

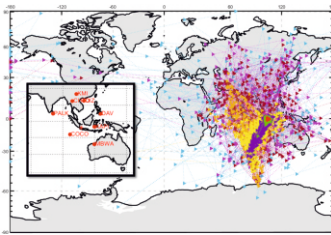
# Rapid Hazard Assessment

7

## Earthquake Location using Genetic Algorithm

Ajit Kundu\* and Pratap Mane

Seismology Division, Electronics and Instrumentation Group, Bhabha Atomic Research Centre, Trombay – 400085, INDIA



Optimization in 2D search space

### ABSTRACT

For achieving accurate, bias-free earthquake locations genetic algorithm (GA) has been used. This algorithm minimizes the misfit between observed and calculated arrival times of P waves recorded on a multi-station network. The location parameters of 16 earthquakes that occurred at and around the Sumatra region have been estimated using P onsets recorded at around 10 stations of the Global Seismographic Network (GSN) surrounding the Sumatra region. The study shows that most of the estimates are accurate enough except for one event which reclines far outside the network. This method is very efficient for estimating the location of events that occurs well inside the recording network.

KEYWORDS: Earthquake, Genetic algorithm, Global Seismographic Network

### Introduction

The most fundamental problem in seismology is the estimation of earthquake location. The earthquake location is defined by the earthquake hypocentre and the origin time. The hypocentre is the physical location, usually given in latitude, longitude, and depth below the surface. Reliable knowledge of hypocentre and origin time has tremendous importance for the studies such as the determination of earthquake magnitude, focal mechanism, and stress conditions around the source which are crucial for rapid hazard assessment and emergency responses.

The shape of the seismograms depends vastly on the relative positions of the recording instruments and source. The process of earthquake location using these seismograms is an inverse problem. The depth-dependent gradients and discontinuities in the velocity distribution within the Earth make this inverse problem highly non-linear and non-trivial to solve.

Mathematical fundamentals of earthquake location methods were established more than a century ago. These are namely, graphical methods [1], simple grid searches [2] and the linearized inversion method [3]. Since then various location methods have been developed to improve the accuracy of the location. These improvements have enabled semi-automation of the event location process, which is employed at many earthquake centres worldwide (e.g. the International Seismological Centre (ISC), the United States Geological Survey (USGS), and National Earthquake Information Centre (NEIC)).

Based on the number of earthquakes to locate, location techniques can be broadly classified into two groups, namely, single-event location techniques [4-10] that locate one event at a time and multiple-event location techniques [11-13] that locate multiple events simultaneously. Depending on the spatial type, there are the absolute location that is determined with respect to a fixed geographic coordinate system and a

fixed time standard, and the relative location that is computed with respect to another event (an earthquake or explosion) which might have an uncertain absolute location [4]. Moreover, relying on the methodology employed, location techniques can be linear, nonlinear and probabilistic.

The genetic algorithm (GA) is a global optimization technique and a useful tool for solving non-linear problems with many local minima or maxima. It was developed by Holland [14]. For the present study, GA has been used to minimize the misfit between observed and calculated arrival times to determine earthquake location parameters. The novel feature of GA is that it works with many estimates of the location at once and uses the properties of the cluster of estimates to drive the optimization process and reach finally to the global minimum. In this study, 16 events of various depths and magnitudes ( $M_w > 6$ ) from in and out of the Sumatra region have been selected. The selected (around 10) recording stations of GSN well surrounds the Sumatra region. GA generates a pool of location parameters, uses a 3D velocity model to compute arrival times (i.e., P onsets) at the selected stations and minimizes the misfit between observed and calculated arrival times to estimate the location parameters. The study reveals that the estimated parameters are accurate

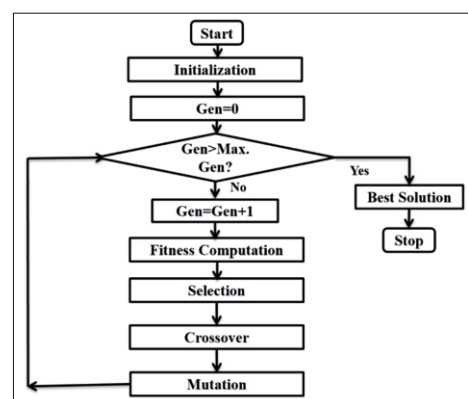


Fig.1: Working of GA. Gen and Max. Gen represent generation and maximum value of generation respectively.

