Selection of Genetic Algorithm Operators for Urban Drainage Model Parameter Optimisation

N. R. Siriwardene and B. J. C. Perera

School of Architectural, Civil and Mechanical Engineering, Victoria University,
PO Box 14428 MCMC, Melbourne, Victoria 8001, Australia,
E-mail: nilmini.siriwardene@students.vu.edu.au

Abstract : Recently, Genetic Algorithm (GA) has proven to be successful and efficient in identifying the optimal parameters for water resource modelling applications. However, in order to produce efficient and robust solutions, proper selection of GA operators is necessary for the application, before conducting the model parameter optimisation. General guidelines are available for standard GA optimisation applications. However, there is no specific guidance available for selecting GA operators for urban drainage model parameter optimisation. Therefore, the sensitivity of these operators are analysed through numerical experiments by repetitive simulation considering one GA operator at a time, by integrating GA and urban drainage modelling software. It was found that models with a small number of parameters (i.e. two or less) could be optimised with any GA operator set in urban drainage modelling. However, the proper selection of GA operators is vital to the convergence of the optimum model parameters, for large number of parameters (i.e. five or more) in urban drainage modelling.

Keywords: *Genetic Algorithm; Urban Drainage; Parameter Optimisation; Modelling*

1. INTRODUCTION

Management of stormwater runoff from urban catchments has become an increasingly important environmental issue and stormwater drainage is a major part of this overall stormwater management. The development of efficient stormwater drainage systems is still necessary due to continued urban development. The most practical way of designing these systems is by the application of mathematical models, which consider complex hydrological (eg. rainfall, infiltration, overland flow, evaporation) and hydraulic (eg. pipe and open channel flow) processes of urban areas.

The accuracy of these models depends on the correct selection of the model parameter values. Some of these values can be physically measured, but other parameters such as depression storage, flow roughness etc. are difficult to measure. Therefore, these parameters, which are impossible or difficult to measure physically, have to be estimated through model calibration. The model calibration is done through an iterative process by comparing model predictions with observations, until the two sets match with each other within a reasonable accuracy.

There are several methods available to calibrate mathematical models ranging from trial and error to optimisation methods. Traditionally, the model calibration is done through trial and error. With this method, the model parameters are estimated by experienced modellers starting with educated guesses and refining these guesses by comparing observed and modelled hydrographs. However, this method is subjective, time consuming and can also miss the optimum parameter set. On the other hand, computer based automatic optimisation methods have proven to be robust and efficient. In this project, one of the most popular automatic calibration optimisation methods known as genetic algorithm (GA) is used to calibrate the urban drainage models. Even though the GA has been recognized as a robust optimisation method for estimating model parameters in many fields, it has not been used widely for urban drainage models.

GA operators, such as parameter representation, population size, selection methods and crossover and mutation rates play an important role on the convergence of the optimum model parameter set. Davis (1991) reported that the optimum GA operator set varies according to the application. Franchini and Geleati (1997) studied the effects of GA operators in detail in their rainfall-runoff model calibration study and reported that a robust GA operator range was adequate, as it did not have any significant effect on the optimum model parameter set. Wardlaw and Sharif (1999) and Ng (2001) conducted comprehensive GA operator studies in their optimal reservoir system operation and water quality model parameter optimisation studies respectively, and arrived at different optimum GA operators. The above studies show that there are no clear conclusions regarding the optimum GA operators to be used in model parameter optimisation. Therefore, a detailed study was conducted to determine the optimum GA operators before attempting the model parameter optimisation in urban drainage modelling.

2. GENETIC ALGORITHM (GA)

GA is a widely used probabilistic search method originally developed by Holland (1975) and later refined by many others. It is a robust technique and uses a computer based iterative process that employs the mechanics of natural selection and natural genetics to select the optimum parameter set for the given problem.

