ORIGINAL ARTICLE

An Application of Genetic Algorithms in Seismic Refraction Interpretation

Chaiwat LERSVIRIYANANTAKUL[1], Pattara AIYARAK[1] and Warawutti LOHAWIJARN[2]

[1]Innovation in Physics and Intellectual Properties Research Unit,
[2]Geophysics Research Unit, Department of Physics, Faculty of Science,
Prince of Songkla University, Hat Yai, Songkhla 90112, Thailand

ABSTRACT

Seismic refraction is a conventional geophysical method for determining geological structures by making use of velocity differences of propagating waves in various rock types. Data interpretation to determine the structures are based on travel time and distance from the source to each receiver (geophone) presented in the form of a t-x graph. In addition to the many conventional methods, our paper evaluates here a Genetic Algorithm (GA) for the interpretation of seismic refraction data. The studied subsurface structures were 2-layer and 3-layer models with horizontal and dipping planar interfaces. Actual field tests were also used for the evaluation. Comparing the GA interpretation technique with the conventional Seismic Interpretation Program (SIP), the results showed that the GA can process seismic refraction data well and gives virtually exact solutions for the synthetic structures. On actual field data, the results are quite similar to those from SIP.

Keywords: Genetic algorithms, seismic refraction, velocity inversion, travel time inversion

INTRODUCTION

Seismic refraction technique may be used for determining the properties of geological rock structures, such as the thickness of each layer, and also the existence and location of faults and anomalous velocity zones in the earth's crust. The seismic data sets are composed of the travel times (travel time, t) of the first arriving seismic waves from a source to an array of receivers (i.e. geophones) at a distance, x, on the surface. They are usually displayed in the form of a graph called a t-x graph.

From the t-x graph, certain parameters such as wave propagation velocity, thickness and dip of geological layers can be determined. Several interpretation methods are available e.g. the intercept time method, the wavefront reconstructure method, the ray tracing method, and the generalized reciprocal time method [1].

In this paper, we present an additional method for analyzing seismic data, called Genetic Algorithms (GA). The GA method adopts a kind of genetic process in order to solve the equation, and starts with a random solution. The next generated solutions are initially random, but gradually tend to become realistic solutions. After a large number of generations, the best generated solution is found. In order to arrive at such a solution, the Seismic Inversion technique was used. The underlying structures, i.e. the solutions, were generated and converted into t-x graphs. After that the generated and actual graphs were compared. A better matched graph, i.e. higher fitness, has a higher score and is therefore, more likely to be the correct solution.

MATERIALS AND METHODS

The Genetic Algorithm method

The GA method was designed to work on non-linear, multi-modal and sometimes poorly understood problems, and still yield results [2], and mimic natural selection and evolution [3]. Examples of the use of GA are the traveling-salesman problem [4], the school timetable problem [5], robot motion planning [6], pattern recognition in the immune system [7], crossword puzzles [8], reconstructing occlusal surfaces of teeth [9] and fault matching [10].

A typical algorithm might make use of the following operational: [11]
1. A number, or population, of guesses of a possible solution to the problem;

2. A way of calculating how well the individual solution fits within the population;
3. A method for mixing fragments of the most reasonable solutions to form a new, on average even better solution; and
4. A mutation operator to avoid permanent loss of diversity within the solution.

The GA method is started by randomizing solutions of the problem. Each solution is encoded and called a chromosome. Each chromosome is determined and given a mark, or fitness number. A higher mark typically implies a better answer or solution. The population of solutions will then move to the next generation using genetic processes through selection, crossover and mutation. The resulting generation will contain a better or at least the same fitness. After several such generations, the solution found is expected to be exact or close to the real solution.

In order to apply GA for solving the problem, each parameter is encoded in binary. Each bit is called a gene as shown in **Figure 1** and the overall parameters are called individuals.
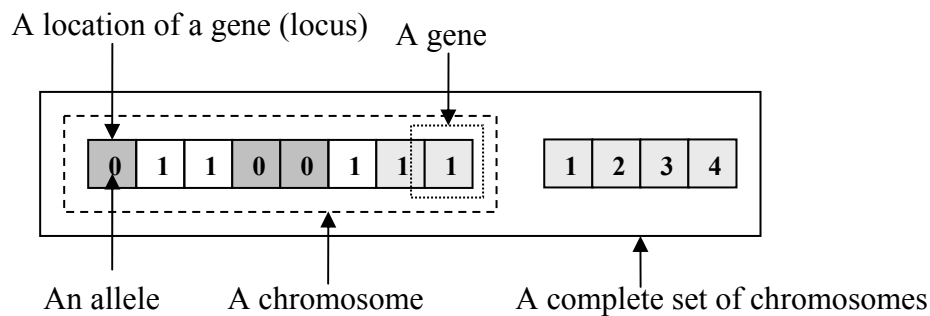

Figure 1 The chromosome components.

To evaluate the fitness of a solution, GA requires an objective function. A particular group of chromosomes is selected from the population to generate offspring by defined genetic operations. The fitness of the offspring is evaluated in a similar fashion to their parents. The chromosomes in the current population are then replaced by their offspring. In the GA process, there are mainly three operations in order to generate the new offspring, i.e. selection, crossover and mutation.

Selection

As in natural selection, individuals with a higher or sometimes, depending on the problem, lower fitness have a higher opportunity to be selected and mate in order to create a new population set.

Crossover

This method generates a new population from an old one. It is starts by mating individuals and exchanging genes as shown in **Figure 2**.
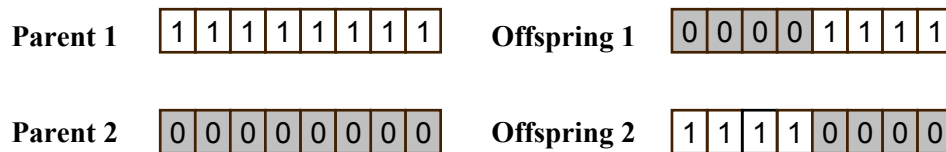
| Parent 1 | 1 1 1 1 1 1 1 1 | Offspring 1 | 0 0 0 0 1 1 1 1 |
| Parent 2 | 0 0 0 0 0 0 0 0 | Offspring 2 | 1 1 1 1 0 0 0 0 |

Figure 2 One point crossover.

Mutation

In order to increase the diversity of the gene pool, i.e. improving the chance to find a proper answer, a mutation process is needed. This method is simply done by flipping an arbitrary bit in the chromosome as shown in **Figure 3**.

| Individual | 0 0 0 0 0 0 0 0 | Mutated individual | 0 0 1 0 0 0 0 0 |

Figure 3 Bit-flipped mutation.

A GA cycle is repeated for a fixed number of generations, or until no further improvement is observed. The best chromosome generated during the search is the final result of the GA. The flowchart of the GA process is shown in **Figure 4**.

