

PAPER • OPEN ACCESS

Binary Coded Genetic Algorithm (BCGA) with Multi-Point Cross-Over for Magnetotelluric (MT) 1D Data Inversion

To cite this article: E Wijanarko and H Grandis 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **318** 012029

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Binary Coded Genetic Algorithm (BCGA) with Multi-Point Cross-Over for Magnetotelluric (MT) 1D Data Inversion

E Wijanarko^{1,2*} and H Grandis²

¹ Research and Development Centre for Oil and Gas Technology “LEMIGAS”

² Faculty of Mining and Petroleum Engineering, Institut Teknologi Bandung (ITB),
Jalan Ganesha 10 Bandung 40132, Indonesia

*e-mail: akhedy_87@students.itb.ac.id

Abstract. The magnetotelluric (MT) 1D modelling has been continuously receiving interest due to its effectiveness in obtaining overall subsurface resistivity image of an investigated area. The advances in computational resources allow increasing implementations of non-linear inversion using global search approach, such as Genetic Algorithm (GA). The genetic algorithm adopts the process of natural selection (survival for the fittest) and genetic transformation, i.e. selection, reproduction, mutation and population change, to solve the optimization problem. This paper discusses GA application in MT 1D modelling using binary coding representation with multi-point cross-over, i.e. one for every model parameter. The model parameters are resistivity and thickness of homogeneously horizontal layers. The algorithm parameters need to be set to work properly, i.e. population size, number of genes, number of bits, crossover probability, mutation probability and number of generations. Despite binary coded GA (BCGA) drawbacks discussed in the literatures, we found that binary representation allows relatively extensive exploration of the search space. Test using synthetic data from three-layered synthetic models lead to satisfactory results, in terms of synthetic model recovery and data misfit comparable with the noise level contained the synthetic data.

1. Introduction

Magnetotelluric (MT) method is a frequency-domain electromagnetic (EM) technique that uses natural electric and magnetic fields as signal source to estimate the electrical resistivity of the Earth's subsurface. The basic concepts of MT were first developed independently by Tikhonov, Cagniard and Rikitake in 1950s [1]. Information on the subsurface resistivity structure plays important role in geothermal exploration due to resistivity-temperature relationship in rock formations [2]. Resistivity contrast between sedimentary layers and basement can also be assessed by MT in hydrocarbon exploration [3]. Resistivity changes occurred in oil and gas reservoir can also be monitored to some extent by using MT technique [4].

Highly non-linear function relating MT data (impedance variation with period) with model parameters (resistivity and thickness of layers) even for simple 1D model often leads to convergence problems in inversion method based on local approach. The iterative linearization techniques fail to find the global minimum of the misfit function if the starting model is too far from the true model [5]. To overcome the flaws of the classical linear inversion method, inversion method with global approach such as Monte Carlo based algorithms were employed for MT 1D and 2D inversions [6]. In this paper we describe an implementation of GA that has been used to solve non-linear optimization problems in many disciplines including geophysics [7]. In particular, a basic binary coded GA (BCGA) with multi-point cross-over was employed to obtain the best fit models. We used synthetic data associated with



simple three-layered models with *a priori* number of layers similar and different from (i.e. more than) the synthetic models.

2. Method

The Genetic Algorithm (GA) is an optimization method that adopts the evolutionary behaviour of biological systems, developed among others by Holland [8] and Goldberg [9]. In a binary coded GA, model parameters representing a solution to the optimization problem are encoded by binary strings of 0's and 1's referred to as a chromosome. The algorithm starts with a population consisting of a set of chromosomes randomly selected within the search space. The population undergoes genetic operations, i.e. selection, cross-over and mutation, leading to a new population that would be better than the old population.

Selection

Chromosomes from a population are selected based on their fitness with reference to the objective function to form pairs of parents. This selection process mimics the principles of "survival of the fittest". In the case of MT 1D inversion, the misfit between the observed data and calculated data for a particular model can be expressed by,

$$E = \frac{1}{2N} \sum_{i=1}^N \log \left(\frac{\rho_{a,i}^{cal}}{\rho_{a,i}^{obs}} \right)^2 + \beta (\phi_i^{cal} - \phi_i^{obs})^2 \quad (1)$$

where ρ , and ϕ are apparent resistivity and phase respectively, β is the weighting factor for phase relative to apparent resistivity data, while N is the number of periods in the data. One of possibilities to convert misfit to fitness of the model is simply by using,

$$F = \exp(-E) \quad (2)$$

The roulette wheel mechanism [8,9] can be used to select chromosomes based on their fitness. In principle, chromosomes or models with lower misfit, hence higher fitness, would have higher probability to be selected as parents for subsequent generations.