The genetic algorithm vocabulary is adopted from natural genetics. The model parameter set is defined as a 'chromosome', while each parameter is known as a 'gene'. The representation or coding of chromosomes has a large impact on search performance, as the optimisation is performed on this representation. There are two main types of parameter coding methods available, which are bit string coding and real-value parameter coding. In real-value coding each parameter is represented by its real-value. There are two types of bit string coding methods available, namely binary and gray coding, which use similar concepts. In binary coding, each parameter is encoded into strings with binary digits (i.e. 0 and 1) and arranged linearly to form a chromosome. Gray coding, which is an enhancement of binary coding, increases the search performance as it avoids the 'Hamming Cliffs' problem associated with binary coding (Ng, 2001). The bit string coding is widely used by GA researchers because of its simplicity and also GA theory was initially developed on this basis. In the bit string coding method, the genes present in the chromosome can be represented with varying string lengths (i.e. number of bits) according to their parameter range and required precision of parameters. However, some of the recent research work reported that real-value coding is superior to bit string coding (eg. Wardlaw and Sharif, 1999).

Each GA run consists of a number of successive populations of chromosomes, which are possible solutions (or parameter sets) of the given problem. Population size is the number of chromosomes present in a population. At the start of the GA optimisation, the user has to define the population size and the number of model parameters that need to be optimised and their ranges. The initial population is then generated at random or using a heuristic technique. The latter method is based on prior knowledge of the parameters and hence provides a good initial estimate and rapid convergence. The advantage of the random method is that it prevents premature convergence to an incorrect solution due to insufficient variability in the initial population. A user-defined objective function is used to evaluate each chromosome in the population. These objective functions of the chromosomes indicate the suitability (or fitness) of the parameter set for the given problem. After computing the objective function for each chromosome of the current population, GA operators such as selection, crossover and mutation are used to generate the next population. Several generations are considered in the GA process, until the user defined termination condition is reached.

There are several selection methods available, namely proportionate, ranking and tournament selections, which are described in detail in http://www.geatbx.com/docu/algselct.html. The selection method determines which chromosomes of the current population participate in generating the next population according to their fitness. This process ensures a higher chance of fitter chromosomes passing their genes to the next generation. Once the appropriate chromosomes are selected, they are used to create new chromosomes for the next population by randomly combining two chromosomes, which is called the crossover process. There are several crossover methods available, namely single point, multi point, uniform and shuffle crossover. The "elitist" selection option ensures the fittest chromosome from one population is propagated to the next population without any disturbance. The crossover rate determines the probability that a pair of chromosomes will be subject to the crossover process. A high crossover rate is used to encourage good mixing of the chromosomes.

The mutation operator randomly modifies the chromosomes (eg. in bit string representation '0' to '1' and visa versa) to introduce diversity to the population. A large mutation rate increases the probability of destroying good chromosomes, but will prevent premature convergence.

3. SELECTION OF GA OPERATORS FOR USE IN XP-UDD DRAINAGE MODEL

3.1. XP-UDD Urban Drainage Model

There are several computer software packages available for urban catchment modelling. Most of these software tools use the same equations for modelling hydrological and hydraulic processes of urban catchments. XP-UDD (XP Software 2000, http://www.xpsoftware.com.au /hydraulic.html) is an improved version of SWMM (US Environment Protection Agency, 1992) and its input and output files are in ASCII format, which can be accessed by the GA software or other external software tools. Therefore, XP-UDD was used in this study.

In XP-UDD, the urban catchment is divided into two significant sub-areas namely, impervious and pervious areas. The impervious area includes road surfaces, roofs and other man-made hard surfaces. The pervious area includes bare surfaces, porous pavements, grass courts and lawns. During 'small' storm events, runoff is generally generated only from the impervious area after filling its depression storage, since rain falling on the pervious areas infiltrates into the soil. However, during 'large' storm events, pervious areas contribute to runoff, in addition to impervious areas.

The Horton's infiltration equation was selected in this study to model infiltration in pervious areas, as the parameters of this equation can be determined through soil infiltration tests (but refined during calibration to allow for heterogeneity of the soil). Seven model parameters were identified for calibration of the XP-UDD model, two related to the impervious areas (i.e. percentage of the impervious area -*%A* and the depression storage - DS_i) and the other five related to the pervious areas (i.e. depression storage - DS_p, overland flow roughness of the pervious areas - n_p and the three Horton's soil infiltration parameters - f_o, f_c *and* k).