In **Figure 4**, the GA process begins by randomizing geological rock structures or initial population. Travel time data were calculated from each structure and called, in the program, the calculated t-x data. The input of the program which is the travel time data from the field work was compared with the calculated t-x data to determine the fitness of each structure. The

calculated t-x data are compared with the field work data using a statistical method to determine its fitness. The GA process, as shown in **Figure 4**, is performed until a satisfactory solution is found.
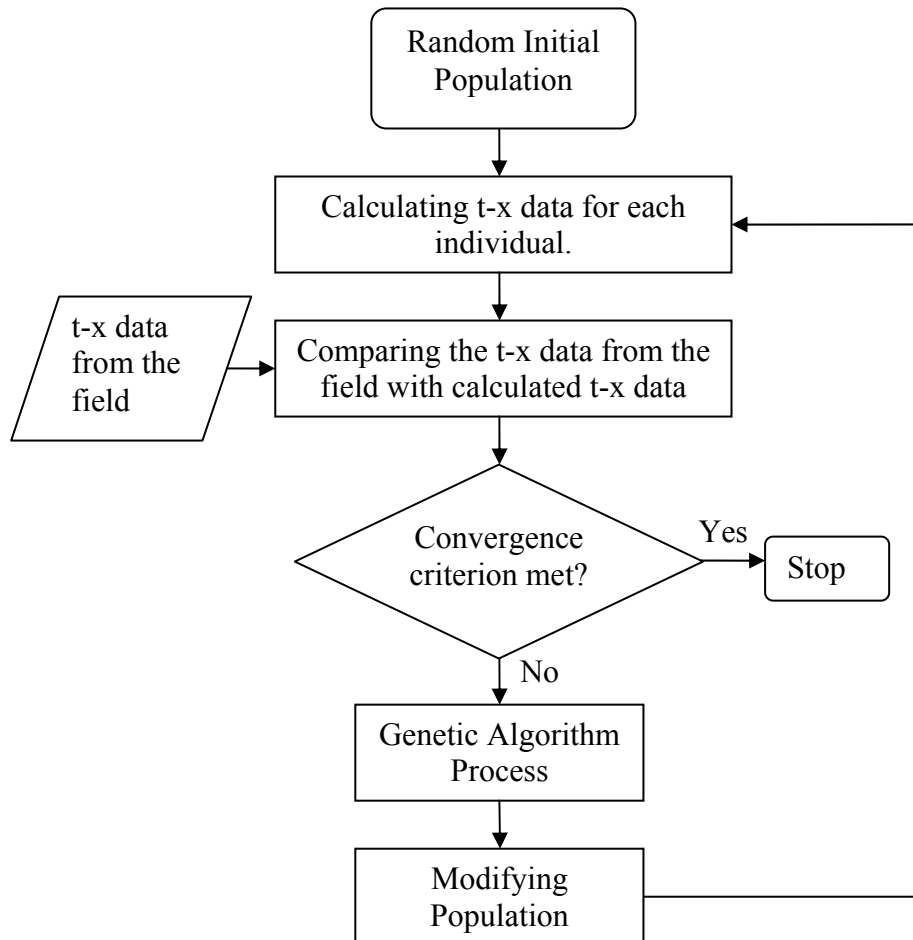

Figure 4 Genetic Algorithm process for refraction interpretation.

Application of GA in seismic interpretation

In this study the GA is applied for the interpretation of seismic refraction data of simple geological structures. We studied the interpretation of 2 and 3-layer planar dipping interfaces. Initially, the solutions, i.e. the underlying structures, were randomly generated. The t-x graphs corresponding to the structures were calculated and through comparison between these graphs and the real data the results were evaluated. The solutions consisted of

the number of layers, velocities, depths and dipping angles of each layer (see **Figure 5**).
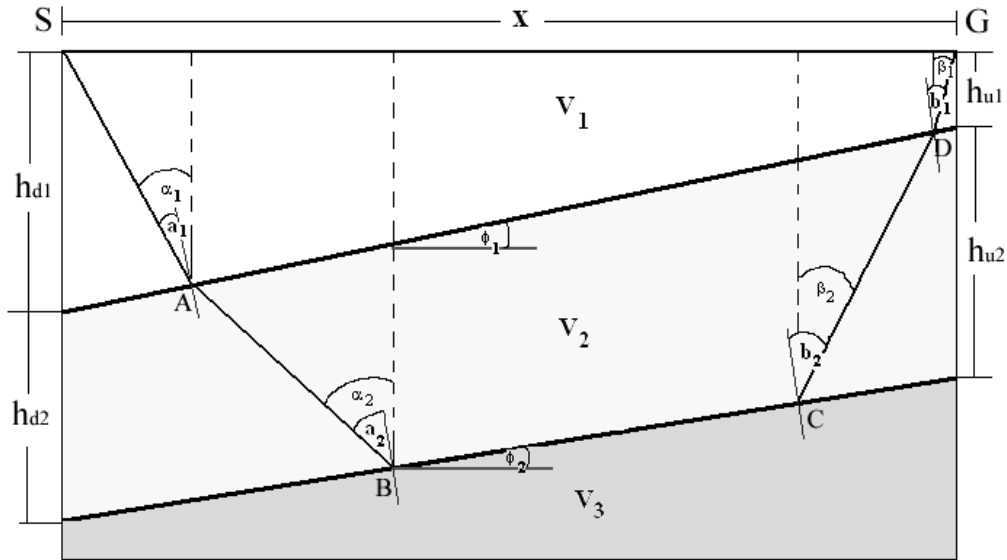

Figure 5 An example of three generated underlying rock structures.

Since we confined the number of layers with dipping angles to three, there are seven parameters: V_1, V_2, V_3, h_1, h_2, ϕ_1 and ϕ_2. Each parameter will be translated into chromosomes and these chromosomes are manipulated using the GA. If the actual structure has two layers, the parameter h_2 will simply be zero. Similarly, when dealing with non-dipping layers, the parameters ϕ_1 and ϕ_2 will be zero. For simplicity, the values of these parameters are rounded up to be integers and converted to chromosomes by converting the decimal values into binary notation.

Each parameter was thus encoded into a binary number i.e. a chromosome. The ranges of parameters were set such that wave velocities, the depths and the dipping angles have a realistic range of 330 - 7200 m/s, 0 - 20 m and $-15°$ to $15°$, respectively. The depth is the measurement of h_{di} and a negative dipping angle means that h_{di} is less than h_{ui}, see **Figure 5**. The length of the chromosome encoding the velocities is 13 bits whereas depths and dipping angles use only 5 bits each.

The calculation of the t-x graph from underlying structures

In order to evaluate the fitness of each individual, the t-x graph for each one has to be calculated. The critical point in calculating the t-x graph is

to find the first arrival of the signal at each geophone. The signal travels from the source to the geophone by several paths, i.e. a direct path and a number of multiple refracted paths. All paths have to be explored in order to find the arrival time; the shortest is selected for the data of the t-x graph.

Direct and refracted signals

The travel times of the paths from source to geophones are of two types: direct and refracted paths. The travel time of the direct path is the time that the signal uses from the source to the geophone directly whereas the travel time of the refracted path, as the name suggests, is the time that the signal needs from the source to the geophone via refractions at the interfaces of the underlying layers. Eq. (1) and (2) give the time from the direct and refracted paths, respectively [12].

$$t(k) = \frac{x}{V_1} \tag{1}$$

$$t_d(k) = \frac{x \sin \beta_1}{V_1} + \sum_{i=1}^{k-1} \frac{h_{di}}{V_i}(\cos \alpha_i + \cos \beta_i) \tag{2}$$

$$\alpha_i = a_i + \phi_i \tag{3}$$

$$\beta_i = b_i - \phi_i \tag{4}$$

where

$t(k)$ is the time at the kth geophone for direct wave.