\*Author for Correspondence: Ajit Kundu  
E-mail: ajitkun@barc.gov.in

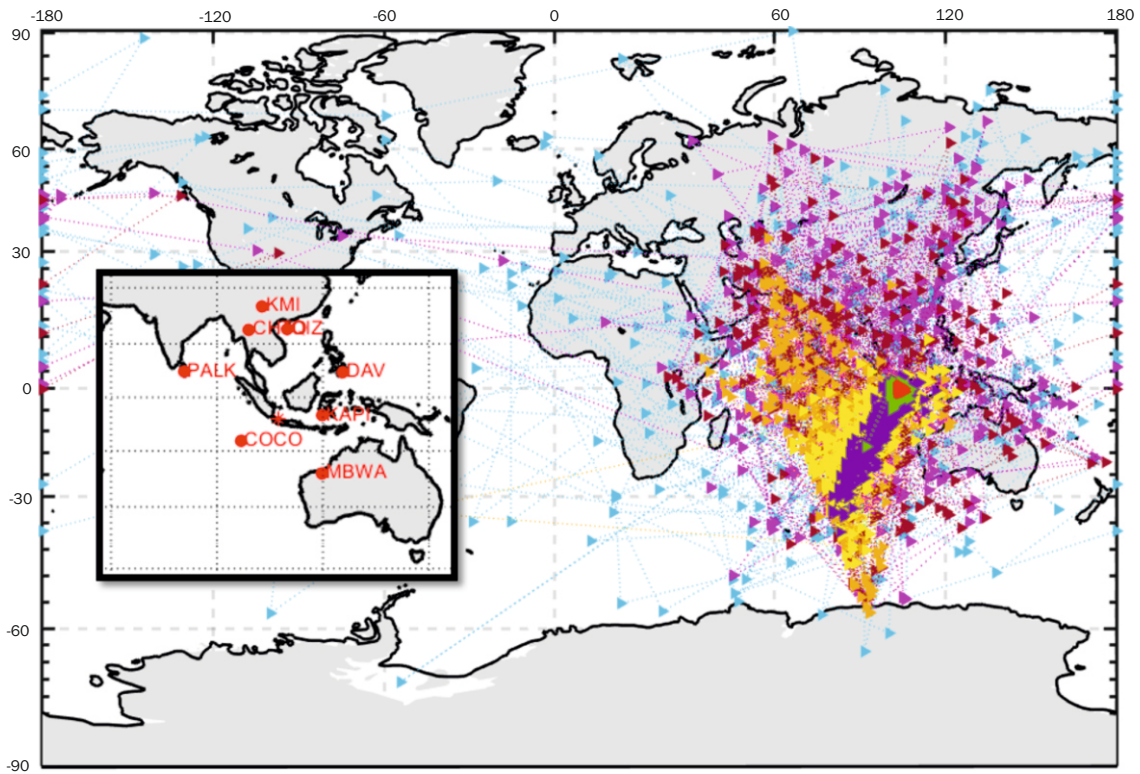


Fig.2: Optimisation in 2D search space. Triangles with colours cyan, magenta, tamarillo, orange, yellow, indigo, green and red represent search points corresponding to generations 1, 2, 3, 5, 6, 7, 9 and 11 respectively.

enough for those events that are well inside or very near to the network. For the events far outside the network, depth and origin time estimations are quite erroneous though the estimated latitude and longitude are reasonably good. This signifies that the events occurring inside the monitoring network are accurately located using this technique.

**Methods**

For accurate estimation of location of seismic events using genetic algorithm, the design of fitness function (FF) plays a vital role. In this study it is formulated as

$$FF = \text{std} (t^o_{arrival} - t^c_{arrival})$$

where std is standard deviation,  $t^o_{arrival}$  is the arrival time of P waves observed at recording stations and  $t^c_{arrival}$  is the calculated arrival times at those stations using the GA estimated locations parameters (i.e., latitude, longitude, depth and origin time) and 3D velocity model.

**Working of GA**

The working of GA has been recounted in a flow chart as portrayed in Fig.1. GA actuates with the random initialization of a population which propagates itself through selected criteria and is changed by the application of well-developed genetic operators. The population consists of chromosomes and the number of chromosomes is called population size. A chromosome is a collection of bits representing the location parameters. The initial search domains for latitude, longitude, depth and origin time are  $-90^\circ$  to  $90^\circ$ ,  $-180^\circ$  to  $180^\circ$ , 0 to 794 km and 30 minutes before the onset of the first coming P wave to the onset of the last coming P wave in a network respectively. These search spaces are typically represented by 8, 9, 10 and 11 bits respectively. So a chromosome is a bit-string of length 38 ( $8+9+10+11=38$ ). The population size is mostly problem dependent. For the present study, it varies between 340 to 345 which are obtained on a trial basis. Thus, GA starts with a population that consists of 340 (say typically) chromosomes each 38-bit long. A single iteration of a genetic

algorithm is based on three stages namely, selection (or replication), crossover (or recombination) and mutation. Selection begins with the evaluation of FF. For this, each chromosome is decoded to obtain the values of the location parameters. With knowledge of observed arrival times and

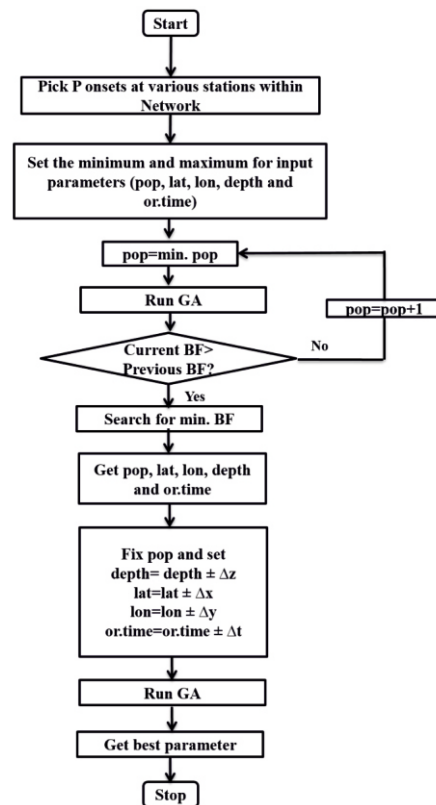


Fig.3: Location estimation method. The parameters pop, min, BF, lat, lon, and or.time represent the population size, minimum, best fitness, latitude, longitude and origin time respectively. The step values  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  and  $\Delta t$  correspond to latitude, longitude, depth and origin time respectively.

