Australian Journal of Basic and Applied Sciences, 4(8): 3354-3359, 2010
ISSN 1991-8178

Review of Genetic Algorithm Model for Suspended Sediment Estimation

[1]Omolbani Mohamad Rezapour, [2]Lee Teang Shui and [3]Amir Ahmad Dehghani

[1]Faculty of Engineering University Putra Malaysia
[2]Faculty of Engineering University Putra Malaysia
[3]Gorgan University of Agricultural Sciences and Natural Resources

Abstract: Number of attempts have been made to relate the amount of sediment transported by a river to flow conditions such as discharge, velocity and shear stress. However, none of the equations derived have received universal acceptance. Usually, either the weight or the concentration of sediment is related to the discharge. Fluxes of suspended materials collected at gauge stations are closely related to flow discharges. In the absence of continuous recorded suspended sediment concentration data, hydrologists have used different models such as rating sediment transport curves to define the water discharge, suspended sediment relationship and to estimate (predict) suspended sediment concentrations for use in flux calculations. Correct estimation of sediment volume carried by a river is very important for many water resources projects. Empirical relations such as sediment rating curves are often applied to determine the average relationship between discharge and suspended sediment load. This type of models generally underestimates or overestimates the amount of sediment. During recent decades, some black box models based on artificial neural networks have been developed to overcome this problem. GA has been applied to a wide range of problems in artificial intelligence, engineering and science applications, industrial, and mechanical models. The main purpose of this paper is literature review of Genetic Algorithm for suspended sediment estimation.

Key words: Genetic Algorithm, suspended sediment, model.

INTRODUCTION

A large part of the sediment in lowland rivers is transported as wash load. As wash load is a non-capacity load it often is modelled using empirical relations, such as the rating curve technique. Sediment rating curves in the form of a power function are derived for several locations along the river Rhine and its main tributaries, using different fitting procedures. Inaccuracies in estimated sediment loads are analysed, and spatial differences in the shapes of the fitted rating curves are related to watershed characteristics. In many lowland rivers a major part of the sediment is transported in suspension.

About 85% of this load consisted of silt and clay, i.e. wash load. It can thus be concluded that wash load plays an important role in the sediment transport in the river. As the finest fraction of the suspended sediment load often is a non-capacity load it cannot be predicted using stream power related sediment transport models. Instead, empirical relations such as sediment rating curves often are applied. A sediment rating curve describes the average relation between discharge and suspended sediment concentration for a certain location. It is well known fact that all reservoirs are designed to a volume known as ‘‘the dead storage’’ to fit the sediment income that will collect over a specified period called the economic life. The underestimation of sediment yield results in insufficient reservoir capacities while the overestimation will lead to over-capacity reservoirs.

Only the appropriate reservoir design and operation is sufficient to justify every effort to determine sediment yield accurately, but in sanitary engineering the prediction of river sediment load has an additional significance, especially if the particles also transport pollutants. On the other hand, the sediment can aggrades channel beds with excess sand and gravel for tens to hundreds of kilometers downstream. Such aggradations promote lateral migration of channels and may cause serious flooding during rainstorms, due to loss of channel capacity necessary to convey floodwaters. The assessment of the volume of sediment carries significance also for the flooding problem.

A number of attempts have been made to relate the amount of sediment transported by a river to flow conditions such as discharge, velocity and shear stress. However, none of the equations derived have received

Corresponding Author: Omolbani Mohamad Rezapour, Faculty of Engineering University Putra Malaysia
E-mail: nrezaii2000@yahoo.com 00601-73601749

universal acceptance. Usually, either the weight or the concentration of sediment is related to the discharge. These two forms are often used interchangeably. Empirical relations such as sediment rating curves are often applied to determine the average relationship between discharge and suspended sediment load. This type of models generally underestimates or overestimates the amount of sediment. During recent decades, some black box models based on artificial neural networks have been developed to overcome this problem. Due to the high costs associated with the Suspended sediment transport by rivers much research over the last 30 years has been dedicated to the development of techniques to minimise the capital costs associated with such infrastructure.

