

# Selection of Genetic Algorithm Operators and Application for Calibration of Dissolved Oxygen Simulation in the River Water Quality Model

Tuan, Nguyen Van

Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University

Mori, Ken

Faculty of Agriculture, Kyushu University

Hirai, Yasumaru

Faculty of Agriculture, Kyushu University

<https://doi.org/10.5109/9334>

---

出版情報：九州大学大学院農学研究院紀要. 52 (2), pp.439-444, 2007-10-29. Faculty of Agriculture, Kyushu University

バージョン：

権利関係：



## Selection of Genetic Algorithm Operators and Application for Calibration of Dissolved Oxygen Simulation in the River Water Quality Model

Nguyen Van TUAN<sup>1</sup>, Ken MORI\* and Yasumaru HIRAI

Laboratory of Bioproduction and Environment Information Sciences, Division of Bioproduction and Environment Information Sciences, Department of Bioproduction and Environmental Science, Faculty of Agriculture, Kyushu University, Fukuoka 812–8581, Japan  
(Received June 27, 2007 and accepted July 17, 2007)

In the overall model development, model calibration is one of the most important tasks. This paper focuses on use of genetic algorithm (GA) in water quality model calibrations and definition of suitable operators which allow GA process run well and generates sets of optimized model parameters for river water quality model. An investigation on the importance of GA operators on model parameter optimization was considered in this study based on the data which collected from the Tatara River, Fukuoka, Japan. Many of numerical experiments on GA operators were conducted to check impacts of different GA operators on efficiency of GA process and goodness of parameter set which will be used for calibrations of water quality model. It was found that most of GA operators have sensitive effects on optimization process, in which the most important operators are rate of crossover and mutation.

Keywords: Genetic algorithm; Optimization; Water quality; Model Calibration

### INTRODUCTION

Successful management of river water quality requires the development and use of river water quality models. Model calibration (or parameter estimation) is an important part of overall model development. Some model parameters can be physically measured, while the remaining model parameters should be estimated through model calibration. Model calibration is generally done through a trial and error iterative process by comparing model predictions with observations. This method is time consuming, and, may be, can miss the 'optimum' model parameter set. Recently, genetic algorithm optimization has proven to be successful and efficient in identifying the 'optimum' parameter set for river water quality models. Genetic Algorithm (GA) is a global optimization technique that is based on the concept of natural selection and genetics [Goldberg, 1989].

In general, the efficiency of GA depends on the proper selection of GA operators, which are essentially the components that make up the overall GA process. The GA operators consist of parameter coding, population initialization, selection of subsequent populations, crossover and mutation.

Although the importance of GA operators on model parameter optimization has been studied in the past, the findings were inconclusive and no guidelines were available to select appropriate GA operators for specific applications such as optimization of river water quality model parameters. These operators can also vary depending on type of water quality and typical condi-

tions of the data set that the GA process works on. Therefore, a specific study applied for the Tatara River was conducted to test the significance of GA operators on the Tatara River water quality model parameter optimization and presented in this paper. In this study, three main operators including size of population, rate of crossover and rate of mutation are considered to point out how they work and the effects of these operators in an attempt to obtain optimum GA operators for river water quality modeling.

### MATERIALS AND METHODS

#### Genetic algorithm

GA is a prominent and powerful optimization technique that has been applied successfully in many disciplines. 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 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 offspring. The selection and sampling operation favors

<sup>1</sup> Laboratory of Bioproduction and Environmental Information Sciences, Division of Bioproduction and Environment Information Sciences, Department of Bioproduction and Environmental Sciences, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University

\* Corresponding author (E-mail: moriken@bpes.kyushu-u.ac.jp)

those parent parameter sets with high fitness value to those of lower fitness value in producing offspring. 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 the convergence of data is achieved.

In this study, a GA program was developed and applied. This program based on main principles of GA and using the reciprocal of the root mean square error of the difference between the model predictions and the observed data as fitness function.

