Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 2908 - 2920, 2005

GROUPING GENETIC ALOGIRHTM IN GIS: A FACILITY LOCATION MODELLING

Ladda Pitaksringkarn
Ph.D. Candidate
Transport Systems Centre
University of South Australia
Adelaide, South Australia
5031, Australia
Fax: +618-8302-1880
E-mail: ladpit@gmail.com or
Ladda.Pitaksringkarn@postgrads.unisa.edu.au

Michael A P Taylor
Professor and Director
Transport Systems Centre
University of South Australia
Adelaide, South Australia
5031, Australia
Fax: +618-8302-1880
E-mail: MAP.Taylor@unisa.edu.au

Abstract:
Jasmine rice 'Hom Mali' in Thai is one of the major commodities in Thailand. Farmers in Northeast region, who are main producers, have to face with several problems causing lost from their crop. One of causes is lacking of accessibility to market offering satisfactory price of Hom Mali leading to be poor logistics system in the area. In this study, therefore, facility location model will be developed in term of profitability optimisation using Genetic Algorithm technique. Tradeoffs between farmer revenues and transport costs are used in order to finding optimal logistics network configuration. The algorithm will be used a given existing conditions to estimate the optimal configuration of facility in study area. Number, location, and capacity of facilities are suggested to provide levels of optimal logistics system in the study area.

Key Words: Facility Location Analysis, Agricultural Logistics, Genetic Algorithm, and GIS Modelling

1. INTRODUCTION

Like other industrial products, supply chain processes of agricultural commodities take account of two logistics activities at planning stage; forecasting demand and designing all facilities for the entire chain. Jasmine rice 'Hom Mali' in Thai is one of the major commodities in Thailand. There are problems in Hom Mali supply chain especially for farmers in Northeast region, who are main producers causing farmers facing losses from their crop. Long distance to destination market is one of the major causes in this problem (Agriculture Economic Research Centre 1996). In fact, there are two other problems hidden behind this. At reached markets, lower prices are offered and also a lack of proper transport systems is provided for the entire area. Because of the over supply to the local consumption, Hom Mali paddy have to be therefore exported outside the area to countrywide or worldwide. Similar to grain systems in South Australia, over the past two decades, South Australian growers had faced with similar problems as over supplying to local consuming. AusBulk was then founded from grower contribution in order to manage and manipulate crops throughout seasons all year round. At present, AusBulk become one of the largest agri – business in Australia specialising in logistics providing, freight, storaging and handling grain. Consequently, AusBulk could be a good model to improve Hom Mali supply chain by developing better logistics performance in the systems. As mentioned above, designing and locating all related facilities is an early major task in order to provide a well perform of logistics in a supply chain management. Related facilities will therefore be located and allocated in this study in order improve logistic performance in Hom Mali supply chain. Production nodes, pick-up points and markets are three basic interacted features involved in paddy supply chain.

2. LOCATION MODELLING AND ANALYSIS

2.1 Logistics in supply chain

Supply chain is a procedure of making over one product from raw materials until being final products delivered to end-users. A supply chain network is divided into three stages including supplier stage, plant stage and distribution stage (Selcuk Erenguc, Simpson & Vakharia 1999). Supplier stage relate to acquiring raw materials and other related materials from different sources supplying to plants where transform, compound and combine materials to be final products in the second stage. Distributing finished products to the final destination as end-users is the final stage in the supply chain. Because of growing interest in supply chain management, recently logistics has acquired great significance in industry(Syam 2002). Logistics involves in stages of supply chain since logistics integrates and manipulate procedures in placing of the right products into the right places at the right time with enhancing of information technology(Pitaksringkarn & Taylor 2004). Generally costs are automatically taken account in every logistic component when processes and services occurred. The effective integration of logistics cost components such as transport costs with facility location models could affect entire supply chain because the two are high interrelated in practice (Syam 2002). Decisions in facility locations and allocating supply/demand are generally concerned in each stage before operating therefore other related logistics activities could be ultimate performance. The quality of logistics processes can therefore be affected from locations of facilities and allocation demanding/supplying in the supply chain configuration.

2.2 Location problem and analysis

Forecasting demand and planning in the nature of all structures including facilities for entire process are two primary logistics activities of supply chain processes. Optimum operation and maintenance costs are likely to be achieved from a well and proper plan for establishing the entire system. Designing of constructions and arrangements of structures is one of the most important tasks for entire process in supply chain. The decision-making in facility locations thus has a significant impact on both quantity and quality of distribution process in the supply chain. The focal point in location modelling is to determine the optimal structures in a supply chain, which comprises of the number, size, and geographic locations of the facilities (Pitaksringkarn & Taylor 2004). Types of facility models are varied depending research focus and application. Fields in facility location modelling studies are summarised in (Klose & Drexl 2005) as follows:

- A configuration of potential facilities to be modelled in the plane, specified network, and discrete locations
- Research nature in the objective which is minimised or maximised
- Capacity constrains

- Single- or multiple-product or stages in analysis
- Static or dynamic facility models

In order to solve location problem, analysis techniques have been used such as optimisation technique and heuristics algorithms. Both techniques provide different approach in solving and may generate different solutions. An exact solution is generally obtained from a mathematical optimisation technique whereas the heuristic techniques can normally generate a near optimal one. Because of problem complexity and size concerns, an optimisation technique may take long time to solve the location problem thus heuristic techniques become more practical in the real world problem.