Overview of Genetic Algorithm:

Every organism has a set of rules, a blueprint so to speak, describing how that organism is built up from the tiny building blocks of life. These rules are encoded in the genes of an organism, which in turn are connected together into long strings called chromosomes. Each gene represents a specific trait of the organism, like eye colour or hair colour, and has several different settings. For example, the settings for a hair colour gene may be blonde, black or auburn. These genes and their settings are usually referred to as an organism's genotype. The physical expression of the genotype - the organism itself - is called the phenotype. When two organisms mate they share their genes. The resultant offspring may end up having half the genes from one parent and half from the other. This process is called recombination. Very occasionally a gene may be mutated. Normally this mutated gene will not affect the development of the phenotype but very occasionally it will be expressed in the organism as a completely new trait. Genetic algorithms were formally introduced in the United States in the 1970s by John Holland at University of Michigan. The Genetic Algorithm is a probabilistic search algorithm that iteratively transforms a set (called a population) of mathematical objects (typically fixed-length binary character strings), each with an associated fitness value, into a new population of offspring objects using the Darwinian principle of natural selection and using operations that are patterned after naturally occurring genetic operations, such as crossover (sexual recombination), mutation and survival of the fittest (Dulay, 2005). GAs are very different from most computer programs, which have well-defined algorithms for coming up with solutions to problems. The genetic algorithm approach is to generate a large number of potential solutions in a search space and “evolve” a solution to the problem.

GA is a prominent and powerful optimisation technique that has been applied successfully in many disciplines (Paz, 1998). It is a robust search technique that is based on concepts of natural selection and genetics. For this reason, the terminology used in GA is borrowed from genetics. Every model has its own model parameters. According to the genetics terminology, each model parameter is a gene, while a complete set of model parameters is a chromosome. Each GA run consists of a number of generations with constant population size of several model parameters sets. The process of GA begins with an initial population of a user-defined number of model parameter variables, of which values were chosen at random or using a pre-defined rule, within a specified parameter range. Each model parameter set is then evaluated by an objective function (e.g. simple least squares) to yield its fitness value. The second and subsequent generations are formed by combining model parameter sets with high fitness value from the previous (or parent) population using selection and sampling, crossover and mutation operations, to produce successively fitter model parameter sets or offsprings. The selection and sampling operation favours those parent parameter sets with high fitness value to those of lower fitness value in producing offsprings. The crossover operator exchanges model parameter values from two randomly selected parent model parameter sets to produce a new parameter set for the current population. The mutation operator adds variability to randomly selected model parameter sets by altering some of the values randomly. Several generations are considered in one GA run, until no further improvement (within a certain tolerance) is achieved in the objective functions in successive generations.

Types of GA Operators:

As discussed above, GA operators consist of parameter representation, population initialisation, selection and sampling, crossover and mutation operators. Many methods have evolved for each of these operations.

Parameter Representation:

Parameter representation is the first process in GA operation. This process discretises and translates the searchable range of the parameter to be optimised into a form that is understood by GA. Binary and grey coding systems are the two most commonly used schemes for parameter representation (Mulligan and Brown, 1998; Vasquez *et al*., 2000) and share the common concept where each parameter value encoded into a string of 0 and 1(Wang, 1991; Liong *et al*., 1995). Grey coding can be considered as an extension of the binary representation, which is designed to overcome a problem experienced in binary coding called ‘Hamming Cliffs’

(Haupt and Haupt, 1998). Some studies have used direct real value vector as parameter representation, which uses the actual parameter value itself within the given parameter range (Wardlaw and Sharif, 1999). No discretisation and no encoding or decoding of the parameter space is required for real value coding, as required by binary and grey coding systems. The real value representation is more consistent and precise, and outperforms both binary and grey coding systems in terms of speed and accuracy (Yoon and Shoemaker, 1999).

Population Initialisation:

The population initialisation is the process of generating initial parameter sets for the first population of the GA run. Two methods, namely random and heuristic are used to initiate the population. The commonly used random method generates parameters randomly without any prior knowledge of the likely 'optimum' parameter set. A 'seed' is used to start the random number generator and the generation of the parameters. Therefore, the seed can be a factor in affecting the search of the 'optimum' parameter set (Wang, 1991; Franchini, 1996). The heuristic method, on the other hand, requires some prior knowledge of the likely 'optimum' parameter set, and therefore, it provides the 'optimum' solution faster. However, there is a possibility of producing a local optimum with this method due to limited diversity in parameter sets within the population.