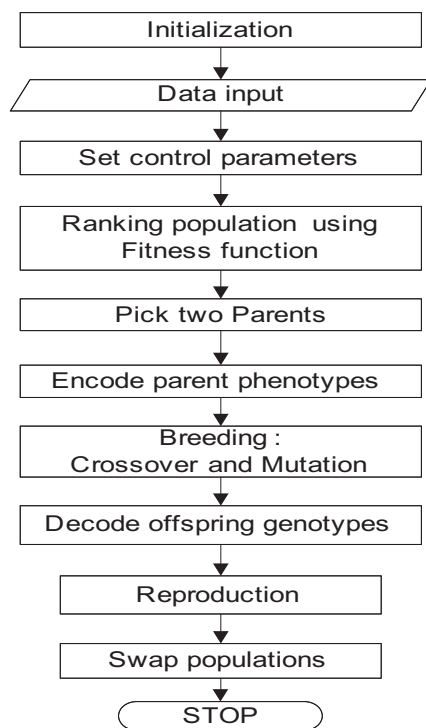


Fig. 1. Flow chart illustration of GA performance.

### The Tatara River

The Tatara River basin is located in Fukuoka prefecture, Japan. The main stem of this river is roughly 21.5 km in length and flow generally from east to west, starting in the mountainous area of Sasaguri Town and Hisayama Town, and finally discharging into the Hakata Bay –Sea of Genkai. The river has an average width of about 3 m at upstream part and about 30 m at downstream. The total catchments area is 101.98 km<sup>2</sup>.

There are total 16 weirs numbered from W1 to W16 (see Fig. 2) across the river. 19 intakes used for purposes of irrigation supply and domestic uses along the river. Water flows into the river through 27 points received waste water from residential area, agricultural drainage, rain water, etc. For the purpose of this study, the Tatara River is divided into 15 river blocks which numbered from B1 to B15 (some examples of river

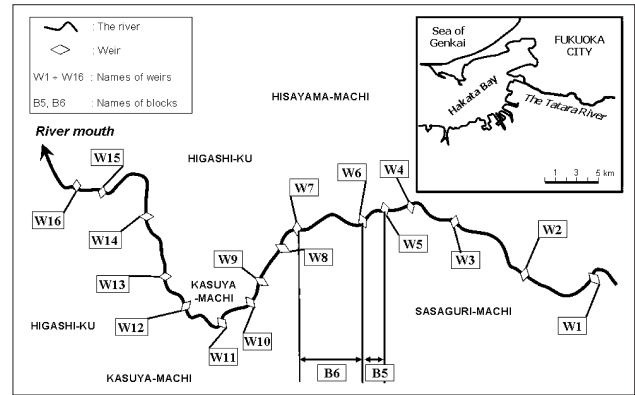


Fig. 2. The Tatara River and the selected river blocks.

blocks are illustrated in Fig. 2); each block is formed by a river segment between two weirs.

### Objectives of the study

The main aims of this study are to test the significance of GA operators on river water quality model parameter optimization and to provide insight into the selection of appropriate GA operators for such optimization studies.

The objectives of study are two river blocks named B5 and B6 chosen from the Tatara River (Fig. 2). Parameter set need to be optimized here are some parameters that could not be measured from field measurement, experiments or analyses in equation of rate of change of dissolved oxygen (DO) concentration.

In equation (1), which used for description of the rate of change of DO, each term represents a major source or sink of DO in the water quality model.

$$\frac{dO}{dt} = K_1(O_s - O) + (\mu - r_2)A - K_3L - \frac{K_3}{d} - r_3N_1 - r_4N_2 \quad (1)$$

where  $O$  = the concentration of dissolved oxygen (mg/l);  $O_s$  = the saturation concentration of dissolved oxygen at the local temperature and pressure (mg/l);  $K_1$  = the reaeration rate in accordance with the Fickian diffusion analogy (1/day);  $K_2$  = carbonaceous BOD de-oxygenation rate (1/day);  $K_3$  = sediment oxygen demand rate (g/m<sup>2</sup>-day);  $r_1$  = the rate of oxygen production per unit of algal photosynthesis (mg-O/mg-A);  $r_2$  = the rate of oxygen uptake per unit of algal respired (mg-O/mg-A);  $r_3$  = the rate of oxygen uptake per unit of ammonia nitrogen oxidation (mg-O/mg-N);  $r_4$  = the rate of oxygen uptake per unit of nitrite nitrogen oxidation (mg-O/mg-N);  $\mu$  = algal growth rate (1/day);  $r$  = algal respiration rate (1/day);  $r_1$  = ammonia oxidation rate coefficient (1/day);  $r_2$  = nitrite oxidation rate coefficient (1/day);  $A$  = algal biomass concentration (mg/l);  $L$  = concentration of ultimate carbonaceous BOD (mg/l);  $d$  = mean stream depth (m);  $N_1$  = ammonia nitrogen concentration (mg/l);  $N_2$  = nitrite nitrogen concentration (mg/l).