$t_d(k)$ is the time at the kth geophone for refracted wave.

h_{di} is the vertical thickness of the ith layer below the source.

V_i is the velocity in the ith layer.

α_i and β_i are the angles between the vertical line and the down- and up-going rays in the ith layer, respectively.

ϕ_i is the dip angle of ith layer.

a_i and b_i are the angles between the normal of the interface and the down- and up-going rays in the ith layer, respectively.

x is the distance from the source to a geophone which is a fixed value for every experiment.

Rejection of the structures

The initially generated structures and the products from the GA need to be verified on their validity, for instance in the case that a structure ends

before the last geophone, as shown in **Figure 6**. The conditions for checking the validity above are shown in Eq. (5) and (6).

For the first layer
$$h_{d1} \geq x \tan \phi_1 \tag{5}$$
For the second layer
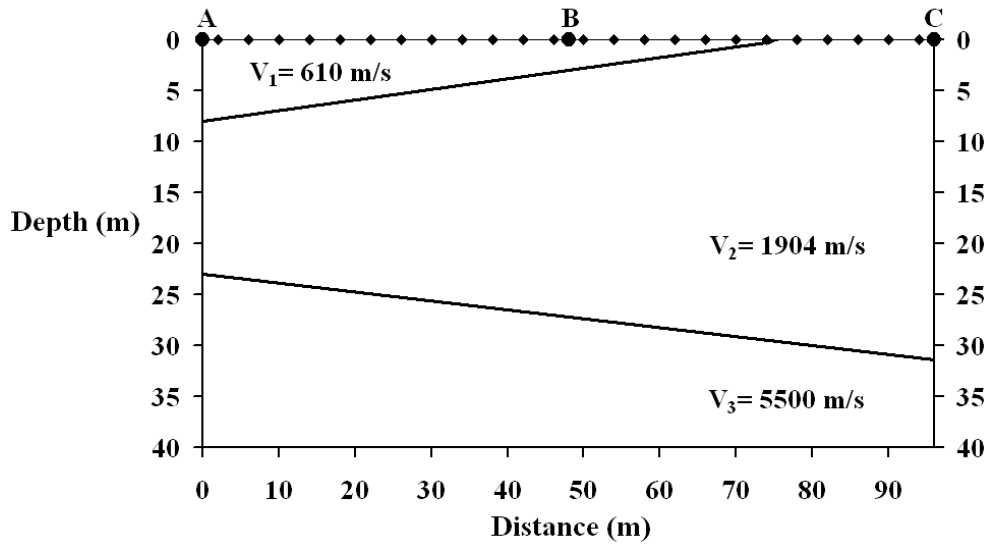$$h_{d2} \geq x(\tan \phi_2 - \tan \phi_1) \tag{6}$$

Figure 6 An example of a rejected structure.

Fitness function

As stated above, the GA requires a so-called fitness function in order to evaluate the individuals. In this study the root-mean-square (RMS) error method is used to evaluate each individual solution. We compared the t-x graphs of the generated structures with the actual or observed results. The closer both graphs, the better are the fitness of that individual. The RMS error can be calculated using Eq. (7).

$$RMSError = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (t_i^{cal} - t_i^{obs})^2} \tag{7}$$

where
n is the number of the observations (geophones).
i is each observation.

t^{obs} and t^{cal} are the observed and calculated travel times, respectively which are determined from Eq. (1) and (2).

The fitness function provides the mechanism for evaluating the status of each chromosome. This is an important link between the GA and the problem. When the error value of each structure is known, the structure which has the lowest error is kept, as it is assumed to be nearest to the real structure. If the structure has a really low error, it means the structure will probably be the correct result.

The algorithm was tested on synthesized and on actual field data. First, we synthesized the 2- and 3-layer structures with and without dipping. The t-x graphs of these structures were calculated manually and then put into the GA as actual data. After testing the validity of our method, field data were tested. The probabilities of crossover, mutation and the population size were also varied in order to find the best result at a fixed running number of generations. We also compared the results from the GA with the well-known and widely used Seismic Interpretation Program (SIP).

RESULTS

Synthesis data model

Five synthesized models with 2-layer and 3-layer structures with and without dipping were generated. They have the following parameters.

Model 1 ($V_1 = 800$, $V_2 = 2500$, $h_1 = 6$, and $\phi_1 = 0$)
Model 2 ($V_1 = 912$, $V_2 = 2640$, $h_1 = 9$, and $\phi_1 = -3$)
Model 3 ($V_1 = 800$, $V_2 = 1604$, $V_3 = 4500$, $h_1 = 6$, $h_2 = 18$, $\phi_1 = 0$ and $\phi_2 = 0$)
Model 4 ($V_1 = 610$, $V_2 = 1904$, $V_3 = 5500$, $h_1 = 6$, $h_2 = 17$, $\phi_1 = -3$ and $\phi_2 = -5$)
Model 5 ($V_1 = 810$, $V_2 = 1840$, $V_3 = 4500$, $h_1 = 8$, $h_2 = 13$, $\phi_1 = 1$ and $\phi_2 = -4$)