Selection and Sampling Operator:

The selection process determines the number of parameter sets in the current generation that participates in generating new parameter sets for the next generation. The parameter sets with the highest fitness value can be expected to have the greatest number of times participating in generating new parameter sets; these parameter sets are called 'parent copies' in this paper. Three selection methods, namely proportionate selection (Grefenstette, 1997a), linear ranking (Grefenstette, 1997b) and tournament selection (Blickle, 1997) are commonly used to determine the number of 'parent copies'. Goldberg and Deb (1991) stated that no one-selection method is superior to the others. The number of 'parent copies' calculated for each parameter set can be a real number or proportionate and linear ranking selection methods, because the number of parent copies selected is a function of the fitness value for each parameter set. On the other hand, under the tournament selection, the number of parent copies is selected through competition regardless of their fitness value; hence translation from real value 'parent copies' to integer value (i.e. sampling in GA terminology) is not required. Therefore, only proportionate and linear ranking selection methods require of undertaking the sampling operator process. The sampling operator translates the number of 'parent copies' from real number to an integer. All methods used for sampling are based on the traditional roulette wheel concept. This concept uses spinning the wheel with slots apportioned according to the real number of 'parent copies' determined from the selection process for each parameter set in the current generation. Five sampling methods have emerged namely, remainder stochastic sampling with replacement, remainder stochastic sampling without replacement, deterministic sampling, remainder stochastic independent sampling and stochastic universal sampling. Of these five methods, stochastic universal sampling is considered as the best sampling method in terms of accuracy, precision and computer time (Baker, 1987).

Crossover and Mutation Operators:

The crossover operator is used to create new parameter sets (i.e. offsprings), by randomly selecting the location of the two parent parameter sets that were selected to participate in the next generation through selection and exchanges parts of chromosome. Three different crossover approaches have been cited in the literature namely, one-point crossover, two-point crossover and uniform crossover (Booker, 1997). Fig. 1 and fig. 2 shows an example each for one-point crossover and uniform crossover methods.

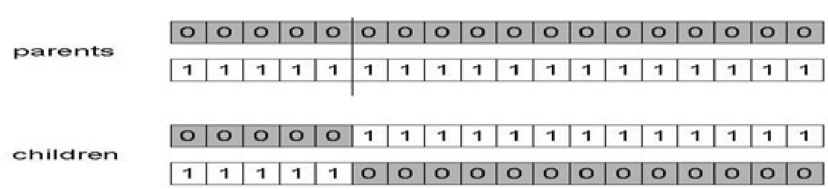

Fig. 1: One-point crossover

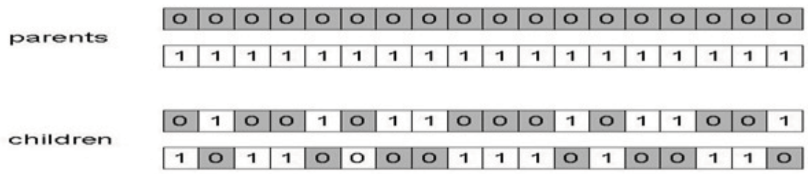

Fig. 2: Uniform crossover

Both one point and two-point crossover operators exchange parts of the chromosome at the randomly selected points of the two parent parameter sets. They differ by the number of cuts made on the parent chromosomes and hence introduce more variety in the exchange process. The uniform crossover operates on randomly selected individual genes of two parent parameter sets and not on parts of the chromosome. The mutation operator is used to add variability to the randomly selected parameter sets, obtained from the above crossover process. For binary and grey coding systems, mutation is done by changing the encoded parameter values (or genes) by flipping 0s to 1s, at one or more randomly chosen location(s) of the encoded chromosome. On the other hand, in real value representation, the values in the randomly selected parameter set are being altered within the feasible parameter range. Two mutation methods are used in real value representation, namely uniform and non-uniform mutations (Michalewicz, 1996). These methods differ from each other by the frequency of mutating the parameters within the generation. In uniform mutation, the number of parameter sets to be mutated is determined by the mutation rate, which is fixed for all generations within one GA run. However, in nonuniform mutation, the mutation rate reduces as the run progresses from one generation to another (Wardlaw and Sharif, 1999).