In this equation, a set of 6 parameters related to reaeration ( $K_1$ ), algal activities ( $r_1, r_2, \mu$ ), ammonia

nitrogen oxidation ( $\text{NO}_3^-$ ,  $\text{NH}_4^+$ ) will be optimized by GA program. Several model runs will be done to find out the best values of GA operators which can be used to generate an optimal data set of above 6 parameters.

### GA operator selection

#### Search space

Table 1 show rates of parameters will be used as search space in GA program.

**Table 1.** Search space used for GA optimization

| Parameter | Search Space | Actual rate |
|-----------|--------------|-------------|
| $\mu$     | 1.710–1.820  | 1.790       |
| $K_1$     | 1.590–1.630  | 1.615       |
| $\mu$     | 3.800–4.100  | 4.000       |
| $\mu$     | 0.930–1.040  | 1.000       |
| $K_1$     | 0.000–1.100  | 0.500       |
| $\mu$     | 0.970–1.005  | 1.000       |

#### Fitness functions

Fitness function is a function that will be used to calculate the fitness of parameter set. In this study, two types of functions, called simple Least Square and Root Mean Squared Error (RMSE), were considered.

Firstly, a simple least square function (2) was used. After some test-runs conducted, the results shown that with a large set of parameters (6 parameters, in this case), use of least square function is not suitable

$$F = \frac{1}{\sum_{i=1}^n (O_i - P_i)^2} \quad (2)$$

where  $F$  = value of fitness,  $O_i$  = observed value,  $P_i$  = predict value,  $n$  = number of pairs of predicted and observed values.

Secondly, the fitness was determined as the reciprocal of the root mean squared error of the difference between the model predictions and the observed data for water quality constituent [G. J. Pelletier *et al.*, 2006] as shown in equation (3).

$$F = \left( \sum_{i=1}^n w_i \right) \left( \frac{1}{\sum_{i=1}^n w_i} \left( \frac{\sum_{j=1}^m \frac{O_{i,j}}{m}}{\left( \frac{\sum_{j=1}^m (P_{i,j} - O_{i,j})^2}{m} \right)^{1/2}} \right) \right) \quad (3)$$

where:  $F$  = value of fitness,  $O_{i,j}$  = observed value,  $P_{i,j}$  = predicted value,  $n$  = number of different state variables,  $m$  = number of pairs of predicted and observed values,  $w_i$  = weighting factor, and  $q$  = number of different state variables included in the reciprocal of the weighted normalized RMSE.

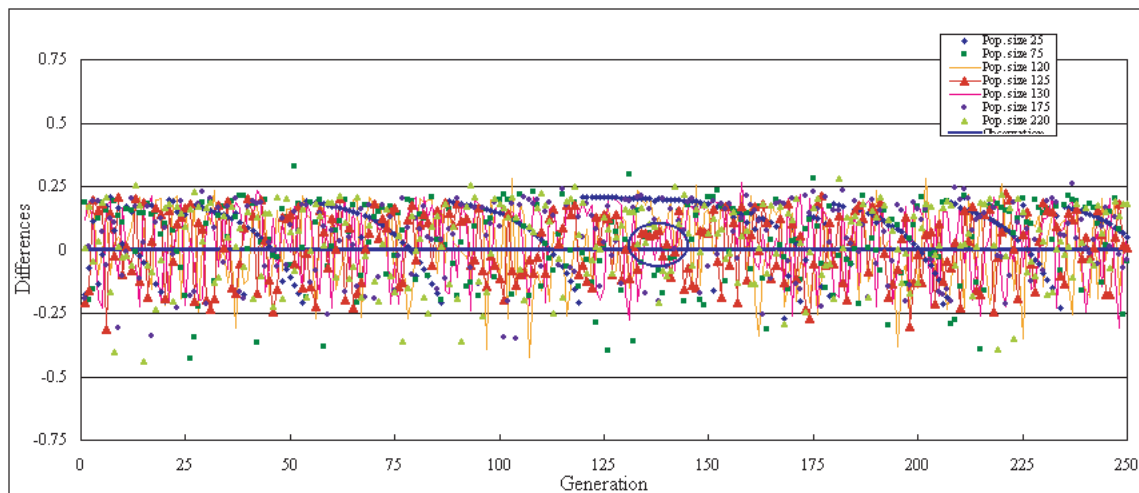
## RESULTS AND DISCUSSION

### Selection of GA operators

#### Population size and number of generations

The first GA operator considered in GA process was the population size, which may have some influence on the 'optimum' model parameter set. Forty different population sizes from 25 to 220 with increasing interval + 5 were tested. The number of generations used for all these population sizes were set at 250, which is big enough to get convergence of model data and not consume much computer time. It was found that the model parameters converged with the population sizes range between 120 and 130, while they did not converge with other population sizes (see Fig. 3).