Cross-Over

New chromosomes (off-spring) are formed by a cross-over operator representing the mating process between selected pair of old chromosomes (parents). The position of bit in the binary string of the parents is determined randomly. Then, the selected bit's position is used as the pivot for the exchange between parts of string of the parents leading to new chromosomes which are combinations of old chromosomes (Figure 1). For multi-point cross-over, more bit positions are used for swapping of chromosomes elements. In MT 1D case, there is one cross-over point randomly selected for every string that represents each model parameter.

Mutation

In the process of natural selection, mutation is a random process whereby a part of a chromosome (allele or gene) is changed. In the binary coded GA, mutation is performed by flipping a randomly chosen element of a chromosome from 0 to 1 or vice versa. Mutation is necessary to explore the search space while avoiding premature convergent to a local minimum.

Algorithm Parameters

In addition to genetic operators, several algorithm parameters need to be set to work properly, i.e. population size, number of genes, number of bits, cross-over probability, mutation probability and number of generations. Some of them can be set as *a priori* with relatively insignificant effect to optimization or inversion results. A big population size explores the search space more exhaustive leading to longer computation time, while too small population size may induce premature convergence.

The number of genes is determined by the number of model parameters. A large number of parameters adds to the dimension of the model space and the complexity of the misfit function topography that would be more difficult to explore. The accuracy of the solution depends on the number of bits of the binary numbers and to the upper and lower bound of the model parameter value to which those binary numbers are mapped.

Selected parents are not automatically eligible to cross-over process. The possibility to obtain off-springs from a pair of parents is determined by the cross-over probability. In general, the cross-over probability is chosen to be relatively high, in the order of 0.8 to 0.9 that means almost all parents will be replaced by their off-springs in the new generation. On the contrary, the probability of mutation to occur is relatively small, only 0.5% to 1% most of the times.

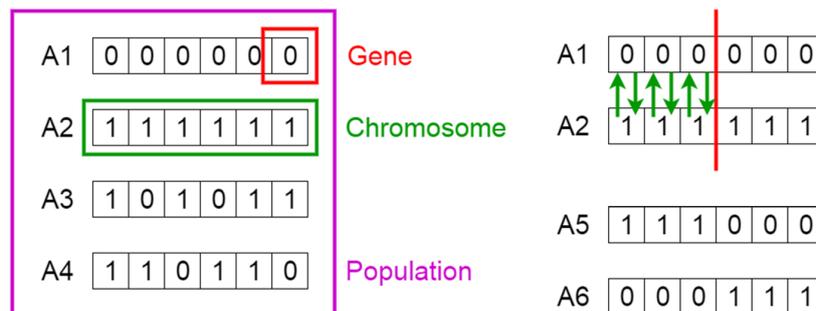


Figure 1. Illustration of gene, chromosome and population (left) and mechanism resulting in off-spring with a single point cross-over (right) in a binary coded GA.

3. Synthetic Models and Data

We used 1D synthetic models to create magnetotelluric (MT) synthetic data to test and validate our GA inversion method. Simple three-layered synthetic models (model-1 and model-2) associated with H-type and K-type sounding curves were used (see Table 1). Synthetic data were generated by performing 1D MT forward modelling algorithm based on the work of Hermance and Pedersen [5]. Gaussian noise with 10% standard deviation from the theoretical (true) values of independent real and imaginary part of the impedance was added to simulate field data. Apparent resistivity and phase synthetic data are in the period range between 0.001 to 1000 sec. with 31 samples regularly distributed in the logarithmic scale. We consider that the synthetic data represent typical full band MT data.

Table 1. Synthetic model parameters.

Layer	Model-1: H-type		Model-2: K-type	
	Resistivity (Ohm.m)	Thickness (m)	Resistivity (Ohm.m)	Thickness (m)
1	100	500	100	500
2	10	1000	1000	1000
3	1000	∞	10	∞

4. Results and Discussion

The GA was applied to invert synthetic data with some pre-determined parameters. The model parameters are resistivity and thickness of layers. The number of layers is set *a priori* similar with the

synthetic models (3 layers) and an over-estimate one (5 layers). The latter is intended to test the capability of the algorithm to handle erroneous guess for the number of layers and the equivalence problems as well. Each model parameter is represented or decoded to a 10-bit binary number that corresponds to a decimal number ranging from 0 to 1023. This interval is mapped 1 to 1000 for layer resistivity and from 100 to 1000 m for layer thickness.