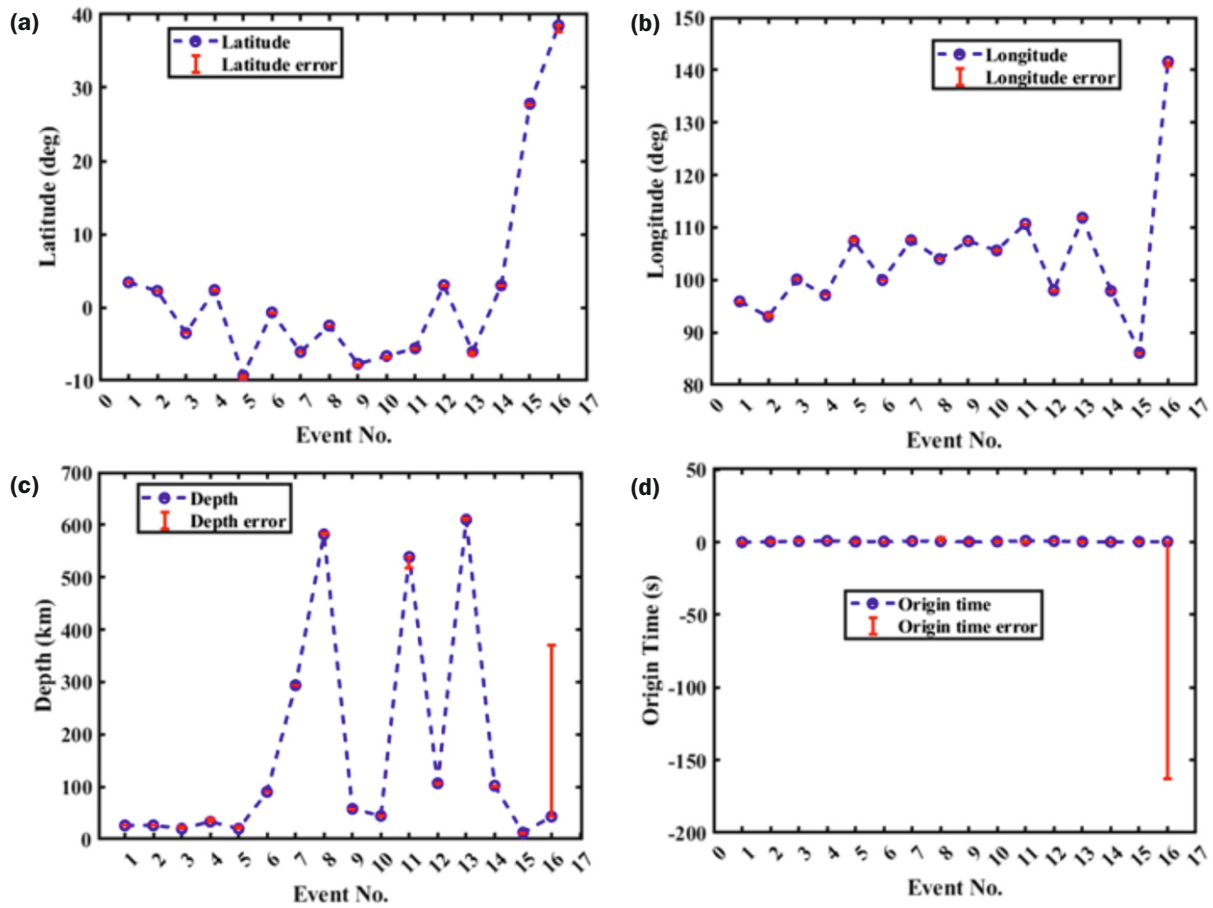


Fig.4: Errors in latitude, longitude, depth and origin time estimation for 16 events are plotted in (a), (b), (c) and (d) respectively. Origin time has been normalized to unity. The positive and negative error bars indicate over-estimation and under-estimation respectively.