2.3 GIS Application in Location modelling

Because of the potentially large number of participants in a supply chain, typical facilities location analysis problems can be very complex and data intensive. Techniques using in modelling location analysis must therefore solve the problems efficiently and effectively. GIS is one of analysis tools that work with both spatial and non-spatial using in analysing as well as displaying in different formats. GIS may be considered as a support tool for planners using to make decision in site and facility selections at strategic or planning stage in a supply chain. The unique capabilities of GIS make it outstandingly useful as an analysis and visualising tool because it allows capture both spatial elements and geographical locations of facilities (Raicu et al. 2002). GIS is used as a decision tool to support logistic and marketing managers to evaluate locating options for facilities based on costs and customer demand (Valchopoulou, Silleos & Manthou 2001). GIS is employed to display results of ranked candidate sites in different scenarios for users to select the best sites. Also, the decision in retail location planning using GIS was examined in (Clarke & Rowley 1995). GIS, as a management tool, can help retailers to understand the catchments of their stores and the geodemographic structure of their catchments. In the other words, the former is to know where customers come from and what customer types of catchment structures capture is the latter. Pattern of customers and facilities can be varied to environmental circumstances. Two changes impacting distribution systems are structural changes and operational changes (Hesse & Rodrigue 2004). In this study, both suppliers and facilities are changed consequently logistics dimension and systems are changed. GIS is then used as an analysis tool for location modelling as well as displaying results.

3. GENETIC ALGORITHM AND GROUPING GENETIC ALGORITHM

Heuristics methods have commonly been used to solve problems in segments of logistics management. Different approaches have been used employed for solving the problems. These approaches have different searching procedures initiating from various assumptions and faiths. As one of heuristics techniques, Genetic Algorithm (GA) is an algorithm that tries to mimic the natural behaviour in searching processes associated with rules of reproducing populations as inheriting good genes (parts) from good performance parents (Pitaksringkarn & Taylor 2004). Many studies have shown that GA provides better solution solving in the facility location problem comparing with other traditional techniques (Houck, Joinest & Kays 1996). Advantages and robustness of GA include speed and reliability, easy to solve different problem formulations, extendable, easy to hybridise and poor data to be identified (Hurley, Moutinho & Stephens 1995; Gonzalez & Rodriguez Fernandez 2000). However, it was pointed out in (Brown & Sumichrast 2003) that the classical GA has difficulties in dividing objects into groups in a grouping problem. Because of high redundant of standard encoding scheme, a crossover operator in the classical GA may produce new children with none reached standard where other operators may cause too much disruption to successful population members. Thus a customised form of GA has been introduced. Grouping genetic algorithm (GGA) is developed from the basis of GA and using for more specific problem i.e. a grouping problem. The key objective of GGA is to encode objects into groups by specified conditions and criteria. GGA makes a use of principle algorithm as well as modified algorithm from classical genetic algorithm. In a basic principal, standard operators are employed to produce new generations. Producing new chromosomes is to recombine parts of previous selected chromosome therefore solutions are maintained within a set of population. Major keys of differences and similarities between GGA and GA are at Table 1. Briefly, GGA is not just a tailored GA but it is an algorithm with distinct difference from the standard GA (Falkenauer 1998).

Table 1. Differences and Similarities of GGA and GA

Similarities	**Differences**
o Maintain a population of solutions o Offspring are created from recombining parts of current solutions o Gene operators are used to create the next generations	o Special chromosome encoding o Operator manners both crossover and mutation o GGA needs problem-specification information to

In general, a GA represents problem solutions as strings of values; these string are called chromosomes. Each chromosome represents a possible solution to the problem. Encoding is a process of interpretation a real world problem into genetic chromosomes and using term of alphabet to represent objects in that objects. A binary character of 0 and 1 is very widely used to represent objects in GA meaning to two choices for example ‘Yes’ and ‘No’ for 1 and 0 respectively (Falkenauer 1998). In fact, depending on conditions and characteristics of that problem, more than two choices could be interpreted for example in the problem of dividing objects into ‘M’ groups which could be more than two groups. Instead of object itself, a group of objects is the unit used in encoding in a grouping problem. There are two principles using in GA encoding (Pham and Karaboga 2000). Firstly, minimal alphabets will be set as choices for an object. Another principle is how the setting of choices performs and the meaning of choice setting. The first set of chromosomes using to produce new chromosomes is called ‘Initial Population’. The next generation combines two parts of existing and new created chromosomes. Fitness function value is employed in order to evaluate how fit of each chromosome. The fitter chromosomes will have higher the surviving probability into the next generation. Reproducing process initiates from selecting two parents from current generation and mating them by using one or two available GA operators. GA tries to emulate natural selection therefore selection two parents relies on chromosome’s value of fitness function. As a result, the better fitness function values will have more chance to be selected as a parent. Once parents are selected, which operators to be used in mating process. Mating is a procedure used in generating one or more child, offspring, from two selected parents. Two standard operators are used in classical GA including crossover operator and mutation operator. Mating using crossover operator generally produces two offspring comprising a part of each parent. Figure1 shows mating procedure using crossover operator. But in GGA crossover operator, each parent part includes groups of objects instead of a bunch of independent objects. Figure 2 shows steps in GGA mating process. Firstly a group of objects is selected for injecting into the other parent. If any objects appear twice, the previous record will be eliminated. New grouping will be performed because of new injection of objects. The empty group will be eliminated whereas the single residual objects will be reinserted to another group. Because of GGA is an algorithm that relies on problem-specific information therefore different problems will vary different condition in reinserting. Offspring will be created again and again in order to seek the best solution. The near optimal will be found in the heuristics techniques including GA and GGA therefore three criteria are used to terminate heuristics algorithm (Lake 2001). If an optimal solution is found, an algorithm will then be terminate or stop running algorithm until a maximum number of iterations is reached. The last criterion is the number of iterations when changes of fitness function not greater than specification value.

