# Genetic algorithm applied to the calibration of a rainfall-infiltration-runoff model

R. Chérif[1], J. L. Robert[2] & R. Lagacé[3]
[1]*High Institute of Environmental Sciences and Technology, Tunisia*
[2]*Department of Civil Engineering, Laval University, Canada*
[3]*Department of Soil and Agro-food Engineering, Laval University, Canada*

## Abstract

Standard optimization methods are generally moderately robust and easily diverge due to the form of the objective function. A genetic algorithm (GA) is used in this work to solve the calibration problems of a coupled model rainfall-infiltration-runoff (RIR). The model parameters are those related to Green and Ampt model infiltration in its general form (hydraulic conductivity and *B*). Experiments are done in the laboratory to collect rainfall runoff data, then the infiltration capacities versus infiltration heights are plotted and the scattered points are fitted to a linear regression, in order to deduce the model parameters. Experimental data are used to test the efficiency and the robustness of the genetic algorithm (GA) (using different crossovers). Moreover, the RIR model parameters values estimated by GA optimization are compared to those obtained by experimental fitting strategy. Finally, synthesized data are used to validate the optimization results.
*Keywords: rainfall-runoff, infiltration, hydraulic conductivity, model, optimization, genetic algorithm, crossover.*

## 1 Introduction

Global rainfall-runoff models (RR) commonly use a number of parameters. Most of these parameters are estimated by calibration strategies [1], so the model results will be dependent of its hypothesis, inputs and parameters. These parameters are generally draining coefficients of reservoirs. However, it will be more adequate to estimate some physical parameters from calibration with runoff

 © 2007 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/RM070181

data. In this context, the parameters of the Green and Ampt infiltration model are generally estimated using soil properties, although, they are usually inadequate and parameter values, calibrated from measured runoff data, can be better [2, 3].

The calibration of global RR models uses generally global optimization methods to estimate the model parameters. However, since the objective function to optimize is characterized by the non-smoothness, local optimums, flat plateau, etc., the standard optimization methods can easily diverge due to the objective function shape. In this work, a rainfall-infiltration-runoff (RIR) reservoir model has been developed. This (RIR) model has the particularity of coupling a model infiltration to the subsurface runoff. This RIR model has the particularity of coupling a model infiltration to the subsurface runoff. The model parameters are those related to Green and Ampt model infiltration in its general form. Experimental and synthetically rainfall-runoff data are used to calibrate these parameters.

In a first step, experimental data were fitted to the Green and Ampt model to estimate parameters values. These values are used to validate RIR model calibration results. In this work, standard calibration methods (Newton, Simplex) have diverged in calibrating RIR model when genetic algorithm (GA) had converged on good parameter's values [4, 5].

GA is an optimization research method that combines probability and natural genetic selection [6]. It differs from other classic calibrating methods by using a coded parameter set rather than the parameter values itself, searching among a population of points and it uses only the values of the objective function and not its derivative values [7]. A simple GA is divided in three parts: reproduction, crossover and mutation. The crossover sort may influence the optimization results and the efficiency of the GA [8]. Three crossovers types were used and the best results were retained [9].

Also, this paper investigates to estimate the physical soil parameter values (or their value orders) from runoff data (not expensive) using a simple genetic algorithm optimization.

# 2 Materials and methods

## 2.1 Experimental measurements

The experiments were conducted using the rain simulator at the hydraulic laboratory of the "École de technologie supérieure (ÉTS)" in Montreal. The experimental watershed is represented by a plastic tank filled with an 18 mm/h hydraulic conductivity medium sand. The variable parameters are rain intensities, durations and initial soil moisture values. For every experiment, humidity measurements, soil moisture values at different depths; rainfall masses; total percolation masses and soil moisture values for different time steps are taken. By applying the mass conservation law, rainfall heights and infiltration quantities are deduced (considering null evaporation). Rain intensities range from 10 to 200 mm/h with duration between 15 and 360 min.