Also as data shown in this figure, the best values of number of generation range from 132 to 140 (area marked by the blue ellipse shape). In this step, initial values chosen for rate of crossover and mutation are 0.85 and 0.01, respectively. From the results, the population size of 125 with 135 generations was chosen as the 'optimum' population size and the number of generations, respectively, and was used in subsequent investigations.



**Fig. 3.** Differences from actual value of rate of change of DO.



### Crossover and mutation rates

In this step, the mutation and crossover rates were considered. Since these two rates simultaneously determine the rate of convergence of model parameters, they should be tested together. Several combinations of rates of crossover and mutation were chosen and tested in order to find out the optimal set of GA operators. Varying increments within the range 0.0025–0.03 were

considered for mutation rate with increasing constant interval is 0.0025, while constant interval of 0.05 was considered for crossover rate within the range 0.40–0.90.

The results of the analysis were presented using spectral plots of the model parameters. These plots, as shown in Fig. 4, were produced with respect to mutation and crossover rates for all six parameters mentioned

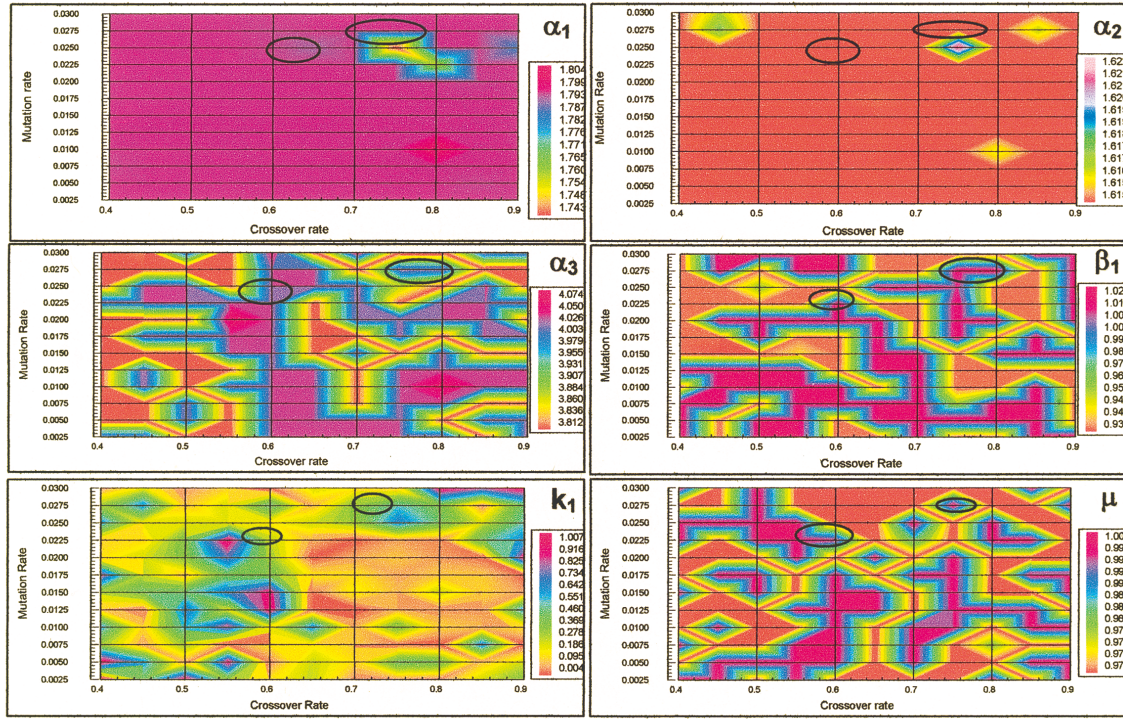


Fig. 4. Mean values of water quality parameters with different combinations of mutation and crossover rates.