Various probabilities of crossover and mutation and population sizes were tested on these synthetic data sets. The best results with a fixed number of generations was found when using a crossover probability ($P_c = 0.8$), mutation probability ($P_m = 0.01$), and a population size 200. **Figure 7** shows the results at different probabilities when the number of generations was fixed and **Figure 8** shows the population sizes with RMS errors.
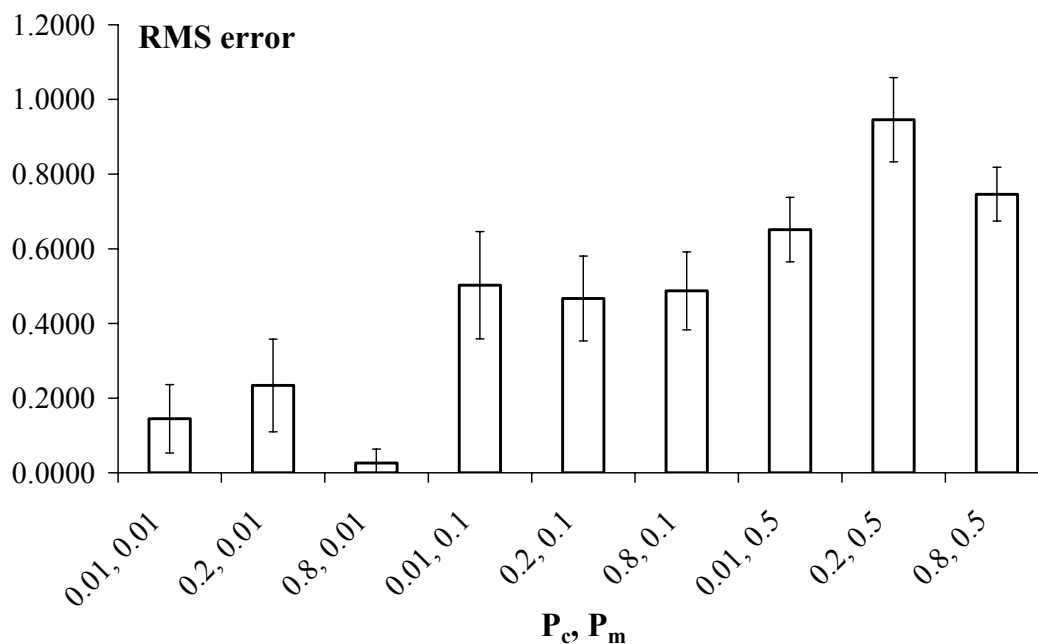
Figure 7 RMS errors from 10,000 generations with different P_c and P_m. (an example of 50 individuals on a 2-layer structure with a dipping interface)
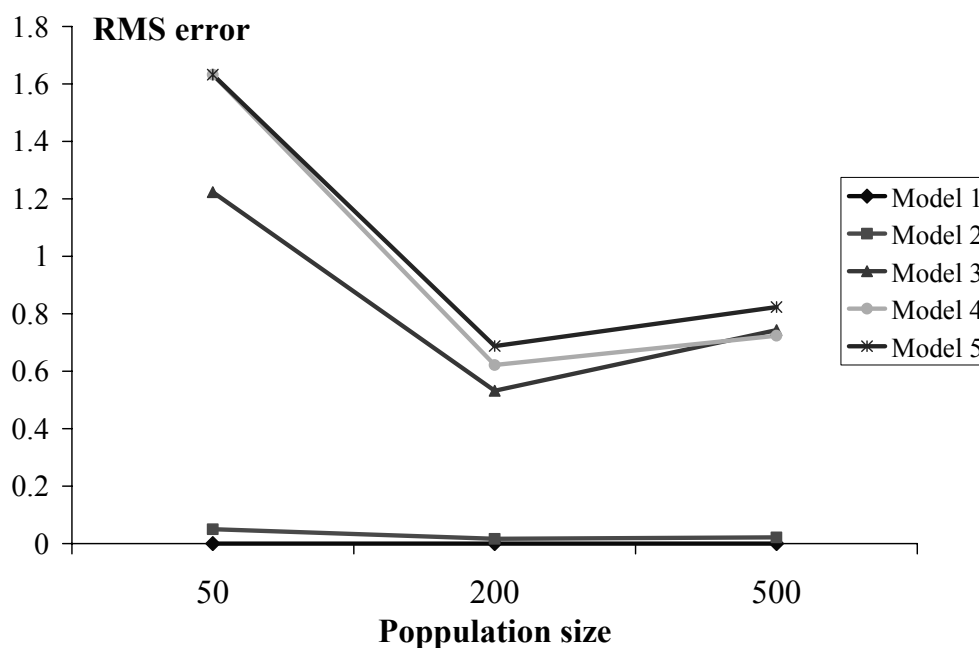

Figure 8 The population size versus the RMS error for different structures.

The results from the GA showed that all models found are exactly the same as the synthesized models. **Figure 9** shows two examples of the results from the GA model. All models were also tested with SIP and although the results were not exactly the same they were very similar (see **Figures 9** and **10**). The only major difference is that SIP tends to interpret the layer with a non-smooth interface.
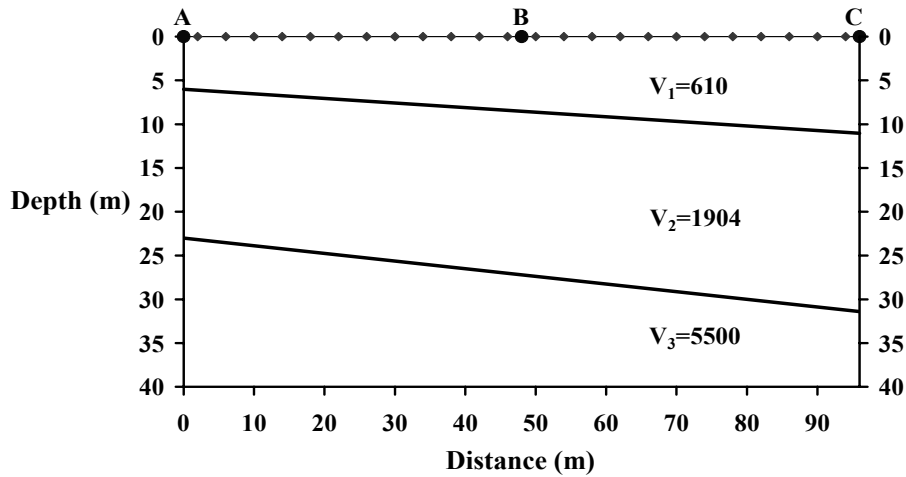
(a) Model 4 ($V_1 = 610$, $V_2 = 1904$, $V_3 = 5500$, $h_1 = 6$, $h_2 = 17$, $\phi_1 = -3$ and $\phi_2 = -5$)
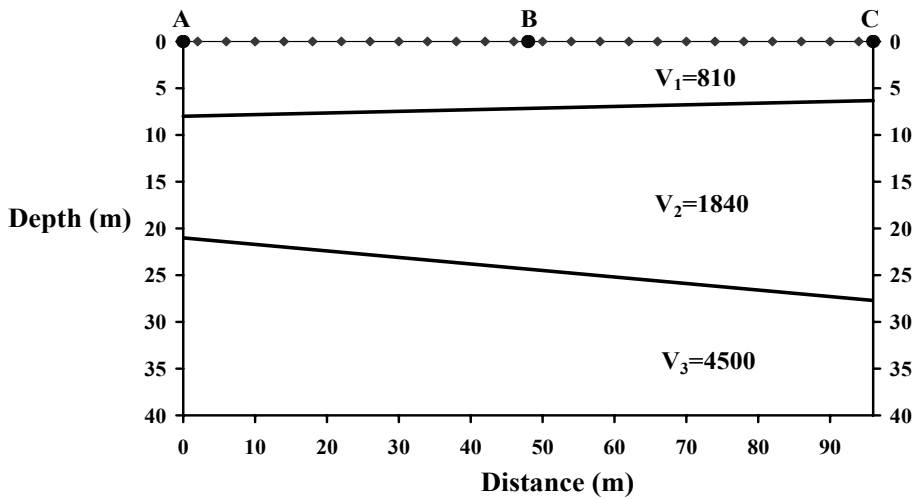
(b) Model 5 ($V_1 = 810$, $V_2 = 1840$, $V_3 = 4500$, $h_1 = 8$, $h_2 = 13$, $\phi_1 = 1$ and $\phi_2 = -4$)