Applications of Genetic Algorithms:

Genetic algorithms are very powerful search tools. By "search" it is meant that GAs is capable of pouring through a large number of potential solutions to find good solutions. Scheduling has been an area where genetic algorithms have proven very useful. The GA searches the space of potential schedules and finds those schedules which are most effective and maximize the desired criteria (such as minimizing idle time). For example, GAs is used by some airlines to schedule their flights (Dulay, 2005). An application of a GA to a financial problem (tactical asset allocation and international equity strategies) resulted in an 82% improvement in portfolio value over a passive benchmark model, and a 48% improvement over a non- GA model used to improve the passive benchmark (Dulay, 2005). GAs have also been applied to problems such as protein motif discovery through multiple sequence alignment and obtaining neural network topologies (Taylor and Agah, 2006).More information on genetic algorithms can be found in Goldberg (1989).

GA has been applied to many problems in hydrology and water resources. Wang (1991) successfully used GA technique to calibrate the conceptual rainfall-runoff models. Liong *et al*. (1995) used GA for peak flow forecasting in a watershed in Singapore. Cieniawski *et al*. (1995) used GA to solve a multi-objective groundwater monitoring problem. Aly and Peralta (1999) used GA for optimal design of an aquifer cleanup system under uncertainty. Aral *et al*. (2001) used progressive GA for the identification of contaminant source location and recharge history for groundwater management. Jain and Srinivasulu (2002) used real-coded GA for estimating parameters of infiltration equations. Samuel and Jha (2003) employed GA to determine aquifer parameters from pumping test data. More recently, Jain *et al*. (2004) used GA for optimal design of composite irrigation channels; Prasad and Park (2004) used multi-objective GA for the optimal design of water distribution networks and Jain and Srinivasulu (2004) used artificial neural networks to develop improved methodologies for rainfall-runoff modelling.)) GA Genetic Algorithms (GA), also have been applied to numerous engineering problems (Rauch, 1999) such as management of water systems (Cai *et al*., 2001), design of water distribution networks (Savic, 1999), optimisation of sewer networks (Parker, 2000), calibration and improvement of urban drainage systems (Liong 1995, James, 2002). In most applications, single objective GAs were used such that only one criterion of optimisation was evaluated at a time and this is mainly due to the inadequate ability of single objective GA to deal with more than one criterion/objective.

S_en and Öztopal (2001) used GAs to predict precipitation occurrence. Moreover, GAs can be easily integrated to other approaches such as fuzzy logic models and neural networks as a training tool that optimize the parameters of the concerned system. They are also very efficient in determining the antecedent parameters, optimum rule numbers and consequent parameters of a fuzzy system. Combination of GAs and Fuzzy logic (FL) algorithms is known as geno-fuzzy methodology which is employed for the optimization of output parameters in modelling sun shine duration and radiation (S_en *et al*, 2004). GAs are also used in the optimization and operation of groundwater resources design uncertainties in the hydraulic conductivity and they also provide solution of more complex nonlinear problems compared with traditional gradient based approaches (Espinoza *et al*. 2005, Hilton *et al*. 2005, Mahinthakumar *et al*. 2005). In order to take the uncertainties into consideration without any basic assumptions such as linearity, normality it is preferred here to employ GAs for the prediction of sediment load from discharge data.

Altunkaynak (2009) was predicted sediment load from discharge measurements by using Genetic Algorithm. He applied GA methodology for discharge and sediment load data that obtained from Mississippi river at Missouri, St. Louis. Whit compare between GAs with Relating Model, He was found that GAs outperform RM in terms of mean relative error (MRE).