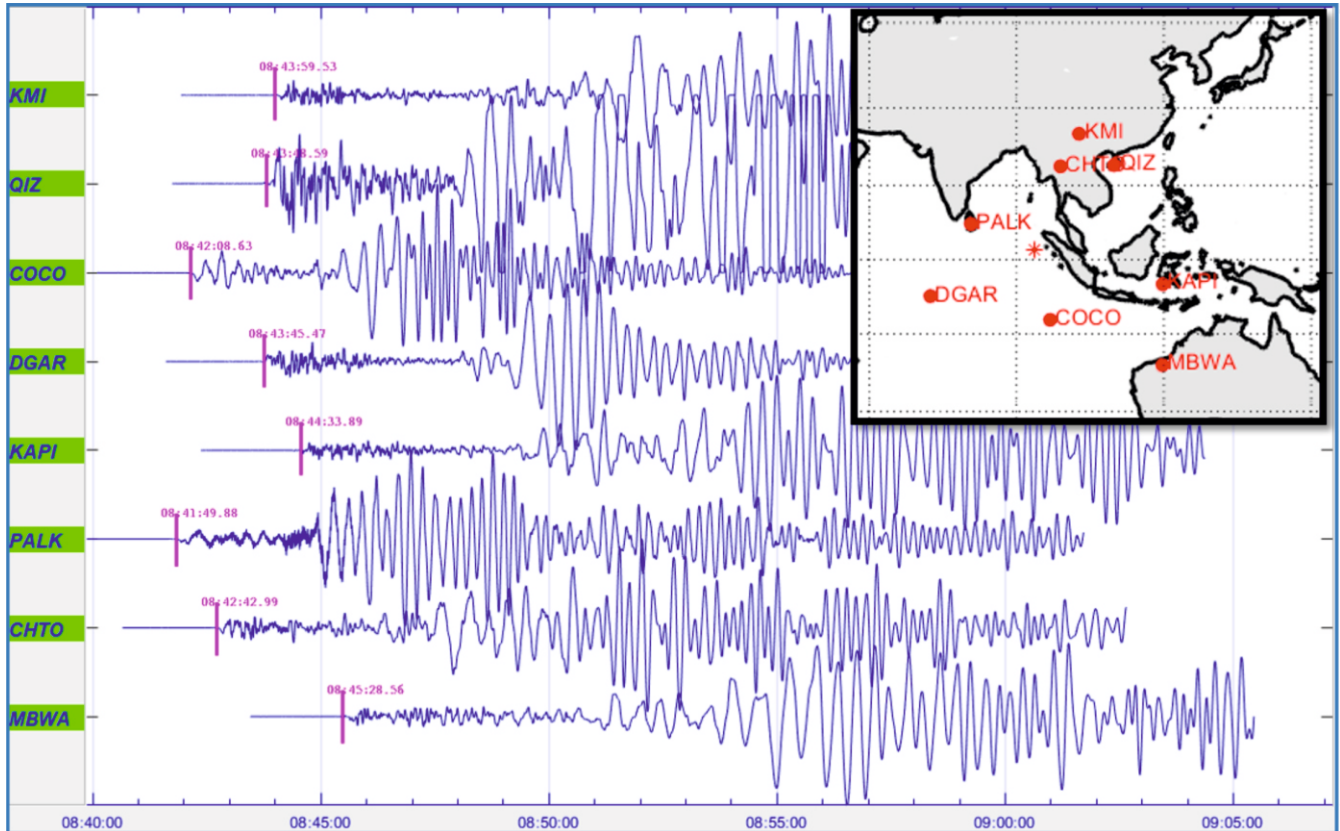


Fig.5: Signal plots with marked P onsets and the network (shown in right corner panel).

calculated arrival times (obtained using decoded location parameters and a 3D velocity model), FF is evaluated and assigned to each of the chromosomes in the population. The

selection of best chromosomes can be done using a variety of selection algorithms such as roulette wheel, proportionate selection, linear rank selection, tournament selection and