A typical urban drainage catchment (i.e. the 'Kew' catchment, in Melbourne metropolitan area in Victoria, Australia) and two hypothetical storm events (i.e. 'small' and 'large') were considered in optimising GA operators in this study. The catchment has an area of 18ha. The 'small' storm considered was the design storm, which had an Annual Recurrence Interval (ARI) of 1 year and storm duration of 30 minutes. This storm produced runoff only from the impervious areas and was used to calibrate the two impervious area parameters. The 'large' storm, which had an ARI of 100 years and 30 minutes duration, generated runoff from both impervious and pervious areas and was used to calibrate the remaining five pervious area parameters after fixing the two impervious area parameters obtained from the 'small' event. Although it is beneficial to consider several storms, it is not feasible to perform an analysis such as this, because of repetitive GA runs and associated high computer time. Typical model parameter values (as given in Section 3.3.e) were assumed to generate the hydrographs due to these two storm events. These hydrographs were considered as the observed hydrographs in optimising GA operators and its assumed parameters as the actual parameters.

Several public domain GA source codes have been developed in Fortran, C/C++, Java and other programming languages in the past, and can be found at http://www.aic.nrl.navy.mil/galist/src/#C. GENESIS GA software has been successfully used in the past by many researchers (eg. Liong et al. 1995 and Ng 2001) and therefore was used in this study. A computer program was developed to link the GENESIS and XP-UDD software tools.

3.2. Methodology Used in Optimising GA Operators

Binary and gray coding options are available in GENESIS for parameter representation. Gray coding was used in this study, as it is an extension of binary coding. Furthermore, the two-point crossover type was used, as it was the only crossover method available in GENESIS.

To study the effects of GA operators in the XP-UDD model, population size, selection type and crossover and mutation rates were varied one at a time, keeping all other operators constant. At the start of these studies, the crossover and mutation rates were kept at 0.6 and 0.001 and the proportionate selection method was used, which are the default values in GENESIS. The objective function used in this study was the minimisation of the sum of square of difference between computed and observed hydrograph ordinates as it had been widely used in many previous studies (eg. Liong et al. 1995) and implicitly allows other important features of the hydrographs such as peak, time to peak and volume to be matched. Population size and number of generations are related by the total number of simulations in one GA run, and therefore these two were studied together.

a) Population Size

Based on the previous work of Franchini and Geleati (1997), population sizes of 75, 100, 125 and 200 were initially investigated for both impervious and pervious area studies with 7500 simulations. Based on these results further investigations were done for population sizes of 10, 25, 50 and 50, 150, 300, 500 for impervious area and pervious area parameter studies respectively. The optimum population size and the number of generations were then selected from these investigations by repetitive GA runs.

b) Number of Optimum Parameter Sets to be Considered From The Final Generation

In a typical GA run, there could be several equally good parameter sets giving the best objective function in the final generation. The objective functions of these sets may differ only by a small margin, though there could be significant differences in their parameters. Therefore, it is not appropriate to select a single parameter set from the final generation. Franchini and Galeati (1997) determined the mean value of the best 20-parameter sets based on objective functions in their rainfall runoff model. Wang (1991) and Liong et al. (1995) selected a single parameter set and Ng (2001) selected the mean value of the best 10 parameter sets based on objective function in their applications. Therefore, the results obtained from the best GA runs were analyzed, to determine how many parameter sets need to be considered from the final generation to determine the optimum parameter set.

c) String Length

In bit string representation, string length of the gene is estimated based on parameter range and required precision of coding, as stated in Section 2. Several population sizes were considered with three different parameter ranges, to investigate the impact of parameter range on parameter convergence, keeping the precision of the parameter values at the required level. These population sizes were 25, 50, 75, 100 and 100, 125, 150 for impervious and pervious area parameter studies, respectively.

d) Selection Type

The proportionate selection and the linear ranking selection method options are only available in GENESIS. Therefore, the effects of these two methods on the convergence to the optimum model parameter set were investigated. Each selection method was studied for crossover rates 0.6 and 0.9 with mutation rates 0.001 and 0.01. These figures were the boundaries of the optimum GA operator ranges defined in the literature.

e) Crossover and Mutation Rate

In this study, the effects of crossover rate (XOR) were first investigated for impervious and pervious area parameters studies. For both studies, the crossover rates ranged from 0.1 to 1, with steps of 0.1 (i.e. 10 crossover rates) were initially investigated, keeping mutation rate at 0.001. Then these results were analysed to produce a narrow range of crossover rates. This narrow range was then used with different mutation rates (MR), to produce suitable crossover and mutation rates for the urban drainage model calibration. In this study, crossover and mutation rates were studied together, as they together determined the convergence.