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

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

Figure 1. Crossover Operator Mating in GA

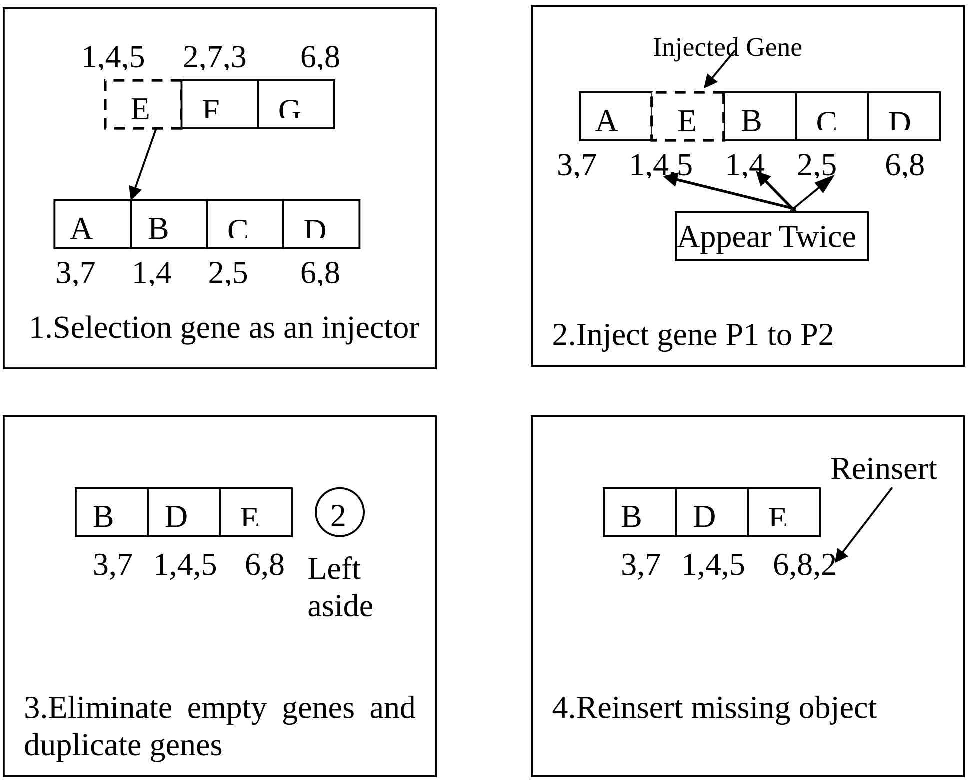

Figure 2 Mating using Crossover in GGA [source: Falkenauer (1998)]

4. METHODOLOGY

In this section GGA procedures of facility location modelling in this study are depicted. Phases and processes of GGA procedures are presented in Figure 4. The algorithm begins with initiating a number of chromosomes as the first set of population. Fitness function values of the initial population are then calculated for using in later selection process. By ranking fitness costs, two parents are then chosen to produce new offspring. Reproducing new offspring employed two operators i.e. crossover and mutation. Within the mating process, only changed genes (object groups) from mating process in both crossover operators and mutation operators are recorded. Thus only changed groups are re-calculated again to determine the fitness function value. Parents are re-selected and new offspring are re-created and added until chromosomes in population table are full. Next step is to find the highest profit value to be the fittest chromosome at that generation. After generations, the algorithm will be terminated if number of generation is reached.

Encoding: In this study, 73 production nodes or source nodes will be geographically divided into groups. A table and geographic encoding of production nodes are shown in Figure 3. In each group, a centroid of a group represents a market or pick-up point.

Group	Object ID
1	13 - 17
2	1 – 12
3	18-39
4	40
5	41 – 59
6	60-73

Figure 3. Geographical and Table Grouping of Production Nodes in Study Area

Initial population: A network distance matrix is formed for using in calculating of group forming. In chromosome building, the number of group is firstly random between one to specified number from unconditional circumstance in order to mix and vary chromosomes. Numbers of production nodes as group members in each group are also random. In a group assembly, starting point is initially selected and adjacent points will then be collected as a member for each group. Once all production nodes are denoted, the nodes are merged into one area for one group and to calculate centroids of groups later.