### 2.2 Model description

The coupled model RIR is a global model with two reservoirs that uses the Green and Ampt [10] infiltration model to estimate the maximum infiltration capacity. This model is composed of two reservoirs; the first one describes the behavior of the water in the surface and the other represents the soil one. The surface of the watershed receives the rain and splits it in infiltration and runoff. The soil reservoir, setup with an initial water height, takes the infiltrated water until it reaches the saturation and the excess percolates. Similarly in the first reservoir, the water is piled up until it reaches a preset threshold then the exceeding water generates the surface runoff. The detailed model flowchart is described in Chérif [9]. The estimation of infiltration capacity (*f*) of the soil is done by the Green and Ampt [10] infiltration model in its general form (eqn.1). So, the model parameters will be: the hydraulic conductivity *K* and the parameter *B* which is a function of the soil humidity, *K* and the maximum level of water that can be absorbed by the soil.

$$f = K + B/h_i \tag{1}$$

$h_i$ : infiltrated water quantities

### 2.3 RIR model calibration

The model calibration is generally based on the optimization of an objective function that relates the observed and simulated outputs of the model [11, 12]. The objective function used in this work is the mean squared errors of the runoff heights, eqn. (2):

$$fob = \Sigma(Rob(i) - Rsim(i))^2 \tag{2}$$

*fob*: objective function value; *Rob*(*i*): observed runoff height at the step *i*;
*Rsim*(*i*): simulated runoff height at the step *i*.

### 2.4 Fitting of experimental data to the Green–Ampt model

Experimental measures of rainfall-runoff are used to determine the infiltration quantities ($h_{ic}$).

In a first step, an exponential equation is fitted through the infiltration time data, [9]:

$$h_{ic} = \mathrm{a}\ \mathrm{t}^{\mathrm{b}} \tag{3}$$

$h_{ic}$ : cumulated infiltrated water height [L], t: time [T].
where *a* and *b* are deduced from fitting strategy.

In a second step, the infiltration capacities values (*f*) are calculated by deriving in time this equation:

$$f = \partial h_{ic} / \partial t \tag{4}$$

*f* : soil infiltration capacity [L $T^{-1}$], For each experiment, $f$ values are plotted in a graph versus $(1/h_{ic})$ values. These scattered points are finally fitted to a linear regression in order to deduce $K$ and $B$ parameters values for each experiment. These parameters values will be used later to validate the optimization results.

Determination coefficient values (D) obtained vary in the interval [0,92; 0,99] showing a good model fitting quality. Table 1 shows the parameters $K$ and $B$ values for all the experiments and the errors calculated in a 5% conveyance interval and the determination coefficients. All hydraulic conductivity values ($K$) varies in the interval [7.51; 16.72] so always inferior to the saturated hydraulic conductivity value ($Ks$ = 18 mm/h).

Table 1: Green and Ampt parameters values and errors obtained by fitting strategy.

| Experiments | $K$ (mm/h) | $B$ (mm²/h) | Error (K) | Error ($B$) | D |
|---|---|---|---|---|---|
| 1 | 8.69 | 689.3 | 4.38 | 53.0 | 0.99 |
| 2 | 16.72 | 253.5 | 2.40 | 25.8 | 0.99 |
| 3 | 10.68 | 102.4 | 0.47 | 3.7 | 0.99 |
| 4 | 12.70 | 159.8 | 0.50 | 7.3 | 0.99 |
| 5 | 14.77 | 69.9 | 0.64 | 8.7 | 0.98 |
| 6 | 9.33 | 97.1 | 1.33 | 19.8 | 0.92 |
| 10 | 4.84 | 57.4 | 1.33 | 19.9 | 0.92 |
| 11 | 11.19 | 21.4 | 0.38 | 3.8 | 0.96 |
| 12 | 15.69 | 31.3 | 0.64 | 4.2 | 0.98 |
| 13 | 10.86 | 25.2 | 0.77 | 5.5 | 0.98 |
| 15 | 14.19 | 34.5 | 0.81 | 11.7 | 0.95 |
| 16 | 9.36 | 73.8 | 2.23 | 12.7 | 0.99 |
| 17 | 11.36 | 27.8 | 0.68 | 4.4 | 0.98 |
| 18 | 11.78 | 55.3 | 0.60 | 4.4 | 0.99 |
| 19 | 7.51 | 80.8 | 1.24 | 7.3 | 0.99 |

### 2.5 Calibration data

In a first step, the calibration of the model parameters is done from the experimental data. The results of the genetic optimization strategy are compared to those obtained by the fitting method of experimental measurements. Then, in a second step, the calibration is made using generated runoff data from different soil parameters, so that, the synthesized runoff data were introduced in the model in order to estimate its parameters ($K$ and $B$). The synthetic runoffs are generated by the RIR model with a rainfall intensity of 40 mm/h.