3.3. Results

(a) Population Size

Figure 1 shows the results obtained for optimising two impervious area parameters with population sizes of 25, 50, 75 and. Population size 10 has not converged to the actual model parameters at all, since it has not enough variation in parameters present. As can be seen from Figure 1, all the parameter sets converged very quickly with a population size of 25 within 1125 simulations (45 generations). However, other population sizes were not able to give similar results with the same number of generations.

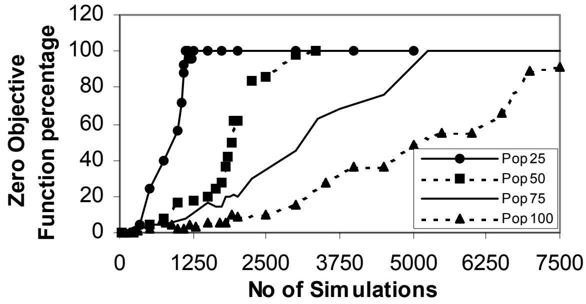

Figure 1. Number of simulations vs. Number of zero objective functions as a % of population (for impervious area parameters)

Figure 2 was produced to illustrate the results of minimum, mean of minimum 5 and mean of minimum 10 objective functions in the final generation, for the pervious area parameter study.

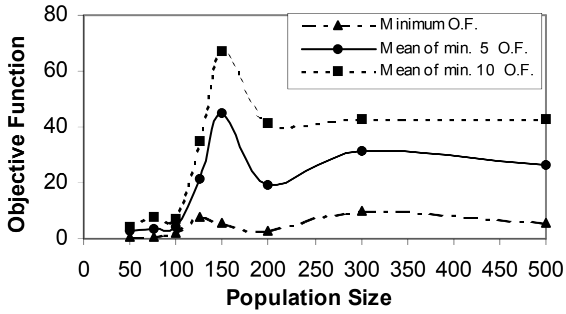

Figure 2. Objective Function value $(liters/sec)^2$ vs. Population size (for pervious area parameters)

Although it can be seen from Figure 2 that the population sizes of 50, 75 and 100 were equally good in terms of objective function, only the population size of 100 converged all five model parameters accurately. Based on the above results, the population sizes of 25 (with 1200 simulations) and 100 (with 7500 simulations) were identified as the optimum population sizes (and number of simulations) for optimising impervious and pervious area parameters respectively in this study. These results were used in the following sections, except in (c).

b) Number of Optimum Parameter Sets to be Considered From The Final Generation

As stated earlier, all parameter sets in the impervious area parameter study reached the actual values in the final generation and therefore need not to be studied. The remaining five pervious area parameters were studied and plots were made of these parameters against the number of parameter sets taken from the best results found, which was the final generation of the population size of 100. Figure 3 illustrates an example of the results obtained for f_c (whose actual parameter was 10 mm/h).

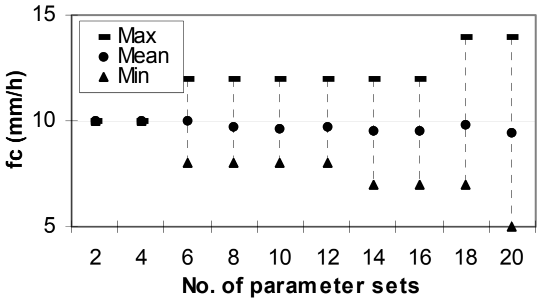

Figure 3. Number of parameter sets vs**.** saturated soil infiltration rate (i.e. f_c)

These plots illustrate the minimum, mean and maximum parameter values against the number of parameter sets (i.e. based on the minimum objective functions) taken from the final generation. It was found that number of parameter sets beyond six deviated significantly from the actual parameter values. Therefore, the mean of the best five parameter sets based on objective function value from the final generation was considered as the value of the optimum parameter set for the pervious area study.

c) String Length

Three different string lengths (which implicitly account for the parameter range and precision) were assumed for both studies to study the effect of the string length on algorithm convergence. The computed string lengths (SL) of the chromosomes were 10, 16, 20 and 38, 48, 58 in impervious and pervious area parameter studies respectively.