Fitness function evaluation: This study tries to maximise farmer profit from their paddy. Profit is derived from a difference of farmer revenues and paddy transport costs as shown in (1) – (3). Two factors related in revenue estimation are paddy types and market types. Sale percentage and price of both fresh and dry paddy impact to revenue of paddy sales. Generally, dry paddy price is higher than the fresh one but the sale ratio between both paddy types depending on farmers' financial conditions. Different types of markets offer different level of paddy price. By market operation, market can be categorised into five types including mills, local shops, commodity central markets, farmer cooperation, and designated farmer banks. However, by market area, in this study markets are divided into two types i.e. inside markets and outside markets. An inside market is a centroid of each group while outside markets locate at city centre of five provinces. Two network distances therefore are generated from GIS tool as shown in Figure5. Two distances are used to calculate costs that paddies are shipped from production nodes to both inside market and outside market. Transportation costs comprise of shipping costs and drying costs. Dry paddy has a higher price than the fresh one however there are higher costs in drying paddy. Costs occurring after harvesting paddy in drying process are costs from moving paddy to production field to storages. Shipping costs vary to distance whereas drying costs are charged by paddy volume. Two assumptions are used in calculation of fitness function as follows:

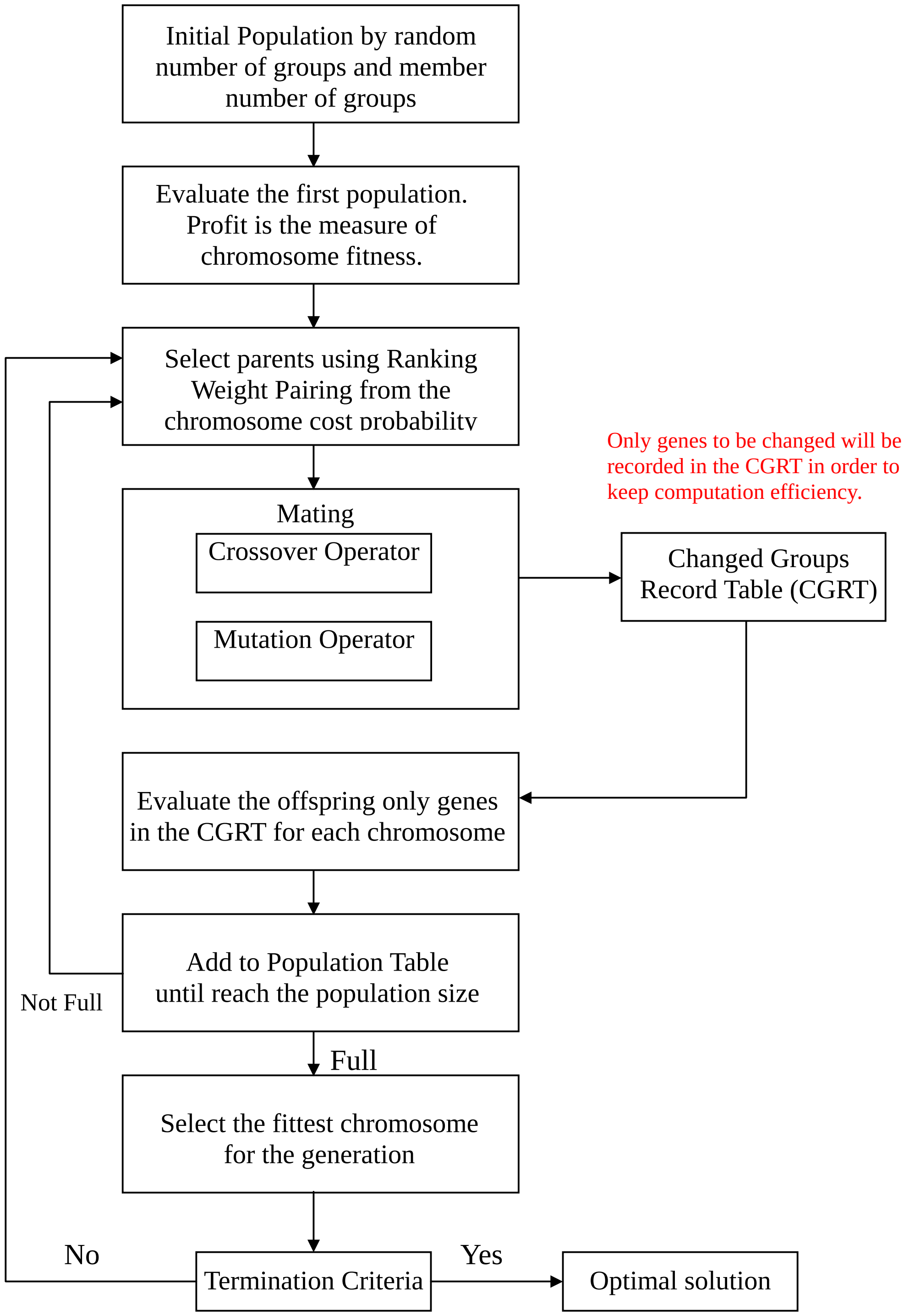

Figure 4. Diagram of GGA Procedures

- Only one outside market is a representative for each province and locates in central city area of each city. A centroid of a group represents as an inside market.
- Each production node sells inside market paddy to an inside market representative or its centroid and sells outside market paddy to its province outside market representative.