Figure 9 Results from GA which is exactly the same as the synthesized models.
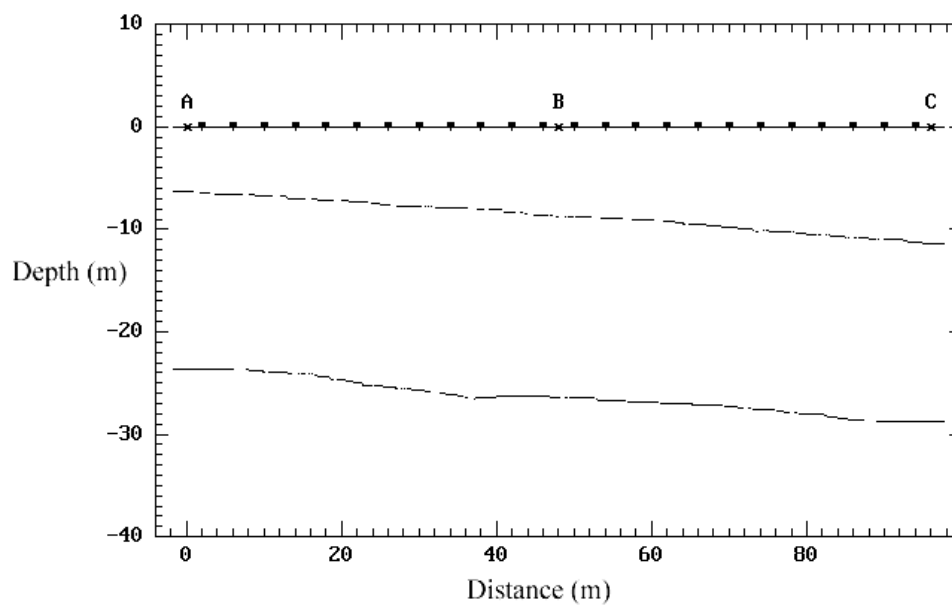
(a) Model 4
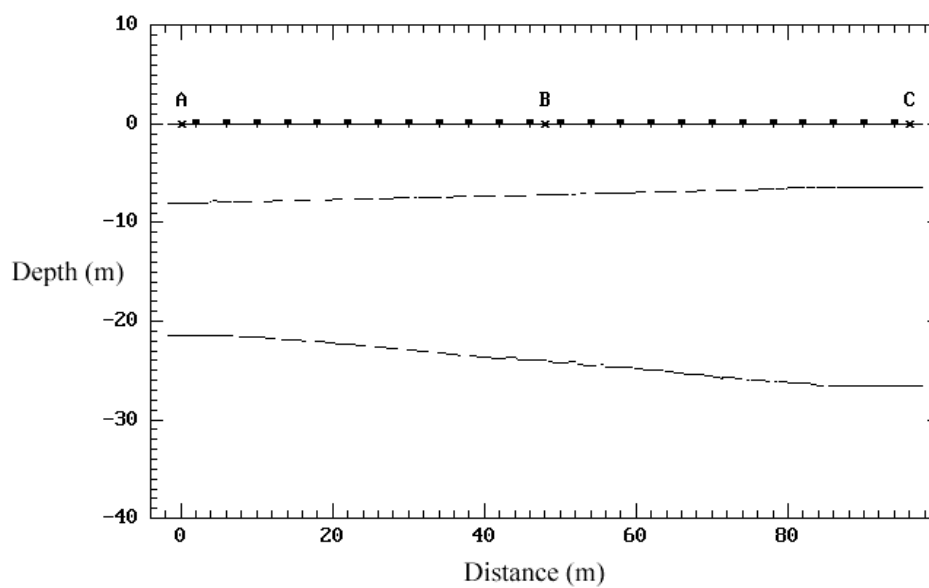
(b) Model 5

**Figure 10** The results of model 4 and model 5 using SIP.

Field data

Field data were collected at the Hat Yai campus of Prince of Songkla University. Two lines were measured on the football field another located near the reservoir. Line 1 and 2 used 12 geophones and 3 source points, while Line 3 used 24 geophones and 5 source points. When using SIP to interpret the data, the result from Line 1 showed that there are two layers and the velocities are 472 and 1646 m/s, respectively. The thickness under sources A, B and C are 3.7, 4.1 and 4.4 m, respectively. Line 2 is 5 meters to the north and parallel to Line 1. The velocities are 514 and 1666 m/s, respectively. The thickness under sources A, B and C are 3.8, 4.3 and 4.6 m, respectively. In Line 3 we found that there are two layers with the velocities of 362 and 1989 m/s, respectively and the thickness underneath the source points B, C and D are 2.4, 1.9 and 1.6 m, respectively. The results from SIP are shown in **Figure 11**.
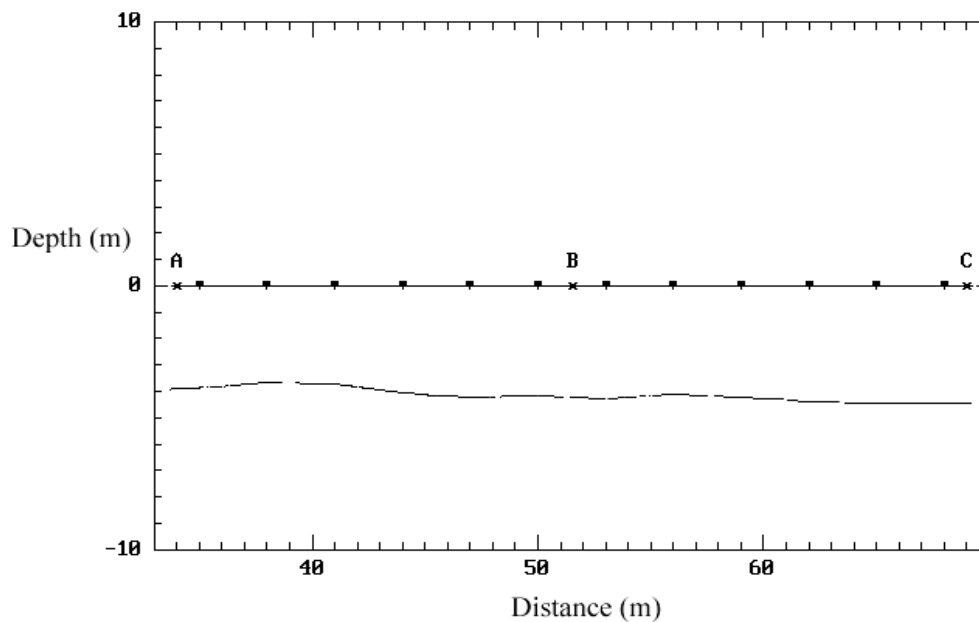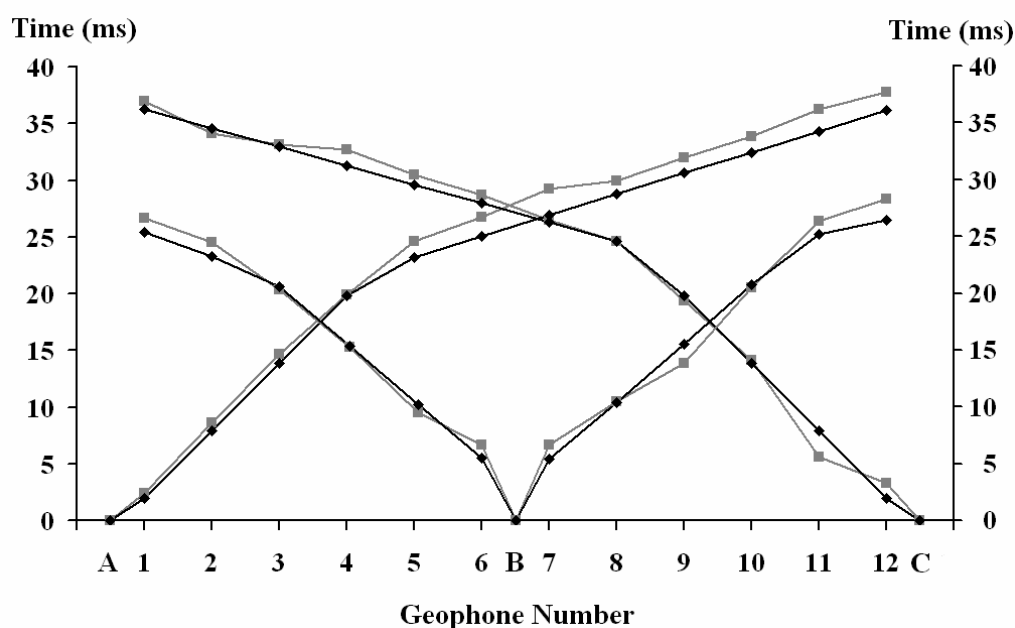