After several preliminary tests, the population size of 200 with 500 generations are considered as adequate for MT 1D inverse problem. The cross-over and mutation probabilities are 0.9 and 0.01 respectively. The results with *a priori* number of layer identical to the synthetic models, i.e. 3 layers, are quite similar with the synthetic models associated both for H-type and K-type sounding curves. Therefore, they are of little interest to be presented here. Several attempts were also performed to invert the synthetic data using *a priori* number of layer greater than the synthetic models, i.e. 6 layers. The results were not satisfactory, i.e. the synthetic models were not recovered by the inverse models. In such case, the dimension of the model space was too high such that the exploration performed by the GA was not sufficiently exhaustive. It is also possible that the equivalence problems were too difficult to handle by such configuration of layers. Bigger population size with more generations might resolve the problems with larger *a priori* number of layers. However, other *a priori* values should also be set accordingly, i.e. *a priori* interval for minimum and maximum of the model parameter values.

The use of models with 5 layers are considered as an intermediate trade-off between too small and too large *a priori* number of layers. The results of inversion with 5 layers are presented in Figure 2 and Figure 3 for H-type (model-1) and K-type (model-2) sounding curves respectively. The inverse models are average models from all models at the last generation. The relative misfit between calculated response of the inverse model from the synthetic data are 5.99% for model-1 and 6.99% for model-2. The difficulties in resolving the depth and resistivity of the last layer in model-1 is due to screening effect of the low resistivity layer of the second layer (Figure 2). Similar problems are also observed with model-2 resulting in a relatively higher misfit (Figure3). However, the GA for 1D MT inversion can handle problems related to incorrect *a priori* number of layers and equivalence relatively well.

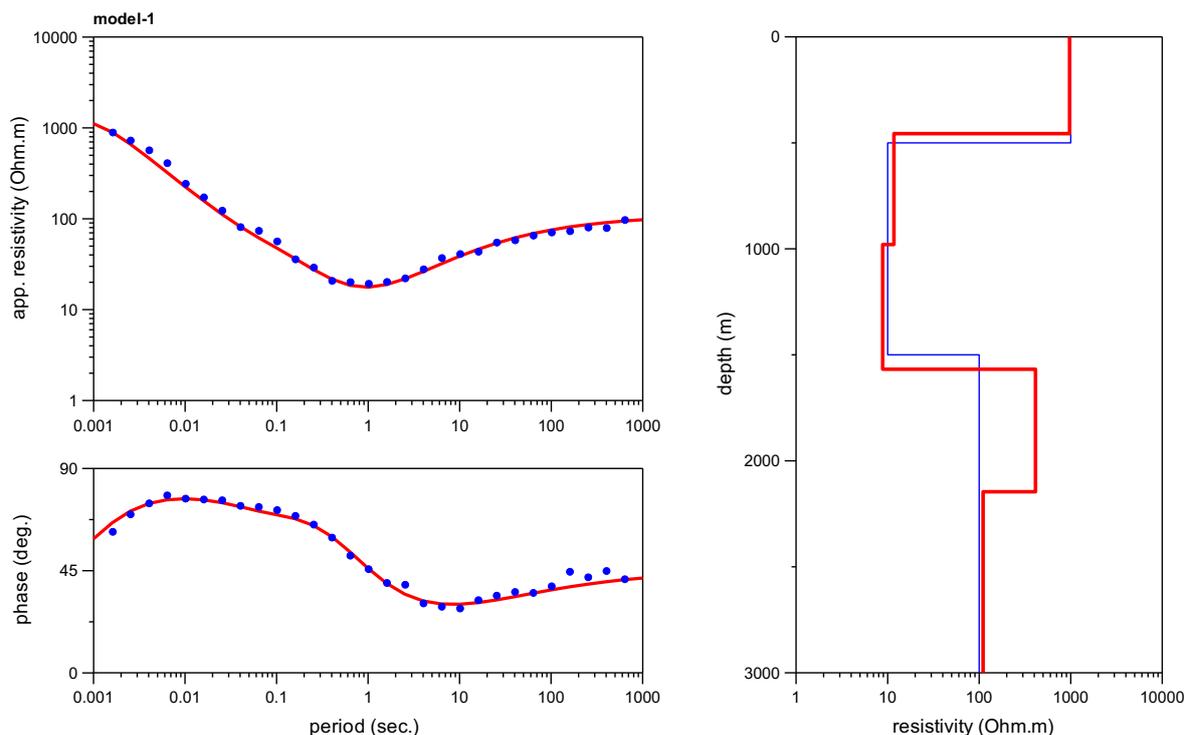


Figure 2. Comparison of synthetic data in blue dots and inverse model response in red line (left) and comparison of inverse model in red with the synthetic model in thin blue line (right) from inversion of synthetic data with H-type sounding curve (model-1).