$$Profit = Revenue - TransportCost \tag{1}$$

$$\begin{aligned} Revenue = & \sum_i \left(ProInNode_i\right) \times \left(FPriceIn \times \%FNode_i + DPriceIn \times \%DNode_i\right) \\ & + \sum_i \left(ProOutNode_i\right) \times \left(FPriceOut \times \%FNode_i + DPriceOut \times \%DNode_i\right) \end{aligned} \tag{2}$$

$$\begin{aligned} TransportCost = & \sum_i \left(ProInNode_i \times InDistNode_i \times TransRate\right) + \\ & + \sum_i \left(ProOutNode_i \times OutDistNode_i \times TransRate\right) \\ & + \sum_i \left(ProNode_i \times \%DNode_i \times DryRate\right) \end{aligned} \tag{3}$$

Denotation

Survey database: Production Node level

Pro*Node$_i$* = Production at Node$_i$

Pro*InNode$_i$* = Product to be sold to an inside market at Node$_i$

Pro*OutNode$_i$* = Paddy to be sold to an outside market at Node$_i$

Survey database: Province level

Each production node has the same value in the same province

% DNode$_i$ = Percent of Paddy sold as dry crop at Node$_i$

%FNode$_i$ = Percent of Paddy sold as fresh crop at Node$_i$

TransRate = Transport cost B/ton-km

DryRate = Drying cost B/ton

FPriceIn = Price of fresh paddy at inside market

DPriceIn = Price of dry paddy at inside market

FPriceOut = Price of fresh paddy at outside market

DPriceOut = Price of dry paddy at outside market

GIS calculation

InDistNode$_i$ = Distance from Node$_i$ to its centroid

OutDistNode$_i$ = Distance from Node$_i$ to its province outside market

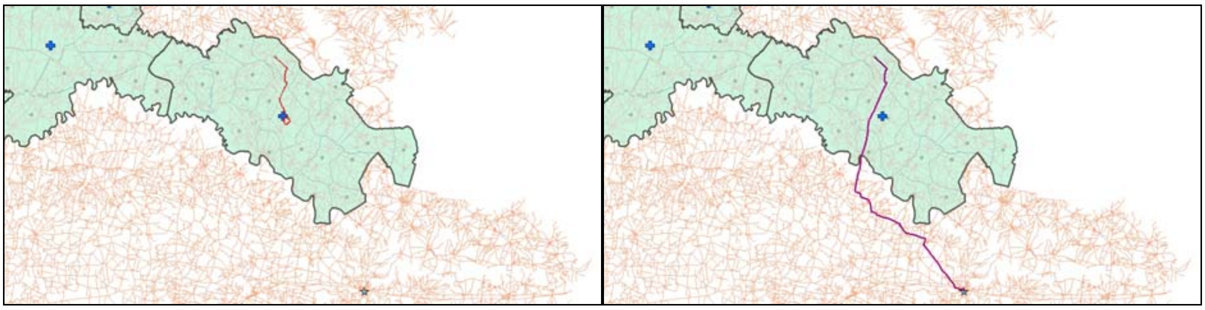

Figure 5. Distances from Production Nodes to Inside Market and Outside Market

Selection: A process to select two chromosomes, as parents using in producing new chromosome is pairing process. A chromosome with a higher value of fitness function has more chance to be selected as a parent but this is still depended on selection techniques. Haulp and Haulp (1998) categorises selection techniques into four types as follows:

- Simple paring: Two chromosomes recorded in population lists with generating sequence are selected as parents from top to bottom list i.e. 1&2, 3&4, and so on.
- Random pairing: Chromosomes are ranked in term of costs within population list. Two numbers from a uniform random are then generated to be an identification number of two parents.
- Weight random pairing (Roulette wheel): Ranking by costs of chromosomes in population list is used in this pairing technique. Probability of each chromosome is calculated to cost weight. If a randomly generated number is greater than accumulated probability of which chromosome, that chromosome will be selected as a parent.
- Tournament selection is the method trying to mimic natural mating as competition by forming mating pool for subset of chromosomes then finding the best two from the set to be parents.

Genetic operators: Operators are used as an engine tool in producing offspring for the next generation. In this study, two genetic operators including crossover operator and mutation operator are used in GGA.

Crossover operator

A crossover operator in GGA is different from the standard one because of difference in encoding scheme. Instead of select crossing point to separate chromosomes, group(s) of objects, so called injector(s), are selected in order to swap between two parents. In the Figure6, injector in Parent1 is selected to replace the previous grouping in Parent2. Offspring is generated from parts for both parents. Group number replacement is used in GGA crossover. 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. Consequently, group(s) will be created or eliminated to be new offspring as shown in Figure6.