Nasseri *et al* (2008) was used, feed-forward type networks developed to simulate the rainfall field and a so called back propagation (BP) algorithm coupled with genetic algorithm (GA) to train and optimize the networks. The technique was implemented to forecast rainfall for a number of times using rainfall hyetograph of recording rain gauges in the Upper Parramatta catchment in the western suburbs of Sydney, Australia. Results of the study showed the structuring of ANN network with the input parameter selection, when coupled with GA, performed better compared to similar work of using ANN alone.

Cheng *et al* (2002) calibrated the model that consists of two parts: water balance parameter and runoff routing parameter calibration. The former is based on a simple genetic algorithm (GA). The latter is based on a new method which combines a fuzzy optimal model (FOM) with a GA for solving the multiple objective runoff routing parameters calibration problem. Their parameter calibration includes optimization of multiple objectives: (1) peak discharge, (2) peak time and (3) total runoff volume. Results of their study show that the hybrid methodology of GAs and the FOM is not only capable of exploiting more the important characteristics of floods but also efficient and robust.

Tayfur *et al*. (2008) developed a genetic algorithm model to predict flow rates at sites receiving significant lateral inflow. They predicts flow rate at a downstream station from flow stage measured at upstream and downstream stations on Tiber River, Italy. Their study developed a genetic algorithm model to predict flow rates at sites receiving significant lateral inflow. Their results showed that the GA model produced satisfactory results and it was superior over the most recently developed rating curve method.

REFERENCES

Altunkaynak, A., 2009. Sediment load prediction by genetic algorithms. Journal of Advances in Engineering Software, 40: 928-934.

Aly, A.H., R.C. Peralta, 1999. Optimal design of aquifer cleanup systems under uncertainty using a neural network and genetic algorithm. Water Resour. Res., 35(8): 2523-2532.

Aral, M.M., J. Guan, M.L. Maslia, 2001. Identification of contaminant source location and release history in aquifers. J. Hydrol. Eng., ASCE, 6(3): 225-234.

Baker, J., 1987. Reducing bias and inefficiency in the selection in the selection algorithm. Second International Conference on Genetic Algorithms, July 28-31, 1987, Massachusetts Institute of Technology, Cambridge, MA, pp: 14-21.

Blickle, T., 1997. Tournament selection. In: B.ack, T., Fogel, D., Michalewicz, Z. (Eds.), Handbook on Evolutionary Computation (Vol. Part C, Section 2.3). Institute of Physics Publishing, Oxford University Press, New York, pp: 1-4.

Booker, L., 1997. Recombination. In: B.ack, T., Fogel, D., Michalewicz, Z. (Eds.), Handbook on Evolutionary Computation (Vol. Part C, Section 3.3). Institute of Physics Publishing, Oxford University Press, New York, pp: 1-11.

Cai, X., D.C. McKinney, L.S. Lasdon, 2001. Solving nonlinear water management models using a combined genetic algorithm and linear programming approach. Adv Water Resour, 24(6): 667-76.

Chenga, C.T., C.P. Oub, K.W. Chauc, 2002. Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration. Journal of Hydrology, 268: 72-86.

Espinoza, F.P., B.S. Minsker, D.E. Goldberg, 2005. Adaptive hybrid genetic algorithm for groundwater remediation design. J Water Res Plng Mgmt., 131(1): 14-24.

Franchini, M., 1996. Use of a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. Hydrological Sciences, 41(1): 21-39.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimisation and Machine Learning. Addison-Wesley Publishing Company, Massachusetts.

Goldberg, D.E., K. Deb, 1991. A comparative analysis of selection schemes used in genetic algorithms. In: Rawlins, G.J. (Ed.), Foundations of Genetic Algorithms. Morgan Kaufmann, San Mateo, CA, pp: 69-93.

Grefenstette, J., 1997a. Proportional selection and sampling algorithms. In: B.ack, T., Fogel, D., Michalewicz, Z. (Eds.), Handbook of Evolutionary Computation (Vol. Part C, Section 2.2). Institute of Physics Publishing, Oxford University Press, New York, pp: 1-7.

Grefenstette, J., 1997b. Rank-based selection. In: B.ack, T., Fogel, D., Michalewicz, Z. (Eds.), Handbook of Evolutionary Computation (Vol. Part C, Section 2.4). Institute of Physics Publishing, Oxford University Press, New York, pp: 1-6.