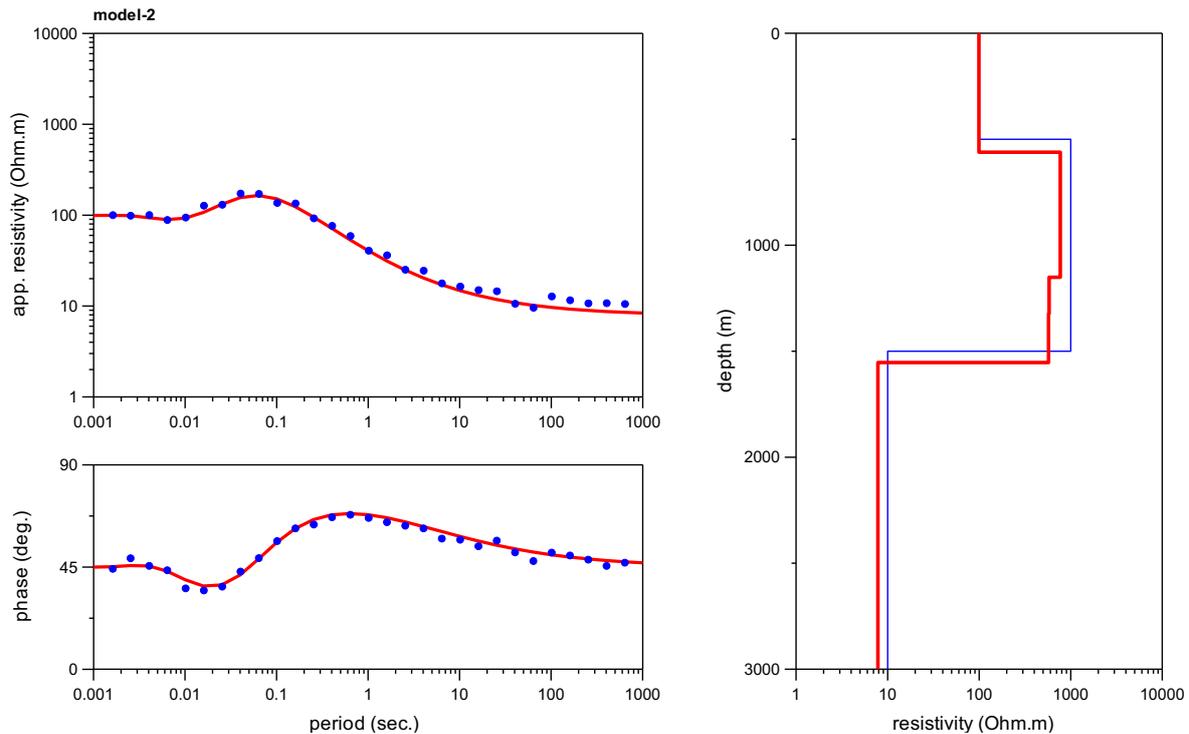


Figure 3. Comparison of synthetic data in blue dots and inverse model response in red line (left) and comparison of inverse model in red with the synthetic model in thin blue line (right) from inversion of synthetic data with K-type sounding curve (model-1).

5. Conclusion

The genetic algorithm solves 1D MT inverse problem with a global approach, i.e. without linearization. With a large number of chromosomes involved in GA at every iteration or generation, it is possible to estimate the uncertainty of the model parameters by calculating the variability or standard deviation of models at the last generation. However, the variability of models at the last generation is too small, i.e. all chromosomes are quite similar, such that they can lead to very small but unrealistic standard deviation of the model parameters.

Applications of the GA to invert synthetic data result in satisfactory inverse models and data fit, even for erroneous *a priori* number of layer involved. In fact, the number of layers can be estimated relatively easy from the data (sounding curve) and usually the least number of layers are preferred. Despite binary coded GA (BCGA) drawbacks discussed in the literatures, we found that binary representation allows relatively extensive exploration of the search space.

References

- [1] Simpson F and Bahr K 2005 *Practical Magnetotelluric* Cambridge University Press.
- [2] Muñoz G 2014 *Surv. Geophys.* **35** 101-122.
- [3] Sarvandani M M, Kalateh A N, Unsworth M and Majidi A 2017 *J. Petrol. Sci. Eng.* **149** 25-39.
- [4] Ogaya X, Ledo J, Queralt P, Jones A G and Marcuello A 2016 *J. Appl. Geophys.* **132** 100-115.
- [5] Pedersen J and Hermance J F 1986 *Surv. Geophys.* **8** 187-231.
- [6] Grandis H, Menvielle M and Roussignol M 2013 *Int. J. Geoph.* 531473.
- [7] Boschetti F, Dentith M and List R 1997 *Geophys. Prosp.* **45**(3) 461-478.
- [8] Holland J H 1975 *Adaptation in Natural and Artificial Systems* University Michigan Press.
- [9] Goldberg D 1989 *Genetic Algorithm in Search Optimization and Machine Learning* Addison-Wesley.