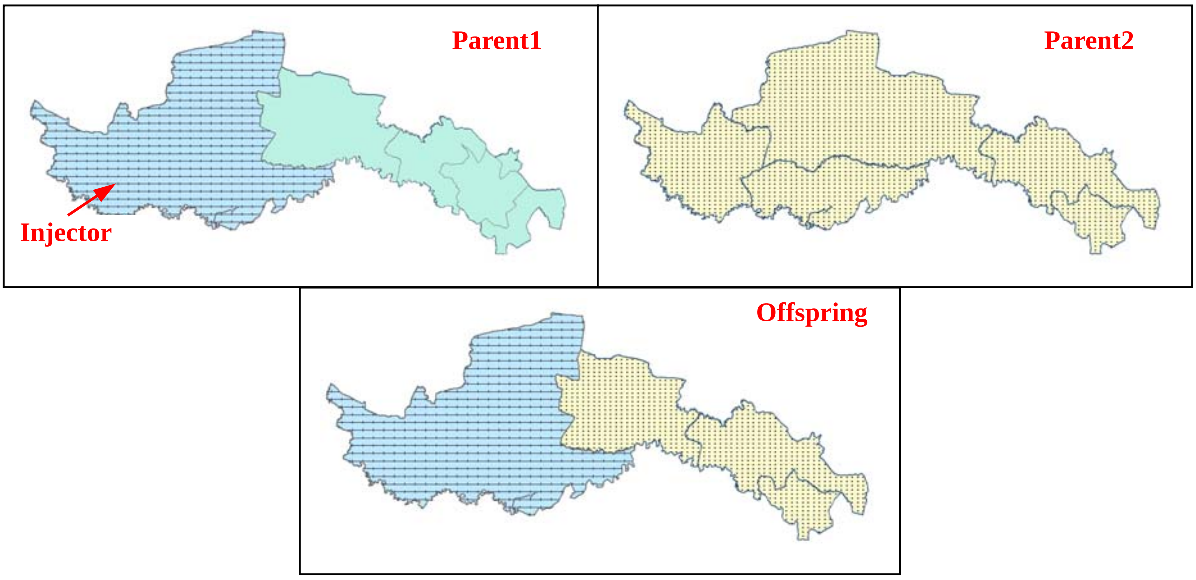

Figure 6. Crossover Operator

Mutation

Working with a single object, a standard mutation is the technique to change object value from one to another. However, a GGA mutation operator for grouping problems must work with groups rather than objects. In mutation process, nodes as parts of group are selected. Because of grouping condition, selected nodes will be mutated and belonged to the adjacent group as shown in Figure7. There are three types of mutation in GGA including eliminating an existing group, swapping groups and creating new group. A mutation operator provides new space for searching in an algorithm which is so called 'exploring' search space.

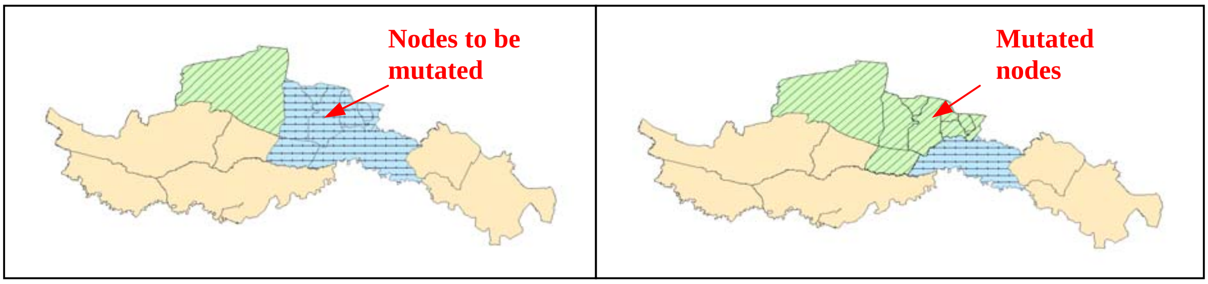

Figure 7. Swapping Mutation

5. CASE STUDIES AND DATA ANALYSIS

Tong Kula Ronghai; know as the Kula area, is the major production area and also the most famous area in producing Homa Mali rice in Thailand (AERC 1996 and Chinsuwan 1999). Basic database including area, production and marketing database are provided from the study of KKU and DoAE (1996) for descriptive data while Department of Environmental Quality Promotion, Ministry of Natural Resources and Environment has provided geographical database. The Kula area is in the Northeast region of Thailand. It covers five provinces comprising ten districts, 760 villages in 73 sub – districts. A sub-district or called 'Tambon' is used as a zonal representation of production and contains both spatial and attribute data as shown in Figure8. In this study, Hom Mali paddy trades are divided into two categories: within area and outside area. This information will be used to estimate trips of paddy shippings for existing condition.