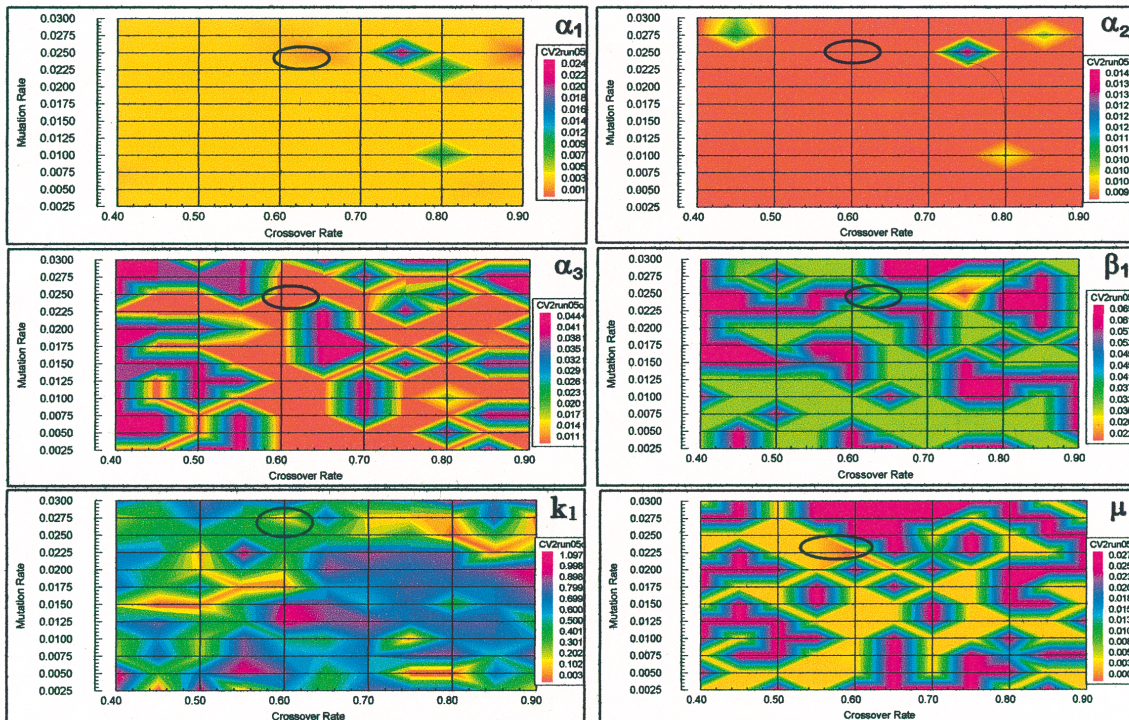


Fig. 5. Coefficient of variation of water quality parameters with different combinations of mutation and crossover rates.



above.

By comparison of these values of six parameters with the actual values as shown in Table 1, some best ranges of result were found as marked by the ellipse shapes in Fig. 4. From this result, the best combination of crossover probability ( $P_c$ ) and mutation probability ( $P_m$ ) were designed in ranges of  $P_c$  from 0.74 to 0.76,  $P_m$  from 0.0272 to 0.0276, and  $P_c$  from 0.58 to 0.62,  $P_m$  from 0.0224 to 0.0227.

In order to find the optimal set of GA operators, one more coefficient has been checked in this study, called coefficient of variation ( $CV$ ).  $CV$  coefficient calculates differences between calculated values of each parameter and the corresponding mean values, respectively.

As the results shown in Fig. 5, most of parameters have convergences with values of  $P_c$  ranged from 0.55 to 0.63 and  $P_m$  ranged from 0.022 to 0.023. Comparing with combinations of  $P_c$  and  $P_m$  found when checking convergences of mean values of water quality parameter (Fig. 4), the best combination of crossover and mutation rates are now defined, with  $P_c$  ranged from 0.58 to 0.62 and  $P_m$  from 0.0224 to 0.0227.

#### The optimal set of GA operators

From the results shown in above sections, the optimal set of GA operators was decided with number of generation is 135, population size of 125, crossover rate is 60% and mutation rate is 2.25%.