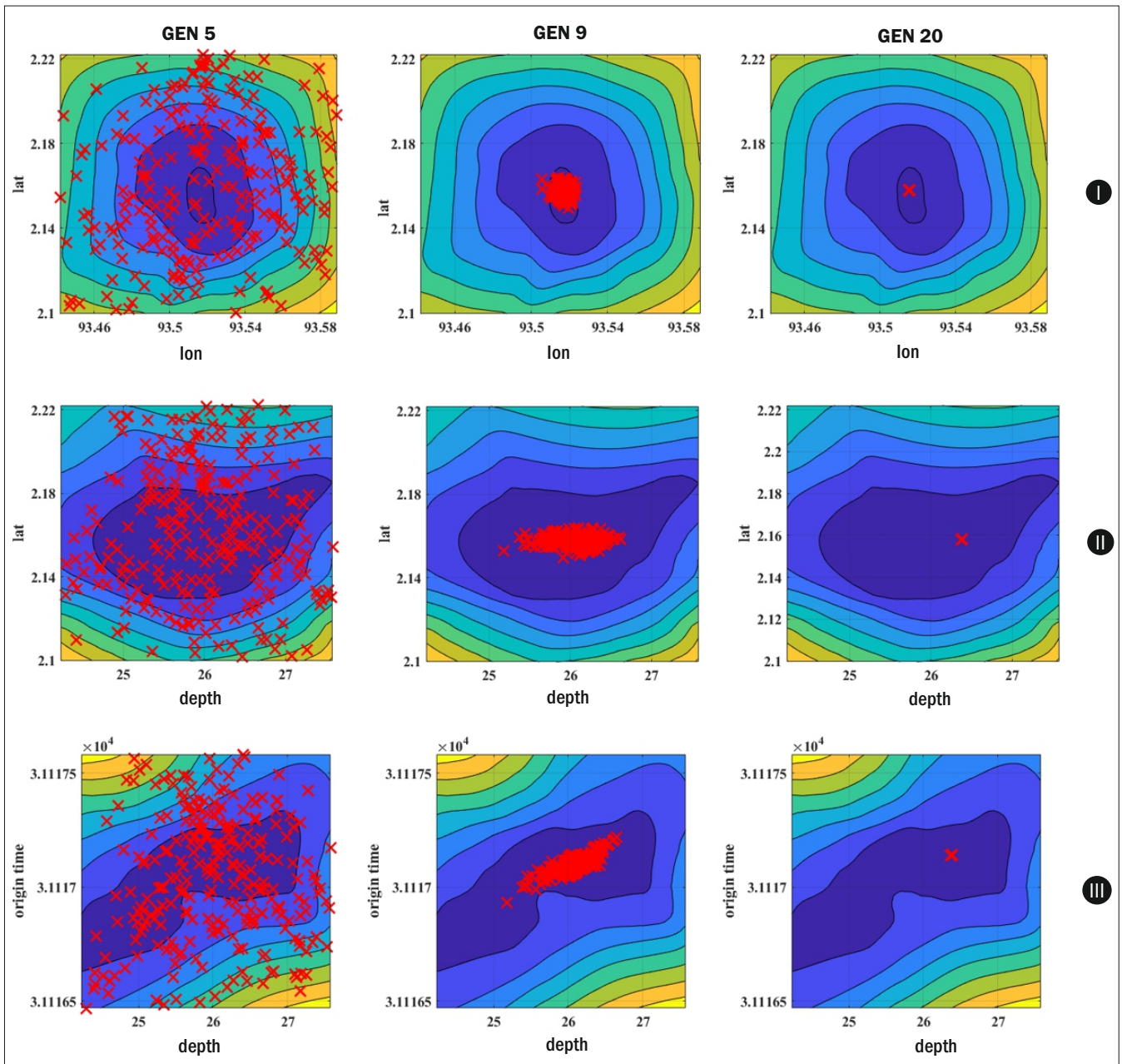


Fig.6: Horizontal panels I, II and III correspond to search spaces latitude-longitude, latitude-depth and origin time-depth respectively for three generation stages (GEN=5, 9 and 20).

stochastic remainder selection [15]. For the present study, the most popular tournament selection algorithm has been used. The primary objective of the selection stage is to emphasize better chromosomes in a population. It does not create any new chromosomes; instead, it eliminates the bad chromosomes and selects relatively good chromosomes from a population and makes multiple copies of them. The creation of new chromosomes is performed by crossover and mutation. The crossover stage basically produces better off-springs. It selects two parents based on crossover probability and randomly picks a point (called crossover point) between two genes to slice each of the parent chromosomes into two parts. Then it combines the first part of the first parent and the second part of the second parent to get the first off spring. Similarly, it combines the first part of the second parent and the second part of the first parent to get the second offspring. These off-springs will not be identical to any particular parent provided the parents are distinct but will instead combine the parental traits in a novel manner. The off-springs are expected to be better solutions (or chromosomes) as they are a result of

the recombination of highly fit parents. It is noted that in the present study the crossover operation uses the intermediate crossover function with probability 0.95. However, every crossover may not create a better solution. Bad solutions, if created, are likely to get eliminated during the next selection process and good solutions are likely to get more chances to achieve a crossover with other good solutions in subsequent generations. Thus more and more solutions in the population are likely to have similar chromosomes. Crossover is used to create new solutions from the population's genetic information. However, these new solutions have no un-inherited or new inheritance information and the number of alleles (e.g., Blue eyes (eye colour), brown hair (hair colour)) is constantly decreasing. This process results in the contraction of the population to one point without searching the promising part of the search space and GA may get trapped in local minima. Whereas, the one goal of every learning algorithm is to search always in regions not viewed before. Therefore, it is necessary to enlarge the information contained in the population. This goal is achieved by using mutation. The

Table 1: Event parameters as reported by IRIS along with the parameters estimated by GA.

Evt No.	Date	IRIS					GA			
		Or. Time	Lat (deg)	Lon (deg)	Depth (km)	Mw	Or. Time	Lat (deg)	Lon (deg)	Depth (km)
1	2004 -12-26	00:58:52	3.4125	95.9012	26.1	9	00:58:51.44	3.421	95.681	25.97
2	2012 -04-11	08:38:37	2.2376	93.0144	26.3	8.6	08:38:37.14	2.158	93.515	26.37
3	2010 -10-25	14:42:22	-3.5248	100.104	20	7.8	14:42:23.22	-3.465	100.19	23.73
4	2010 -04-06	22:15:02	2.3601	97.1113	33.4	7.8	22:15:01.63	2.276	97.075	39.78
5	2006 -07-17	08:19:26	-9.3178	107.424	20	7.7	08:19:26.71	-9.77	107.86	24.21
6	2009 -09-30	10:16:10	-0.7071	99.9678	90.2	7.6	10:16:9.45	-0.846	99.865	90.61
7	2007 -08-08	17:04:58	-6.089	107.584	293.8	7.5	17:04:58.22	-6.071	107.79	292.69
8	2004 -07-25	14:35:17	-2.4931	103.975	581.9	7.3	14:35:19.61	-2.489	104.05	585.43
9	2009 -09-02	07:55:01	-7.7346	107.411	57.8	7	07:55:01.19	-8.002	107.39	56.99
10	2000 -10-25	09:32:24	-6.655	105.619	44.3	6.8	09:32:24.17	-7.066	105.45	46.9
11	2020 -07-06	22:54:47	-5.5956	110.695	538.73	6.7	22:54:45.53	-5.821	110.69	517
12	2011 -09-05	17:55:12	3.0253	97.9991	106.6	6.7	17:55:12.31	2.781	97.904	108.02
13	2019 -09-19	07:06:33	-6.0708	111.842	610	6.1	07:06:33.66	-6.589	111.7	613.61
14	2012 -06-23	04:34:53	3.0022	97.9116	101.9	6.1	04:34:53.29	2.855	97.862	98.67
15	2015 -05-12	07:05:18	27.802	86.126	12.3	7.2	07:05:18.35	27.576	86.168	10.062
16	2021 -03-20	09:09:44	38.5	141.6	43	7	09:07:00.61	37.61	140.7	370