### 2.6 The steps of calculation of a simple genetic algorithm

A simple genetic algorithm uses, generally, the notion of an adaptation function. This function is, generally, the function to optimize (known as fitness) which is the objective function of the RIR model in this study. The genetic algorithm used in this work, is a simple algorithm that is composed of the following steps:

1. a population of couples of parameters (*K* and *B*) are randomly chosen in the search space (with a size m);
2. the objective functions are calculated for all the population;
3. the normalized geometric ranking strategy is used to select a new generation that probabilistically favors the points which have the minimum objective function;
4. some crossovers are done between the members of the generation;
5. a random mutation is done in the new generations;
6. the steps 2, 3, 4, 5 are repeated until the maximum number of iterations is attempt.

The selection method used in this work is the normalized geometric ranking method, [13], this method only requires the evaluation function to map the solutions to a partially ordered set. Three popular types of crossovers: simple, arithmetic, and heuristic are used in this work.

# 3 Optimization results and discussion

## 3.1 Analysis of the influence of the crossover on the parameters *K* and *B* optimized and robustness study

Solomatine [8] defines two main performance indicators: the efficiency measured by the number of evaluations needed and how close the algorithm gets to the global optimum and the robustness measured by the number of successes in finding the global minimum, or at least approaching it sufficiently closely. Figure 1 holds the number of iterations required for each crossover type to converge (for all experiments). We note that for the majority of experiments the crossover type does not affect the GA efficiency, fig. 1.

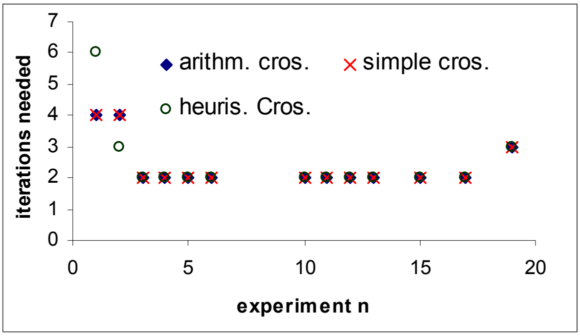


Figure 1: Iterations number needed for different crossovers.

Table 2 shows the optimized *K* values from experiments data using three crossovers sorts. These values are calculated with a 5% confidence level and results are illustrated in table 3. This table shows that the optimized *K* values are independent from the crossover type if we consider the confidence interval.

If compared to the $K$ values obtained from fitting procedure, these optimized $K$ are considered well found, considering the confidence interval, for the experiments (2, 3,10, 12, 13,17 and 18); as for experiments number 4, 5, 6, 15, and 16 they are estimated with a maximum difference equal to 15% of the fitted $K$ value that we consider acceptable. As for experiments (1 and 19), $K$ values are estimated with a difference (from fitted values) of 40% so not well found. So 60% of experiments gave us a good estimation for $K$.

As for, $B$ values, they are estimated in a large confidence interval. Taking these confidence intervals, the $B$ values obtained from different crossover types can be considered equal. However, these optimized values are different from the $B$ values obtained from fitting strategy: these results could be explained by the objective function shape and interactions between $K$ and $B$ parameters. Finally, the GA robustness and results are independent from the crossovers sorts in this case.

Table 2: Optimized $K$ and $B$ values obtained from the three crossovers sorts (experimental data).

