911	Individual 5

Selection table

OBJECTID	Parent1	Parent2
1	Individual 1	Individual 5
2	Individual 4	Individual 6
3	Individual 2	Individual 4

Figure 9 Drawing results and Selection table

Case study

Thong Kula RongHai, often called the Kula area, is the major production area and also the most famous area in producing Jasmine rice (Hom Mali rice) in Thailand (Agriculture Economic Research Centre, 1996, Chinsuwan, 1999). Most of the databases used in the analysis were provided from the study by the Agriculture Economic Research Centre (1996) for enhancing Hom Mali production and marketing in the Kula area. The area covers five provinces including 73 sub-districts, 760 villages. A sub-district or called 'Tambon' is used as the zonal representation of production and contains both spatial and attribute data as shown in Figure 10. There were 59 villages (479 households) participated in the 1996 survey. Basic information on rice growers and their areas were collected, as well as details of machinery used in the area. GIS databases for the research project were provided by the Thai Department of Environmental Quality Promotion and the Ministry of Natural Resources and Environment. The GIS data are available as individual the provincial level. The database for each province contains spatial and general attribute data of the road network and the administration district area. The two databases provide the basic inputs for the model.

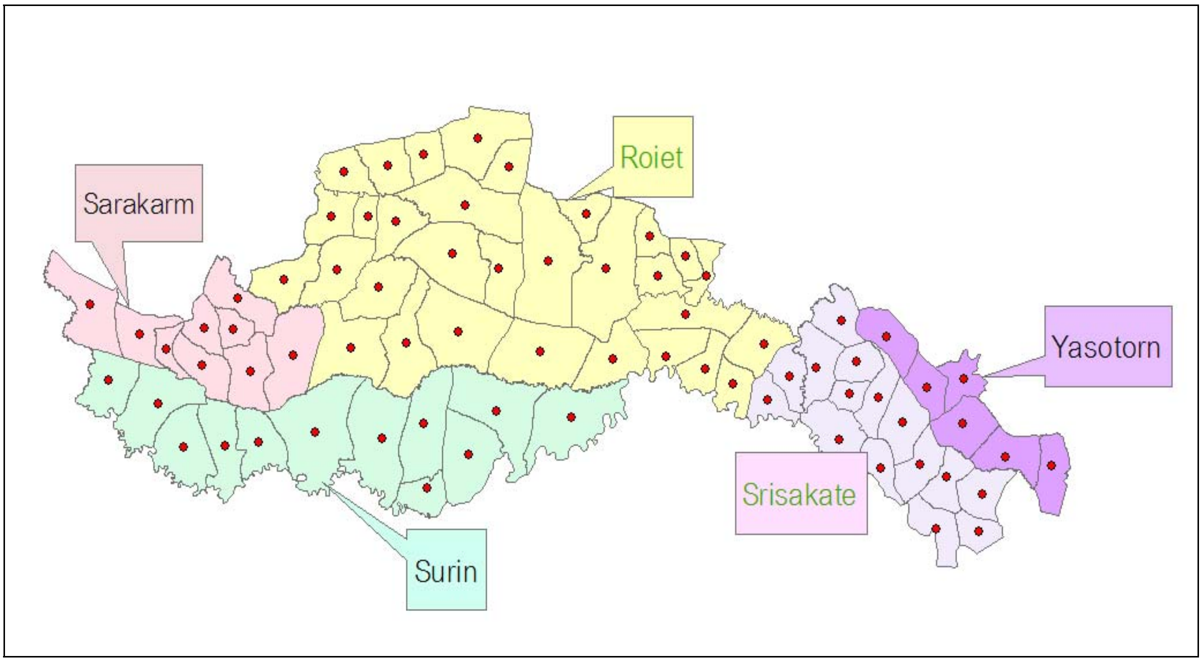

Figure 10 Geographic database of case study, Kula area

Results

Separate groups of production nodes denote each chromosome in the GGA used in the location modelling analysis. Within one chromosome, two layers and one table in ArcGIS are created, which include one polygon, one point layer and one attribute table (as presented in Figure 11). A polygon layer is an outcome of dissolving production nodes that have same group number. Furthermore, a point