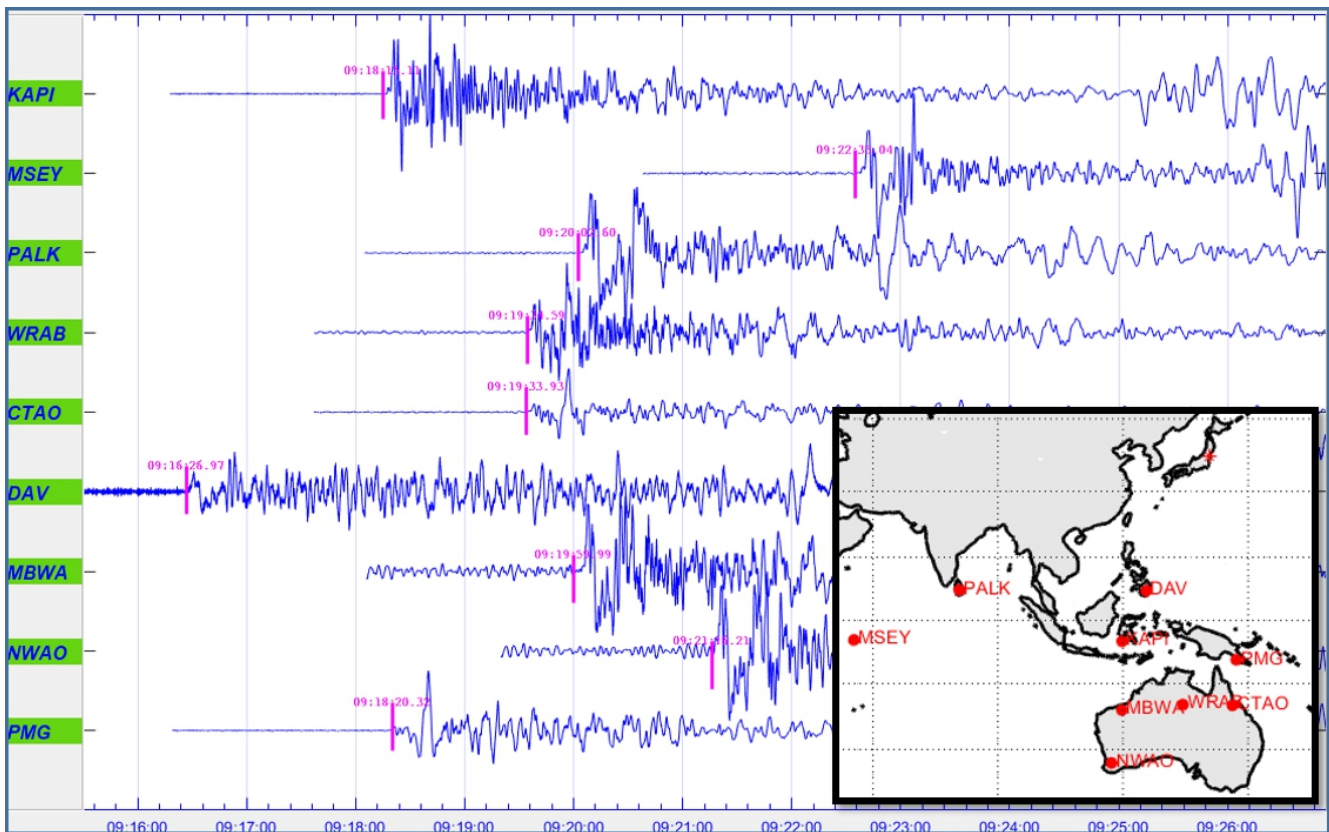


Fig. 7: Signal plots with marked P onsets and the network (shown in right corner panel).

mutation is carried out by flipping some bits of the chromosomes based on mutation probability. For the present study, the mutation operation makes use of the uniform mutation function with probability 0.002. It maintains the diversity of the population so that GA would not get trapped in partial solutions. The algorithm stops when stopping criteria (such as the number of generations reaches maximum generation) is met providing the best solutions i.e., the best estimates of location parameters.

To visualize the working of GA, a search space optimisation plot has been shown in Fig.2. To start with GA searches the whole space with search points indicated by the triangles with cyan colour. As the generations proceed the search space reduces and finally converge to a single point as shown by the triangle with red colour. The network chosen for this search space optimisation has been shown in the left panel of Fig.2. Using this optimisation capability of GA a method has proposed for earthquake location as described in the next section.