| | simple Cr. | | arithmetic Cr. | | heuristic Cr. | |
|---|---|---|---|---|---|---|
| Exp, | $K$ optimized (mm/h) | $B$ optimized (mm$^2$/h) | $K$ optimized (mm/h) | $B$ optimized (mm$^2$/h) | $K$ optimized (mm/h) | $B$ optimized (mm$^2$/h) |
| 1 | 18.0 ± 0.02 | 998 ± 1.1 | 17.9 ± 0.11 | 980 ± 11.8 | 18.0 ± 0.04 | 997 ± 1.6 |
| 2 | 18.0 ± 0.02 | 999 ± 0.8 | 17.9± 0.08 | 983 ±10.3 | 18.0 ± 0.01 | 997 ± 1.0 |
| 3 | 10.0 ± 0.21 | 37 ± 23.6 | 9.9 ± 0.15 | 54 ± 17.7 | 9.8 ± 0.23 | 67 ± 41.5 |
| 4 | 14.4 ± 0.33 | 914 ± 41.0 | 14.2 ± 0.32 | 939 ± 39.4 | 14.0 ± 0.23 | 965± 28.7 |
| 5 | 11.8 ± 0.28 | 921 ± 33.3 | 11.9 ± 0.49 | 912 ± 57.9 | 11.9 ± 0.49 | 912± 57.9 |
| 6 | 6.6 ± 0.12 | 973 ± 16.3 | 6.7 ± 0.21 | 963 ± 25.8 | 6.5 ± 0.14 | 991± 18.4 |
| 10 | 4.4 ± 0.24 | 37 ± 27.6 | 4.3 ± 0.36 | 53 ± 40.4 | 4.3 ± 0.26 | 53 ± 29 |
| 11 | 5.5 ± 0.87 | 831 ± 104 | 5.5 ± 0.42 | 830± 50 | 5.5 ± 0.69 | 832 ± 82.8 |
| 12 | 16.0 ± 0.37 | 68 ± 42.5 | 15.9 ± 0.29 | 81 ± 33.8 | 16.2 ± 0.23 | 42 ± 26.8 |
| 13 | 10.2 ± 0.39 | 49 ± 45.6 | 10.1 ± 0.26 | 57 ± 29.9 | 10.2 ± 0.21 | 42 ± 24.5 |
| 15 | 12.4 ± 0.29 | 952 ± 34.3 | 12.5± 0.44 | 938± 49.8 | 12.4 ± 0.21 | 957 ± 24.8 |
| 16 | 7.7 ± 0.14 | 25 ± 15.6 | 7.4± 0.38 | 54± 43.1 | 7.5 ± 0.28 | 50 ± 30.8 |
| 17 | 12.5± 0.84 | 155 ± 97.2 | 12.8 ± 0.39 | 115± 45 | 12.9 ± 0.55 | 102 ± 62.4 |
| 18 | 12.2 ± 0.13 | 32 ± 15.2 | 11.9 ± 0.35 | 62 ± 39.7 | 12.1 ± 0.31 | 47 ± 34.7 |
| 19 | 18 ± 0.03 | 998 ± 1.9 | 17.9 ± 0.1 | 988 ± 5.9 | 18 ± 0.01 | 998 ± 2.6 |

### 3.2 Optimization results from synthetic data

In a second step, the model calibration is done using rainfall data generated for different cases. To generate this data, a rainfall of a mean intensity equal to 40 mm/h is introduced in the conceptual rainfall-infiltration-runoff model, and $K$ and $B$ parameters values (table 3) are introduced to generate the runoff synthetic data. Later, this synthetic runoff data are used to calibrate the conceptual rainfall-infiltration-runoff model using the GA optimization method.

$K$ optimized values has converged to the good solution (introduced value), for 80% of cases, with 20% of a maximum related difference, fig. 2. As for $B$ values, 60% of cases have converged with related differences less than 30%, 40% of cases has converged with high related differences (can be explained interactions and high correlation between *K* and *B*). Globally, GA can be considered efficient.

- Related difference on *K* is defined as : abs $[(K-K_{optimised})/K]$
- Related difference on *B* is defined as : abs $[(B-B_{optimised})/B]$

Table 3: Different cases of synthesized values.

| | introduced values | |
|---|---|---|
| Case | *K* (mm/h) | *B* ($mm^2$/h) |
| 1 | 20 | 910 |
| 2 | 10 | 375 |
| 3 | 1 | 80 |
| 4 | 2 | 360 |
| 5 | 20 | 800 |
| 6 | 6 | 480 |
| 7 | 5 | 875 |
| 8 | 1 | 120 |
| 9 | 0,5 | 120 |
| 10 | 1 | 50 |
| 11 | 20 | 120 |

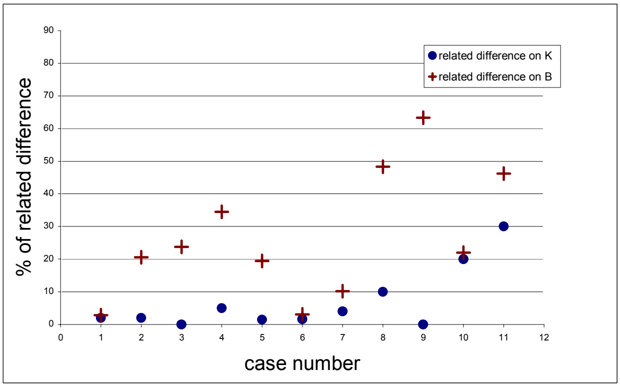


Figure 2: Related differences on *K* and *B*, synthesized optimization (arithmetic crossover).