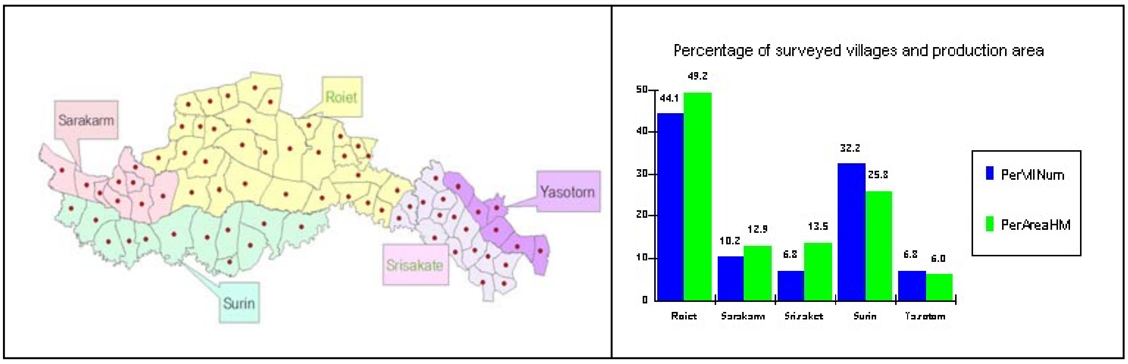

Figure 8. Production Nodes

6. RESUTLS AND DISCUSSION

Table 2. Control Parameters in GGA

Parameters	Description
Initial Population condition	Specify number for initial random selecting group number
Population Size	30 chromosomes
Good Population Size	12 chromosomes
Selection method	Ranking weight pairing
Crossover operator	One random group crossover with crossover rate of 0.98
Mutation operator	Random production nodes with mutate rate of 0.1
Termination criteria	Stopping running if there is no change in ten new offspring

After running experiments, control parameters using in GGA are listed in the Table 2. A designed number as a group number has to be identified for initial random selecting. This selection helps the algorithm in searching the best solution in a specified group number. 30 chromosomes are used in order to maintain a population in one generation where 12 chromosomes represents as good portion in population using for producing the next generation. Ranking by cost weight of chromosome's fitness function is employed as selection technique. 98 and 1 from every 100 offspring are generated from crossover and mutation respectively. One group is used in crossover operator whereas one object is mutated from one group to another group in mutation. After ten offspring created if there is no change in fitness function value, the algorithm is stopped. The last best solution is recommended as the optimal solution.

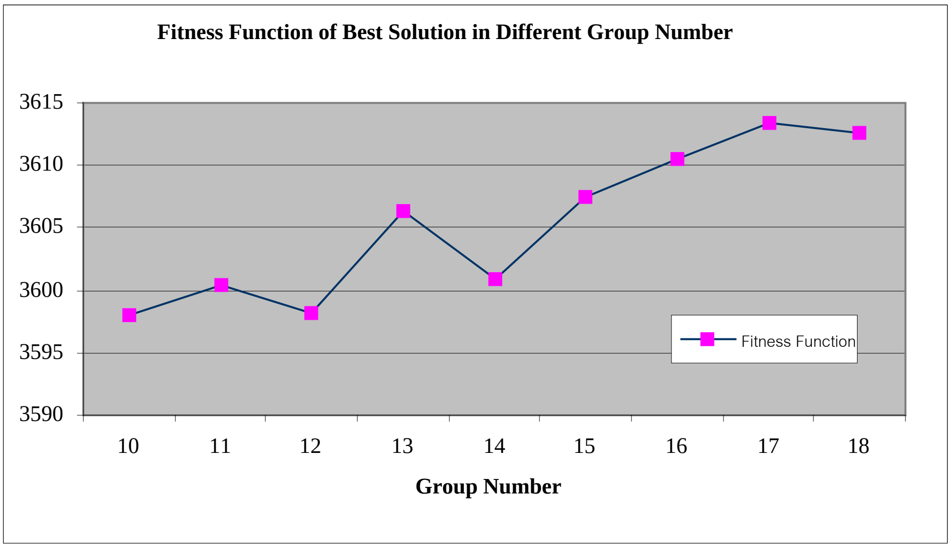

Figure 9. Fitness Function Varies to Group Numbers

Table 3 and Figure 9 show figures and the graph of fitness function value of the best solution in different group number and percent change comparing with the case base one. The trend of the relationship is the higher group number offering the higher the fitness function value. In the GGA model, only grouping condition within area is used to improve logistics performance. There are 14 market destinations in the existing condition or called 'case base' which give the profit of 3542.839 billion baht. Comparing with the 14-group of production nodes from the GGA model, the output from the model gives the better performance at the percent of 1.83. Market/pick-up point position impacts to distance in shipping paddy. In the case base and the model grouping at 14 groups in Figure10. Therefore the grouping of production nodes has an impact on transport costs that vary to distance. This implies that positions of market/pick-up points affect transport costs and profit. If we put market/pick-up points in the right position, profit can be improved (see the case of comparing 14-group of base case with 14-group of algorithm). Because only transport cost is considered to reflect performance of Hom Mali logistic system, shipping distances are reduced as a result of higher number of grouping. Therefore Figure 9 shows the best results for each group number which give different grouping condition but being the best solution in different levels. An optimisation model is a prescriptive model which generally the components of model have to be verified. In this study GGA components are transport costs and revenue which can be verified in order to prove results from this study. However, not only transport cost but also construction costs to set up markets/pick-up points will be considered because generally setting up markets/pick-up points have costs for both infrastructure and operation component. Next step of this study is to consider transport cost and construction costs to determine the number and location of market points for the study area.