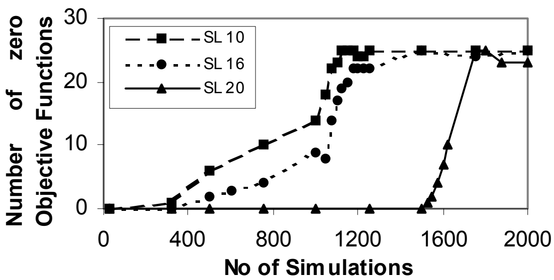

Figure 4. Number of simulations vs. Number of zero objective functions for population size of 25

Figure 4 shows an example of the results obtained for a population size of 25 in the impervious area study. (Similar results were obtained for other population sizes in both studies.) It can be seen from Figure 4 that the longer the string length (i.e. larger parameter range), the longer time was required for convergence, considering the same precisions in parameter values.

d) Selection Type

It was observed that the populations of both impervious and pervious area studies produced minimum objective function values slightly faster when the linear ranking method was used than when the proportionate selection method was used. However, the pervious area model parameters did not converge to the actual values with the linear ranking method. Therefore, the proportionate selection method was used for the rest of the study.

e) Crossover and Mutation Rates

In the impervious area study, 25 chromosomes of the final population were able to converge to zero objective function (i.e. calibrated parameter set reached exactly as the actual or assumed parameter set) in all ten runs (crossover 0.1 to 1). However, in the pervious area study, the crossover rate 0.3 and between 0.5 to 1 only gave the minimum objective function values. When the calibrated model parameters obtained from these GA runs and the actual values were compared, it was found only the actual values corresponding to the crossover rates from 0.6 - 0.9 were matched closely with each other. Therefore, the conclusion was reached that the crossover rate ranging from 0.6 – 0.9 needs to be considered for further study with mutation rates varying from 0.001 to 0.1. The results obtained for five model parameters (based on mean of 5 minimum objective functions) are shown in Table 1. The actual values used for n_p, DS_P, f_o, f_c and k are 0.03, 3mm, 100mm/h, 10mm/h and 0.001 1/sec respectively.

It can be seen from the Table 1 that the crossover rate of 0.6 with 0.001 mutation rate gave the best result based on a comparison of GA-optimised and actual parameter values for this application. These are the default values of GENESIS and therefore they are recommended for use in XP-UDD model parameter calibration. The other acceptable crossover and mutation rates (i.e. based on comparison of GA-optimised and actual parameter values) are shown in bold type in Table 1.

Table 1. Pervious area model parameter values for different crossover and mutation rates