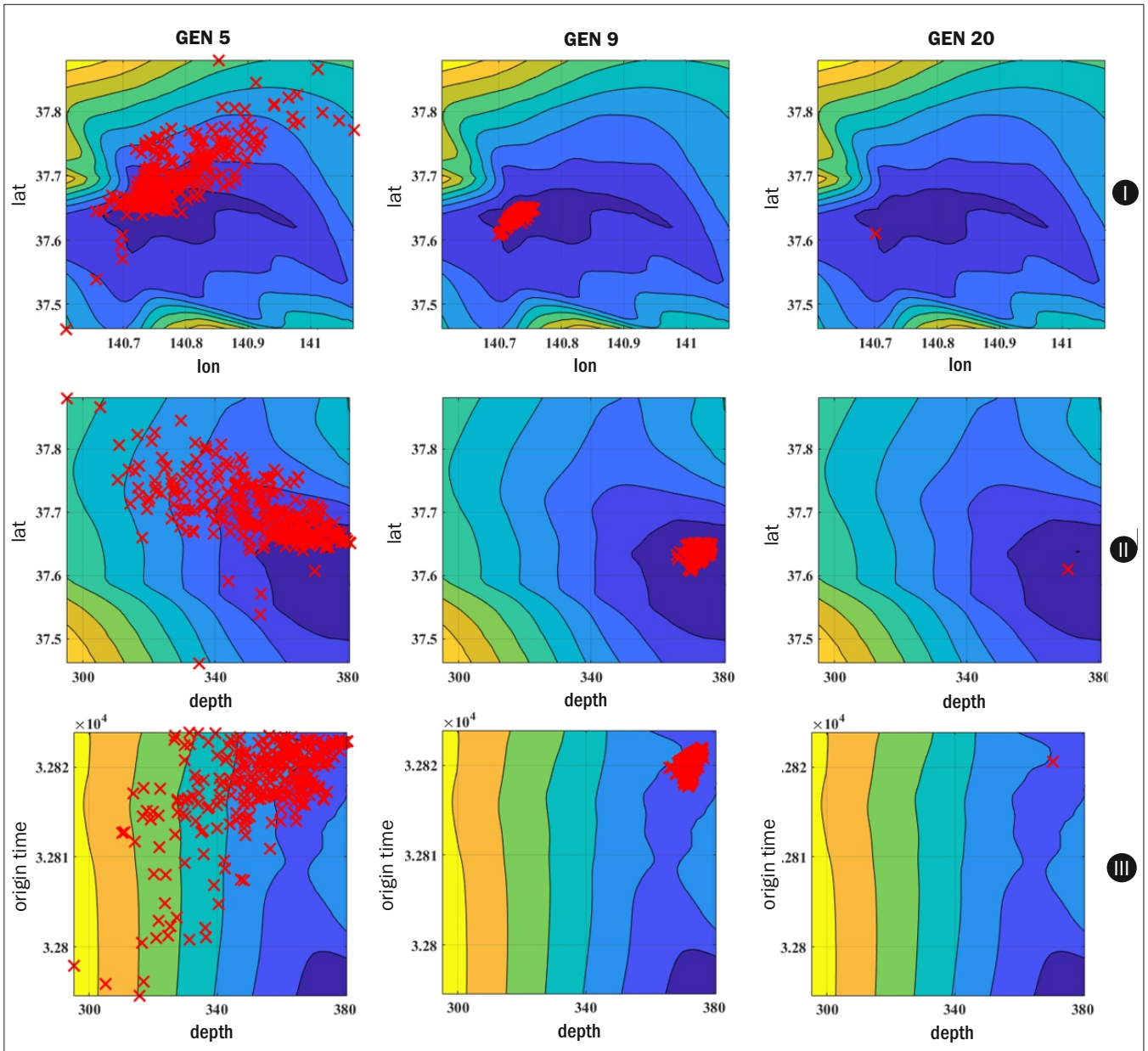


Fig.8: Horizontal panels I, II and III correspond to search spaces latitude-longitude, latitude-depth and origin time-depth respectively for three generation stages (GEN= 5, 9 and 20).

**Location estimation**

The method of estimation of location has been shown in Fig.3. Initially, onsets of P waves recorded at various stations of a network are picked up. Then, the maximum and minimum ranges of the parameters to be estimated are assigned as mentioned in section 2.1. The range for population size has been considered from 340 to 345. Fixing the parameter ranges, GA is started with the lowest population size. The best fitness (BF) value of the fitness function for each step of the population loop is checked. If it is found that the BF value of the current step is greater than that of the previous step the loop is broken and the minimum value of BF is searched to find the optimum values for the parameters considered in the problem. Fixing the population size at optimum value, the latitude, longitude, depth and origin time are reset at  $\text{latitude} \pm \Delta x$ ,  $\text{longitude} \pm \Delta y$ ,  $\text{depth} \pm \Delta z$  and  $\text{origin time} \pm \Delta t$  where  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  and  $\Delta t$  may be typically taken as  $3^\circ$ ,  $3^\circ$ , 20 km and 30 minutes respectively. With these new parameter settings, GA is again run to obtain the final estimates of the location parameters.