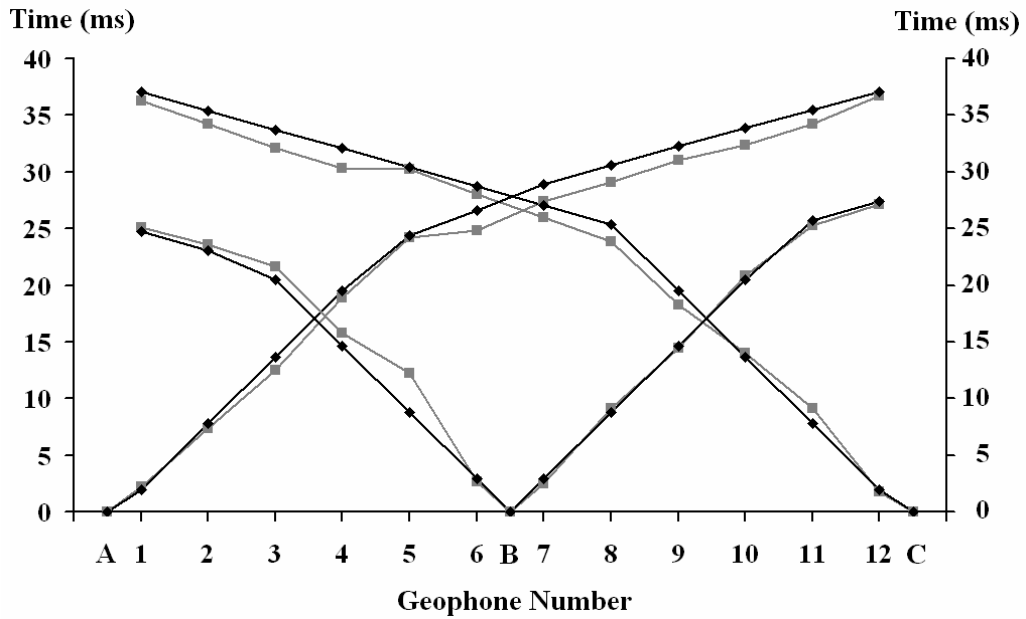(a) The structure from line 1.
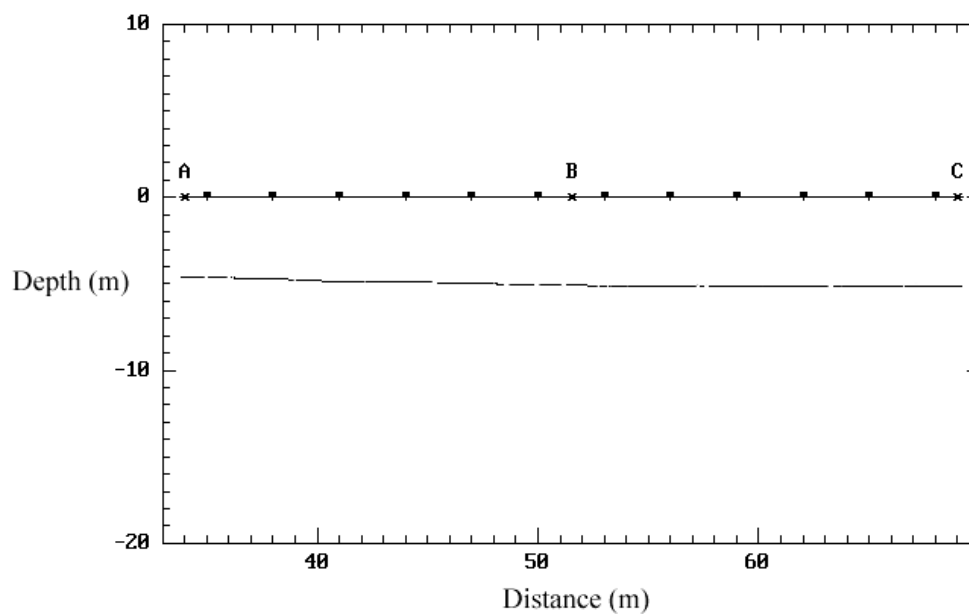
(b) The structure from line 2.
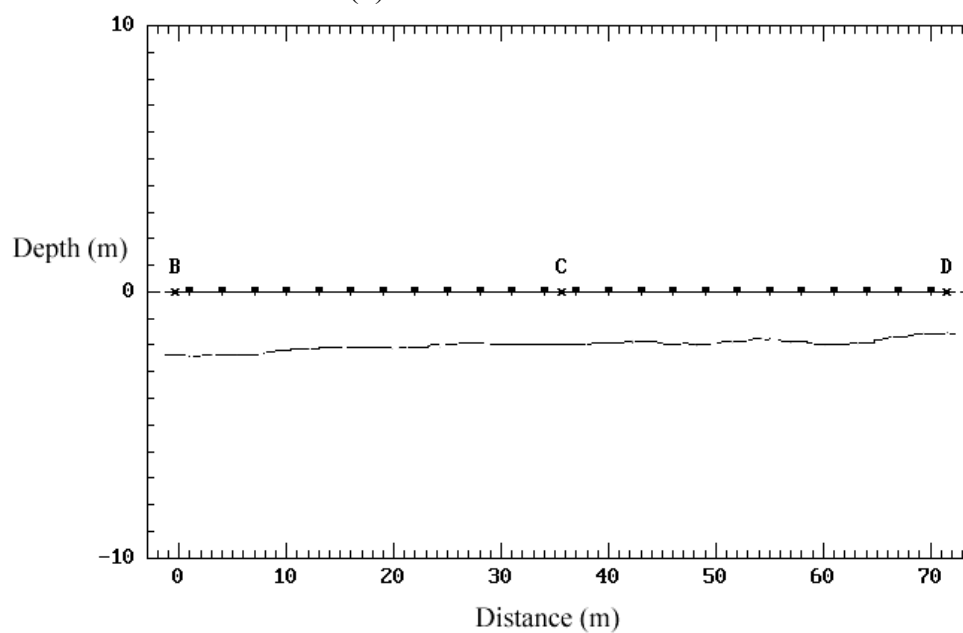
(c) The structure from line 3.