XOR	MR	*np*	*DSp*	*fo*	*fc*	*k*
0.6	**0.001**	**0.029**	**3.1**	**99.4**	**10.0**	**0.001**
	0.002	0.034	2.68	107	11.8	0.0012
	0.004	0.026	2.76	112	13.6	0.0013
	0.006	**0.028**	**3.18**	**99.4**	**13.2**	**0.0011**
	0.008	**0.030**	**3.02**	**104**	**13.4**	**0.0012**
	0.01	0.032	2.72	107	11.6	0.001
	0.05	0.023	2.7	109	12.6	0.001
	0.1	0.039	2.52	106	10.2	0.0011
0.7	0.001	0.029	2.2	120	14.0	0.0014
	0.002	0.024	3.1	111	12.8	0.0013
	0.004	0.060	2.42	107	11.8	0.0012
	0.007	0.034	2.32	115	14.0	0.0013
	0.008	0.029	2.84	110	13.2	0.0013
	0.01	**0.029**	**3.02**	**102**	**13.2**	**0.0012**
	0.05	0.039	2.52	113	9.6	0.0012
	0.1	**0.034**	**2.98**	**95.8**	**11.2**	**0.001**
0.8	**0.001**	**0.031**	**3.12**	**97.8**	**10.6**	**0.001**
	0.002	**0.028**	**3**	**102**	**12.4**	**0.0011**
	0.004	0.034	1.98	121	14.8	0.0014
	0.006	0.027	2.86	107	12.6	0.0012
	0.008	**0.033**	**2.96**	**100**	**12.3**	**0.0011**
	0.01	0.035	2.78	101	13.0	0.0011
	0.05	0.027	2.34	120	13.6	0.001
	0.1	0.050	2.46	107	12.8	0.0013
0.9	**0.001**	**0.028**	**3.02**	**103**	**13.4**	**0.0011**
	0.002	**0.031**	**2.84**	**104**	**12.8**	**0.001**
	0.005	0.038	3.1	98.2	11.8	0.0011
	0.008	**0.030**	**3.26**	**99.4**	**12.4**	**0.0011**
	0.01	**0.030**	**2.96**	**102**	**10.4**	**0.0011**
	0.05	0.033	3.6	92.6	10.2	0.0009
	0.1	**0.031**	**2.92**	**105**	**9.8**	**0.001**

4. SUMMARY AND CONCLUSIONS

Urban drainage models are widely used in urban stormwater planning and management. In order to use these models effectively, it is necessary to calibrate them. Optimisation methods are preferred to the traditional trial and error methods for calibrating such models. Genetic Algorithms (GA) are one of the possible optimisation methods, which is gaining popularity in water resource applications. Before attempting to calibrate urban drainage models using GA, it is necessary to obtain the appropriate GA operators for the study, since there is no guidance available for GA operators to be used in urban drainage modelling.

It was found in this study that GA operators were sensitive to the number of model parameters that needs to be optimised in the application. If the number of parameters to be optimised was small (i.e. 2 or less), GA operators did not play an important role in converging to the optimum model parameter set and therefore general GA operators recommended in literature can be used. Furthermore, small population sizes (i.e. between 25 – 50) are very efficient to use for small number of parameter optimisation in urban drainage modelling. However, for models with large number of parameters (5 or more), GA operators played an important role in converging to the optimum parameter set. In this study, gray coding, population size of 100, proportionate selection, two-point crossover method, elitism, crossover rate of 0.6 and mutation rate of 0.001 gave the best results for the pervious area parameters, and therefore they are recommended for large parameter optimisation in urban drainage models. Furthermore, efficiency of the parameter convergence can be improved by minimising the parameter range and precision of the coding.

5. REFERENCES

Davis, L. Handbook of genetic algorithm, Van Nostrand Reinhold, New York, 1991.

Franchini, M. and Galeati, M., Comparing several genetic algorithm scheme for the calibration of conceptual rainfall-runoff models, *Journal of Hydrological Sciences*, 42(3), 357-379, 1997.

Holland, J. Adaptation in natural and artificial systems, Ann Arbor, The University of Michigan Press, USA, 1975.

Liong, S.Y., Chan, W.T. and Shreeram, J. Peak – flow forecasting with genetic algorithm and SWMM, *Journal of Hydraulic Engineering*, 121(8), 613-617, 1995.

Ng, A.W. Parameter optimisation of river water quality models using genetic algorithms, Ph.D. Thesis, School of the Built Environment, Victoria University, Australia, 2001.

U.S. Environmental Protection Agency (USEPA), Stormwater management model (SWMM), Version 4, User's Manual, Environmental Research Laboratory, Office of Research and Development, Athens, Georgia, 1992.

Wang, Q.J. The Genetic algorithm and its application to calibrating conceptual rainfall-runoff models, *Water Resources Research*, 27(9), 2467-2471, 1991.

Wardlaw, R. and Sharif, M., Evaluation of genetic algorithm for optimum reservoir system operation, Journal of *Water Resources Planning and Management*, 125(1), Jan/Feb, 1999.