**Results**

To study the efficacy of the algorithm we have downloaded the waveform data of 16 events that occurred in and outside the Sumatra region from IRIS (Incorporated Research Institutions for Seismology) website as given in Table 1. Out of 16, 14 data (Events 1 to 14) were from the Sumatra region. Event 15 was from Nepal and Event 16 was from Japan. The recording stations were selected such that the network formed by them well surrounds the first fourteen events. Event 15 was very near to the network whereas Event 16 was approximately  $34^\circ$  away from the periphery of the network. All events were of magnitude (Mw) greater than 6 and depths ranging from 12.3 km to 581.9 km. The location parameters of these events as estimated by GA are also noted in Table 1. The errors involved in the estimation of location parameters are shown in Fig.4. It is observed that the first 15 events are almost accurately located by GA. These events are located inside the network except Event 15 which reclines outside the network but very near the boundary of the network. For event 16, though the latitude and longitude estimates are

reasonable, the depth and origin time estimates are very poor as seen in Fig.4(c) and (d). The reason is that this particular event is very far away from the network for which GA fails to minimize the fitness function reasonably. To visualize GA estimation in search space two events namely, event 2 (well inside the network) and event 16 (far away from the network) have been considered as a case study and are discussed below in detail.

#### **Case study 1: Event inside the network (Event 2)**

The recording network as well as the recorded signals with P onset marking has been shown in Fig.5. The optimization in the search spaces like latitude-longitude, latitude-depth and origin time-depth has been shown through contour plots for generations 5, 9 and 20 respectively in Fig.6. The figure clearly shows that at generation 20 all the search space converges to a single point almost at the middle of the contour leading to an accurate estimate of location parameters. A similar picture has been obtained for all events lying inside or near the boundary of the network.

#### **Case 2: Event outside the network (Event 16)**

The recording network and the recorded signals with P onset marking have been shown in Fig.7. The optimization in the search spaces, namely, latitude-longitude, latitude-depth and origin time-depth has been shown through contour plots for generations 5, 9 and 20 respectively in Fig.8. For this event, though the spread of the search space reduces to a point at generation 20 but not at the centre of the contour, especially in the origin time-depth space leading to poor estimates of depth and origin time. The reason behind this is that event 16 is remotely located from the network.

#### **Conclusions**

The present study takes the advantage of GA in estimating the location parameters. The advantage of GA for earthquake location is that they do not require derivative information and still have both a local and global search character. The study shows that for all the events located inside or near the boundary of the monitoring network, the location estimates are reasonably accurate whereas, for the event lying far outside the network, location estimates especially depth and origin time become erroneous. Thus we strongly believe that GA is a powerful tool for performing robust earthquake locations provided the monitoring region is well surrounded by recording instruments.

#### **Acknowledgements**

Authors are thankful to Incorporated Research Institutions for Seismology (IRIS) for their open source GSN data, which was employed in this study. Authors would like to thank Dr. S. Mukhopadhyay, Shri A. Vijaya Kumar, Shri Ashish Gupta and Shri Abin C. Joy for the useful discussions. Authors also thank the reviewer for his valuable suggestions.

#### **References**

- [1] Milne, J.: Earthquakes and Other Earth Movements. Appelton, New York (1986)
- [2] Reid, H.: The Mechanics of the Earthquake, Vol. II of The California Earthquake of April 18, 1906 (1910)
- [3] Geiger, L.: Probability method for the determination of earthquake epicenters from the arrival time only. Bull. St. Louis. Univ 8, 60–71 (1912).
- [4] Karasözen E, Karasözen B. Earthquake location methods. GEM-International Journal on Geomathematics. 2020 Dec;11(1):1-28.
- [5] Klein, F.: User's guide to HYPONVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes (2002)
- [6] Lee, W., Lahr, J.: HYPO71, a computer program for determining hypocenter, magnitude and first motion pattern of local earthquakes (1972)
- [7] Lin, G., Shearer, P.: The COMPLOC earthquake location package. Seismol. Res. Lett. 77(4), 440–444 (2006). <https://doi.org/10.1785/gssrl.77.4.440>
- [8] Lienert, B., Berg, E., Frazer, N.: "hypocenter": an earthquake location method using centred, scaled and adaptively damped least squares. Bull. Seismol. Soc. Am. 76, 771–783 (1986)
- [9] Pavlis, G.L., Vernon, F., Harvey, D., Quinlan, D.: The generalized earthquake-location GENLOC package: an earthquake-location library. Comput. Geosci. 30(9), 1079–1091 (2004). <https://doi.org/10.1016/j.cageo.2004.06.010>
- [10] Schweitzer, J.: HYPOSAT—an enhanced routine to locate seismic events. Pure Appl. Geophys. 158(1), 277–289 (2001)
- [11] Douglas, A.: Joint epicentre determination. Nature 215, 47–48 (1967)
- [12] Dewey, J.W.: Seismic studies with the method of joint hypocenter determination. Ph.D. thesis, University of California (1971)
- [13] Pujol, J.: Joint event location—the JHD technique and applications to data from local seismic networks. In: Thurber, C.H., Rabinowitz, N. (eds.) Advances in Seismic Event Location, pp. 163–204. Springer, Dordrecht (2000). [https://doi.org/10.1007/978-94-015-9536-0\\_7](https://doi.org/10.1007/978-94-015-9536-0_7)
- [14] Holland, J. H. (1975). Adaption in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor.
- [15] Chaturvedi DK. Soft computing. Studies in Computational intelligence. 2008;103:363-381.