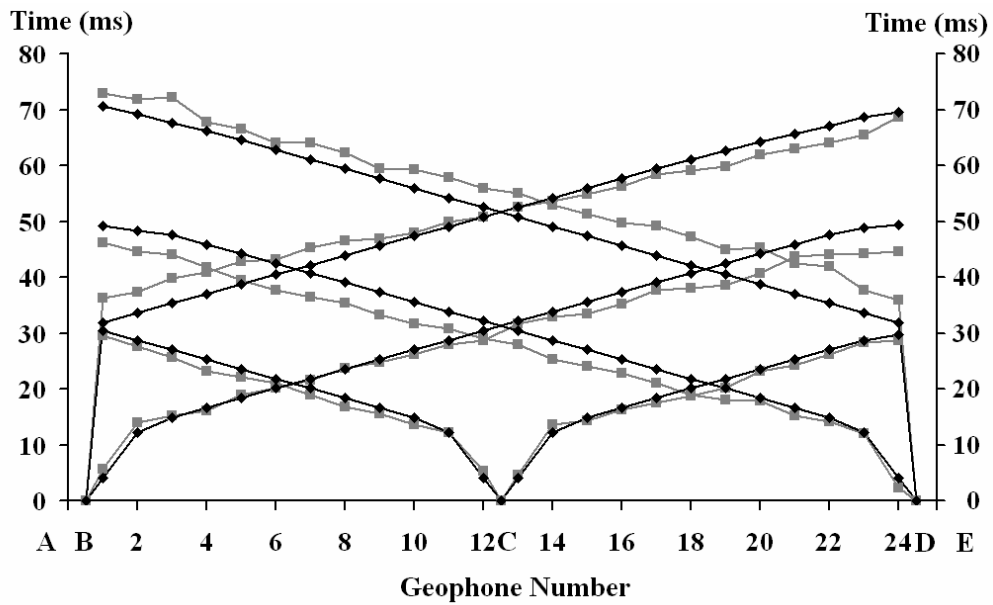**Figure 11** The structures from field data found by SIP.

When using the GA, the results from Line 1 show that there are three layers with velocities of 512, 1378 and 1821 m/s, respectively. The thickness of layer 1 underneath sources A, B and C are 4, 4.3 and 4.6 m while for layer 2 are 4, 2.47 and 0.9 m, respectively. For Line 2, the two layers found had velocities of 504 and 1708 m/s and their thickness underneath sources A, B and C are 4, 4.3, and 4.6 m, respectively. Line 3 was also found with three layers, with velocities being 367, 1744 and 1990 m/s, respectively. The thickness of layer 1 underneath sources B, C and D are 2, 2 and 2 m while for layer 2 they are 13, 11.7 and 10.5 m, respectively. **Figure 12** shows the comparisons between the t-x data from the field work and calculated t-x data while **Figure 13** shows the structure from field data interpreted by GA.
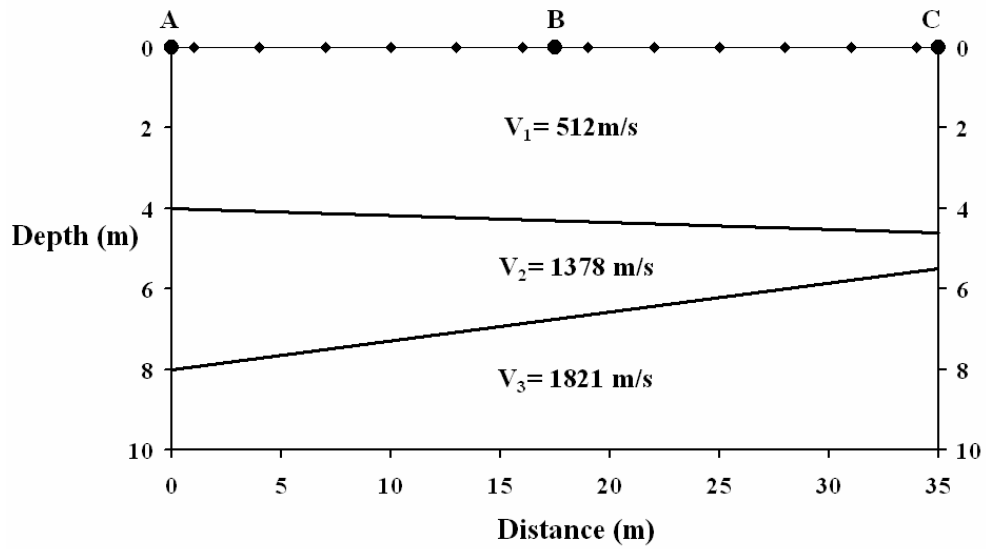

(a) Graph from line 1.
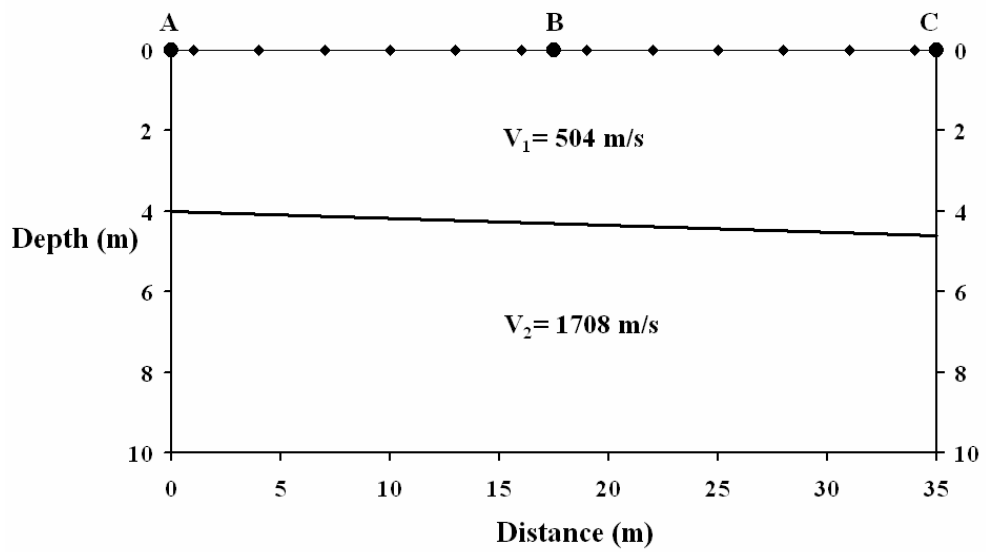
(b) Graph from line 2.