layer represents the centroid of each group in the polygon layer. The centroids denote as a market or pick-up point of a group. Distance is calculated from each production node to the centroid of its group indicating as 'Sink Name'. Transport cost can then be calculated from distance and existing transport rate. Based on an uncompetitive condition in selling price, existing sell price for each province is used to determine selling revenue and profit can then be found in Million baht. The Population Table contains the attribute table name, point layer name, sum profit fitness value and total fitness value.

OID	Sink Name	Distance (km)	Sell Revenue (Million baht)	TransCost (Million baht)	Profit (Million baht)
1	2	11	78.4	16.9	61.5
2	2	7	41.9	6.3	35.6
3	2	11	25.2	5.7	19.5
4	2	11	55.1	11.8	43.4
5	2	18	25.9	7.2	18.7
6	2	24	28.8	10.9	17.9
7	1	20	71.8	22.3	49.5
8	1	14	130.7	36.6	94.1
9	3	12	51.6	11.9	39.7

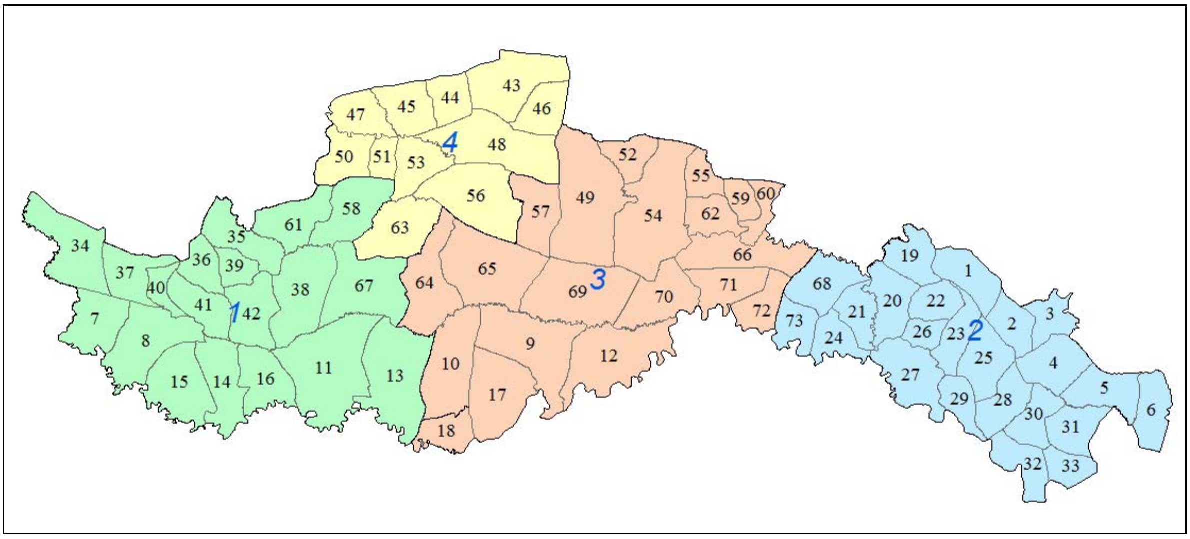

Figure 11 A chromosome classified in groups and a part of Individual Table

Six random numbers between one and the upper limit number (22898.7.4 in this case) are generated and used as six parents, as shown in Figure 12. In the first mating, Individual5 and Individual6 are then assigned as Parent1 and Parent2 respectively. In GGA crossover operator, one group name, four, is selected from Individual5 as an injector group for insert to the Parent2, Individual6. The Injector table lists production nodes having four as a group name. The production nodes in Individual6 having OID same as ‘old OID’ in Injector1 table will then be replaced with four as its group number. A new Offspring table is therefore created and after that the Polygon layer and the Centroid point layer are also generated. Two new Offspring tables 1 and 2 are added into Population Table. Next generation can thus be formed as selecting the sixth top fitness value from all chromosomes including two new chromosomes.