Table 3. Fitness Function in Different Group Number

Group Number	Fitness Function	Percent change
Case Base	3542.839	1.56
10	3598.06453	1.62
11	3600.34478	1.56
12	3598.13433	1.79
13	3606.404	1.64
14	3600.96335	1.83
15	3607.53672	1.91
16	3610.516689	1.99
17	3613.458371	1.97
18	3612.530386	1.56

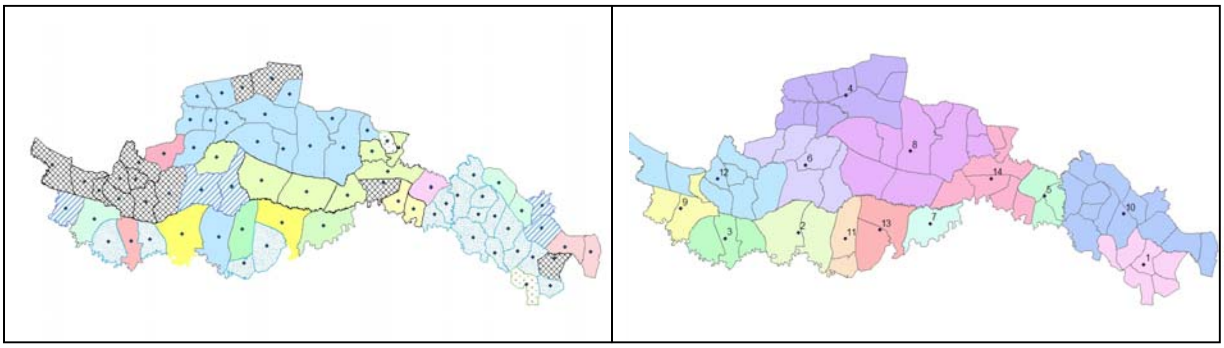

Figure 10. Existing and Algorithm Grouping

REFERENCES

Agriculture Economic Research Centre, A 1996, *Database study for enhancing Hom Mali Rice Production in Tong Kula Ronghai*, Office of Agricultural Economics, Ministry of Agriculture and Cooperative, Bangkok, Thailand.

Brown, EC & Sumichrast, RT 2003, 'Impact of the replacement heuristics in a grouping genetic algorithm', *Computers and Operations Research*.

Clarke, I & Rowley, J 1995, 'A case for spatial decision-support systems in retail location planning', *International Journal of retail and distribution management*, vol. 23, no. 3, pp. 4-10.

Falkenauer, E 1998, *Genetic algorithms and grouping problems*, John Wiley and Sons, New York,Singapore.

Gonzalez, EL & Rodriguez Fernandez, MA 2000, 'Genetic optimisation of a fuzzy distribution model', *International Journal of Physical Distribution and Logistics management*, vol. 30, no. 7/8, pp. 681-696.

Hesse, M & Rodrigue, J-P 2004, 'The transport geography of logistics and freight distribution', *Journal of Transport Geography*, vol. 12, no. 3, pp. 171-184.

Houck, CR, Joinest, JA & Kays, MG 1996, 'Comparison of genetic algorithms, random restart and two-opt switching for solving large location-allocation problems', *Computers and Operations Research*, vol. 23, no. 6, pp. 587-596.

Hurley, S, Moutinho, L & Stephens, NM 1995, 'Solving marketing optimisation problems usign genetic algorithm', *European journal of marketing*, vol. 29, no. 4, pp. 39-56.

Klose, A & Drexl, A 2005, 'Facility location models for distribution system design', *European journal of operational research*, vol. 162, no. 1, April, pp. 4-29.

Pitaksringkarn, L & Taylor, MAP 2004, 'Location modelling: Grouping Genetic Algorithm using a Geographic Information Systems platform', paper presented at the 27th Australasian Transport Research Forum, Adelaide, Australia, 29 September - 1 October 2004.

Raicu, R, Taylor, MAP, Zito, R & S., M 2002, *An evaluation of logistics network modelling tools available to South Australian companies: Literature review and initial evaluation*, Transport Systems Centre, University of South Australai, Adelaide.

Selcuk Erenguc, S, Simpson, NC & Vakharia, AJ 1999, 'Integrated production/distribution planning in supply chains: An invited review', *European Journal of Operational Research*, vol. 115, p. 219–236.

Syam, SS 2002, 'A model and methodologies for the location problem with logistical components', *Computers and Opertions Research*, vol. 29, pp. 1173-1193.

Valchopoulou, M, Silleos, G & Manthou, V 2001, 'Geographic information systems in warehouse site selection decisions', *International journal of production economics*, vol. 71, pp. 205-212.