## 4 Conclusion

A genetic algorithm optimization strategy is used to estimate the physical parameters of a global RIR model through a calibration strategy. The model calibration is done from experimental and synthesized data in order to estimate its physical parameters [*K* and *B*] (of general Green and Ampt infiltration model). The experimental data were fitted by deriving the capacity infiltration expression (exponential equation is fitted through the infiltration time data) in order to deduce [*K* and *B*] values. GA optimization results are compared to the experimental results. It is deduced that the genetic algorithm is effective and the soil parameters [*K*, *B*] are well found by the calibration strategy from synthesized data for the majority of the cases (90% for *K* and 60% for *B*). Finally, the genetic algorithm has successfully solved the optimization difficulties due to the objective function shape where other classic methods have failed.

## References

[1] Wang Q. J., The Genetic Algorithm and its Application to Calibrating Conceptual Rainfall-Runoff Models. Water Resources Research, 27(9), pp. 2467-2471, 1991.
[2] Risse, L. M., Nearing, M. A., Savabi, M. R. Determining the Green–Ampt effective hydraulic conductivity from rainfall-runoff data for the WEPP model. Transactions of the ASAE, 37(2), 411-418, 1994.
[3] Risse, L. M., Savabi, M. R., Nearing, M. A., An evaluation of Hydraulic Conductivity prediction routines for WEPP using natural runoff plot data, Am. Soc. Ag. Eng, ASAE *Summer Meeting*, Charlotte, St Joseph: Mich., Paper 92-2142, 1992.
[4] Chérif, R., Robert, J. L., Lagacé, R., Genetic algorithm calibration of hydrologic model. Proc of 5th Int conférence on Hydroinformatics, IWA Publishing: Cardiff, UK, pp:1417-1422, 2002.
[5] Chérif, R., Robert, J. L., Lagacé, R., Optimisation des paramètres Green et Ampt pour un modèle conceptuel Pluie-infiltration-ruissellement. Canadian Biosystems Engineering, 46, pp. 1.7-1.14, 2004
[6] Holland, J. H., Adaptation in Natural and Artificial Systems, pp. 183, University Michigan Press: Ann Arbor, 1975.
[7] Goldberg, D. E., Algorithmes génétiques: exploration, optimisation et apprentissage automatique, Addison-Wesley: France, 1994
[8] Solomatine D. P., Genetic and other global optimization algorithms – Comparison and use in calibration problems. Proc of 3rd Hydroinformatics, eds Babovic and Larsen, Balkema: Rotterdam, pp.1021-1028, 1998.
[9] Chérif, R., Application des algorithmes génétiques pour l'optimisation des paramètres physiques d'un modèle couplé d'infiltration ruissellement, Ph.D. Thesis, Université Laval, Québec, Canada, 2003.
[10] Green, W. H and Ampt, C. A., Studies on soil physics, I flow of air and water through soils. J. Agr. Sci., 4, pp. 1-24, 1911.
[11] Sorooshian, S., Arfi, F., Response Surface Parameter Sensitivity Analysis Methods for Post Calibration Studies, water resources research, 18 (5), pp. 1531-1538, 1982.

[12] Vicky, L. Freedman, Vicente, L. Lopez, Mariano, Hernandez. Parameter Identifiability for Catchment Scale-Erosion Modelling: A comparison of optimization algorithms. Journal of hydrology, pp.83-97, 1998.
[13] Joines, I. and Houk, C., On the use of non-stationary penalty functions to solve constrained optimization problems with genetic algorithms. *Proc. of the 1st Int. Conf. Evolutionary Computation IEEE*, Orlando: FL, USA, pp. 579-584, 1994.