Haupt, R., S. Haupt, 1998. Practical Genetic Algorithms. A Wiley Inter-Science Publications, New York.

Hilton, A.B.C., T.B. Culver, 2005. Remediation design under uncertainty using genetic algorithms. J Water Res Plng Mgmt., 131(1): 25-34.

Jain, A., S. Srinivasulu, 2002. Calibration of infiltration parameters using genetic algorithm. Proceedings of HYDRO 2002: Conference On Hydraulics, Water Resources and Ocean Eng., December 16-17, IIT Bombay, Bombay, India.

Jain, A., S. Srinivasulu, 2004. Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms, and artificial neural network techniques. Water Resour. Res., 40(4): W04302.

Liong, S.Y., W.T. Chan, J. Shree Ram, 1995. Peak-flow forecasting with genetic algorithms and SWMM. J Hydr Eng, ASCE, 121(8): 613-7.

Michalewicz, Z., 1996. Genetic Algorithms+Data Structures=Evolution Program. Springer, New York.

Mulligan, A.E., L.C. Brown, 1998. Genetic algorithm for calibrating water quality models. ASCE Journal of Environmental Engineering, 124(3): 202-211.

Nasseri, M., K. Asghari and M.J. Abedini, 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Systems with Applications, 35: 1415-1421.

Parker, M., D.A. Savic, G.A. Walters, Z. Kapelan, 2000. SEWERNET: a genetic algorithm application for optimizing urban drainage systems. In: Int conf on urban drainage via Internet, 2000. Available from: http://www.dhi.cz/ICUDI.

Paz, E., 1998. A survey of parallel genetic algorithms, Internet, http:// www.www.illigal.ge.uiuc.edu.

Prasad, T.D., N.J. Park, 2004. Multi-objective genetic algorithm for design of water distribution networks. J. Water Resour. Plan. Manag., ASCE, 130(1): 73-82.

Rauch, W., P. Harremoes, 1999.On the potential of genetic algorithms in urban drainage modeling. Urban Water, 1: 79-89.

Samuel, M.P., M.K. Jha, 2003. Estimation of aquifer parameters from pumping test data by genetic algorithm optimization technique. J. Irrig. Drain. Eng., ASCE, 129(5): 348-359.

Savic, D., G. Walters, 1999. Genetic algorithms for the least-cost design of water distribution networks. J Water Resour Plann Manage ASCE, 123(2).

S_en, Z., A. Öztopal, 2001. Genetic algorithms for the classification and prediction of precipitation occurrence. Hydrol Sci J., 46(2): 255-68.

S_en, Z., A. Öztopal, A.D. Sahin, 2004. Solar irradiation estimation from sunshine duration by geno-fuzzy partial approach. Energy Sources, 26(4): 377-86.

Tayfur, G., T. Moramarco, 2008. Predicting hourly-based flow discharge hydrographs from level data using genetic algorithms. Journal of Hydrology, 352: 77- 93.

Taylor and A. Agah, 2008. Data Mining and Genetic Algorithms: Finding Hidden Meaning in Biological and Biomedical Data. Computational Intelligence in Biology: Current Trends and Open Problems. Smolinski, Tomasz G., Milanova, Mariofanna M., and Hassanien, Aboul-Ella (Eds.).

Vasquez, J., H. Maier, B. Lence, B. Tolson, R. Foschi, 2000. Achieving water quality system reliability using genetic algorithms. ASCE Journal of Environmental Engineering, 126(10): 954-962.

Wang, Q., 1991. The genetic algorithm and its application to calibrating conceptual rainfall-unoff models. Water Resources Research, 27(9): 2467-2471.

Wardlaw, R., M. Sharif, 1999. Evaluation of genetic algorithms for optimal reservoir system operation. ASCE Journal of Water Resources Planning and Management, 125(1): 25-33.

Yoon, J., C. Shoemaker, 1999. Comparison of optimisation methods for groundwater bioremediation. ASCE Journal of Water Resources Planning and Management, 125(1): 54-63.