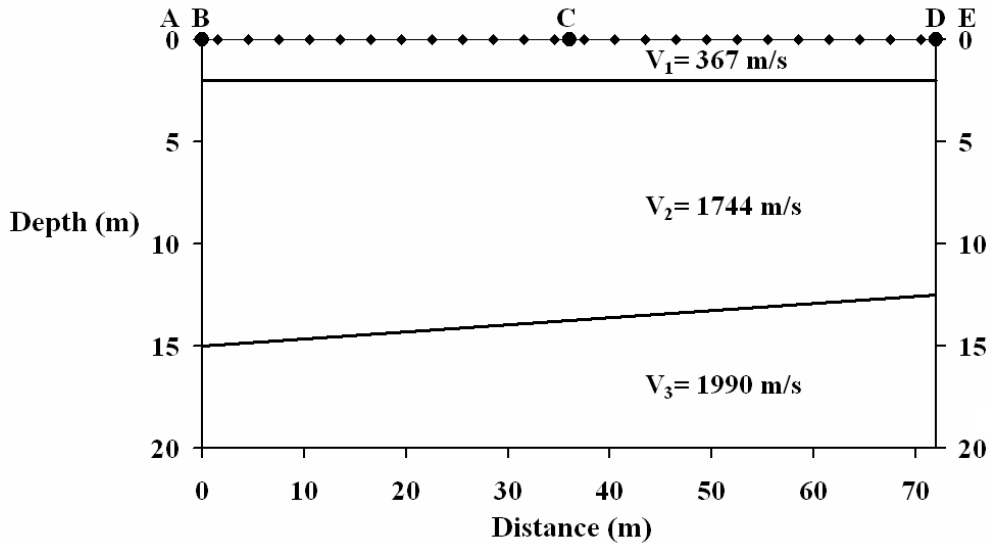

(c) Graph from line 3.

| | The actual t-x graph from field | | The calculated t-x graph from GA |

Figure 12 The comparisons between t-x data from field work and calculation.

(a) The structure from line 1.

(b) The structure from line 2.

(c) The structure from line 3.

Figure 13 The structure from field data found by GA.

DISCUSSION AND CONCLUSIONS

The results showed that SIP and GA give similar results. But in line 1 and 3, the structures from SIP and GA have a difference in the number of layers. This is because the SIP interpretation uses the opinion of interpreters for the number of layers while GA finds the best structure according to the data. In SIP, the number of layers was inputted as two because from the graph it implied as such. However, a different interpreter may give a different number of layers. The advantages of GA can be seen clearly here as there is no bias in the interpretation. GA finds the best fit data regardless of the user's opinion and is the option for the interpretation when experts are not available.

For some structures where the interfaces are not smooth, the results from SIP seem to be more accurate because of the limitation of the fitness function of the GA method, while with smooth interfaces, such as those in the synthetic data, GA appears to be more accurate. If the fitness function of GA was modified to calculate non-smooth interfaces, the result may be expected to be better especially for field data. However, with SIP, the user has to input the number of layers, which varies from user to user, while GA does not require such a parameter. For GA, the only input for the program is the t-x graph and the program will yield results accordingly. If there is no time restriction, the longer the program runs, the better the results will be.

ACKNOWLEDGEMENTS

REFERENCES

[1]   D Palmer. A New Direction for shallow refraction seismology: integrating amplitudes and traveltimes with the refraction convolution section. *Geophys. Prospect.* 2001; **49**, 657-73.

[2]   J Holland. *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975, p. 1.

[3]   Z Michalewiz, ZC Janikow and BJ Krawczyk. A modified genetic algorithm for optimal control problem. *Comput. Math. Appl.* 1992; **23**, 83-94.

[4]   A Homaifar, S Guan and GE Liepins. Schema analysis of the traveling salesman, problem using genetic algorithms. *Complex Syst.* 1992; **6**, 183-217.

[5]   D Abramson and J Abela. 1992, A parallel genetic algorithm for solving the school timetabling problem. *In*: Conference of the 15th Australian Computer Science, Hobart, Australia. 1992, p. 1-11.

[6]   JM Ahuaczin, EG Talbi, P Bessiere and E Mazer. Using genetic algorithms for robot motion planning. *In*: Proceedings of the 10th European Conference on Artificial Intelligence, Vienna, Austria. 1992, p. 671-5.

[7]   S Forrest, B Javornik, RE Smith and AS Perelson. Using genetic algorithm to explore pattern recognition in immune system. *Evol. Comput.* 1993; **1**, 191-211.

[8]   TDM Purdin and G Harris. A genetic algorithm approach to solving crossword puzzles. *In*: Proceedings of the 1993 ACM/SIGAPP symposium on applied computing: states of the art and practice, Indiana, United States. 1993, p. 263-70.

[9]   VV Savchenko and LM Schmitt. Reconstruting occlusal surface of teeth using a genetic algorithm with simulated annealing type selection. *In*: Proceedings of the 6th ACM Symposium on Solid Modeling and Applications, Michigan, United States. 2001, p. 39-46.

[10] M Aurnhammer and KD Tonnies. A genetic algorithm for automated horizon correlation across faults in seismic images. IEEE *Trans. Evol. Comput.* 2005; **9**, 201-10.

[11] DA Coley. *An Introduction to Genetic Algorithms for Scientists and Engineers,* World Scientific Publishing, Singapore, 1999, p. 2.

[12] HJ Stephen. Interpretation of split-spread refraction data in terms of plane dipping layers. *Geophysics* 1976; **41**, 418-24.

บทคัดย่อ

ชัยวัฒน์ เลิศวิริยะนันทกุล[1] ภัทร อัยรักษ์[1] และ วรวุฒิ โลหะวิจารณ์[2]

การประยุกต์ใช้อัลกอลิทึมทางพันธุศาสตร์สำหรับแปลความคลื่นไหวสะเทือนด้านการหักเห

      การค้นหาโครงสร้างธรณีวิทยาใต้ดินโดยใช้คลื่นไหวสะเทือนด้านการหักเหเป็นวิธีการทาง ธรณีฟิสิกส์โดยทั่วไป ซึ่งอาศัยความแตกต่างของความเร็วของคลื่นที่เคลื่อนที่ในชั้นดินที่แตกต่างกัน การวิเคราะห์ข้อมูลดังกล่าวอาศัยเวลาในการเคลื่อนที่ และระยะทางจากต้นกำเนิดคลื่นถึง ตัวรับ สัญญาณ ซึ่งจะแสดงในรูปกราฟ t-x ในงานวิจัยนี้ได้เพิ่มเติมการวิเคราะห์ข้อมูลโดยใช้วิธีทางอัลกอริทึม ทางพันธุศาสตร์ (Genetic Algorithm, GA) ในการแปลความข้อมูลคลื่นไหวสะเทือนด้านการหักเห ใน การศึกษานี้จะศึกษากรณีที่โครงสร้างใต้ดินมีสอง และสามชั้น ทั้งชนิดของรอยต่อแบบขนาน และทำมุม เอียงกับพื้นดิน การประเมินความถูกต้องได้ทำโดยการใช้ข้อมูลในแบบจำลอง และข้อมูลจริง โดยการ เปรียบเทียบผลที่ได้กับโปรแกรมสำเร็จรูปที่นิยมใช้อย่างแพร่หลายซึ่งได้แก่ Seismic Interpretation Program (SIP) ผลที่ได้พบว่า GA สามารถแปลความของข้อมูลได้ดีสำหรับข้อมูลที่สังเคราะห์ขึ้น และมี ความใกล้เคียงกับผลที่ได้จาก SIP ในกรณีการวิเคราะห์ข้อมูลจริงจากภาคสนาม

[1] หน่วยวิจัยนวัตกรรมทางฟิสิกส์และทรัพย์สินทางปัญญา ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ อำเภอหาดใหญ่ จังหวัดสงขลา 90112

[2] หน่วยวิจัยธรณีฟิสิกส์ ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ อำเภอหาดใหญ่ จังหวัดสงขลา 90112