Population Table: Generation #1

OID	Group No.	Chromosome	Sink	Fitness Function	% Change	Sum Fitness
1	8	Individual1	Centroid1	3862.8	1.84	3862.8
2	4	Individual2	Centroid2	3751.6	-1.09	7614.4
3	10	Individual3	Centroid3	3860.3	1.77	11474.7
4	3	Individual4	Centroid4	3783.7	-0.25	15258.4
5	6	Individual5	Centroid5	3864.3	1.88	19122.6
6	4	Individual6	Centroid6	3776.0	-0.45	22898.7
	Base Case			3793.1		

Selection table

Draw No.	Random Value	Sum Fitness Value	Selected Chromosome
1	0.75	17173.5	Individual 5
2	0.92	21066.2	Individual 6
3	0.31	7098.4	Individual 2
4	0.60	13738.8	Individual 4
5	0.68	15570.6	Individual 6
6	0.48	10991.0	Individual 3

Injector1 table

OID	Sink Name	Distance	Sell Revenue	Transport Cost	Profit
11	4	3.12	161	4.4	156.6
13	4	11.20	139.6	31.3	108.3
14	4	11.07	100.2	22.2	78
15	4	16.57	70.9	18.3	52.6
16	4	5.79	116.4	15	101.4
18	4	18.99	48.3	14.3	34

Figure 12 Population table, selection table, and injector table

Discussion and conclusions

A preliminary version of facility location model has been developed using Grouping Genetic Algorithm technique, and implemented in ArcGIS Geographic Information Systems software package. Both transport cost and selling price from existing situation are employed to calculate result of the base case and the profits are 3793.1 Million baht. In GGA modelling, programmatically generated distance from grouping condition and existing transport rates are used to calculate transport costs except the sell revenue remains the same as from the existing condition. By running a number of chromosomes, number and size of groups individually vary the profit fitness function. Primary running results showed the higher number of groups has the better fitness function because of lower transport costs. However, to be more realistic, construction costs will be considered and a different transport service including a new transport rate will also be introduced as an alternative using to improve the entire logistics performance in this part of supply chain. And because of the variety of problems in the real world and the problem-specific characteristics required in GGA, parameter setting from one problem may not be used for others. Next step of this study is finding control parameters to fine-tune the GGA model. Control parameters including population size, selection procedure, crossover probability and maximum generations can be determined by experimental testing.

Soon after, the complete GGA versions will be ready and used to generate the best solution for different condition.

References

Agriculture Economic Research Centre (1996) Office of Agricultural Economics, Ministry of Agriculture and Cooperative, Bangkok, Thailand, pp 240.

Brown, E C and Sumichrast, R T (2003) *Computers and Operations Research*

Chinsuwan, W (1999) *KKU Research Journal 4*, 1-3

Falkenauer, E (1998) *Genetic Algorithms and Grouping Problems New York:* John Wiley and Sons

Falkenauer, E and Delchambre, A (1992) In *IEEE International Conference on Robotics and Automation: Proceeding,* Vol 2, pp 1186 -1192.

Haupt, R L and Haupt, S E (1998) *Practical Genetic Algorithms,* New York: Wiley-Interscience Publications

Kasilingam, R G (1998) *Logistics and Transportation: Design and Planning,* Boston: Kluwer Academic Publishers

Lake, M. R. (2001) In *School of Civil* Engineering, Queensland University of Technology, Brisbane, pp 299.

Valchopoulou, M., Silleos, G. and Manthou, V. (2001) *International Journal of Production Economics 71*, 205-212