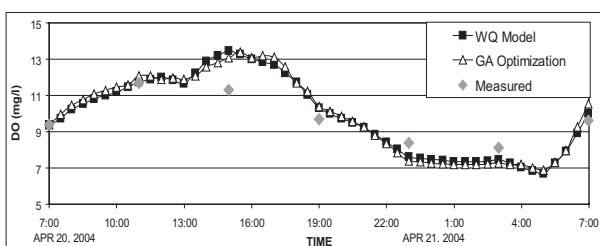
### Calibration of water quality model using GA optimization

Two types of water quality model calibrations for simulation of DO concentration in blocks B5 and B6 of the Tatara River were performed in order to compare results of model calibration using traditional method and optimal parameters sets obtained from GA program.

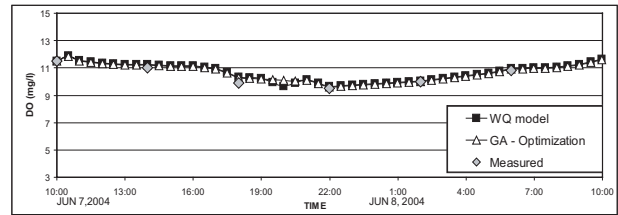
For traditional calibration, here called trial-by-error method, values of 6 parameters were adjusted step-by-step until the model's results have a best fit with observed data set.

For calibration using GA generated optimal data set, a data set, which obtained from GA program based on selected GA operators set, was applied directly into the water quality model.

The results (named WQ model for traditional calibration and GA optimization for calibration using GA) at some calculated point are shown in the pink data series



**Fig. 6.** Concentration of DO at point B5-3, Block B5, the Tatara River calculated by the model using traditional and GA optimization calibrations.



**Fig. 7.** Concentration of DO at point B6-4, Block B6, the Tatara River calculated by the model using traditional and GA optimization calibrations.

in Fig. 6 and Fig. 7. These results shown that simulated data have good fitness with measured data, and data simulated using GA optimization for calibration are also good in comparison to traditional calibration.

### CONCLUSIONS

In this study, a GA program has been developed and applied for calibration of some water quality parameters in the Tatara River. Some main conclusion can be pointed out from this research as following,

- In order to get the best combination of GA operator, not only estimation of the values of parameters, but also coefficient of variation, should be considered.
- Simulated data using GA program for calibration were quite good in comparison to traditional calibration.
- Simulation using GA optimization is a little less correlative with measurement than simulation using traditional calibration. However, it is better in time consumption, and might be better if users have better solutions for GA program.

Because of limitation of data and objectives for application, the results of this study in finding optimal set of GA operators might not be applied well for other cases of study. The main achievements of this study is a quite deeply understanding how is a GA program working, and how to defined an optimal combination of GA operator in order to obtain the best parameters using in calibration of water quality parameters.

In further studies, more cases of study should be conducted, and more references should be considered with expectation of finding out a good solution for application of GA in water quality model calibration.

### REFERENCES

- Brown, L. C. and Barnwell T. O. *The enhanced stream water quality models QUAL2E and QUAL2E-UNCAS: Documentation and user manual*. US Environmental Protection Agency, 1987.
- Charbonneau, P. *An introduction to genetic algorithms for numerical optimization*. NCAR/TN-450 + IA, NCAR Technical Note, National Center for Atmospheric Research Boulder, Colorado, 2002.
- Charbonneau, P. and Knapp B. *A user's guide to Pikaia 1.0*. NCAR/TN-418 + IA, NCAR Technical Note, National Center for Atmospheric Research Boulder, Colorado, 1995.
- Charpa, S. C. *Surface water quality modeling*. McGraw-Hill, 1998.
- Coello, C. A. and Gallegos J. A. F. *Use of genetic algorithms to*

- solve optimal regional water quality management problems*. Proceeding of ACEDC, PEDC, University of Plymouth, UK, 1996.
- Mitchell, A. R. and Griffiths D. F. *The finite difference method in partial differential equations*. Wiley-Interscience Publication, 1985.
- Mulligan, A. E. et al. Genetic Algorithms for Calibrating Water Quality Models. *Journal of Environmental Engineering*, **Vol. 124, No. 3**, ASCE, 1998.
- Nyhoff, L. And Lesstma S. *Fortran 90 for Engineer and Scientists*. Prentice Hall, 1999.
- Nyhoff, L. And Lesstma S. *Introduction to Fortran 90*. Prentice Hall, 1999.
- Orlob, G. T. (ed.). Mathematical modeling of water quality: Stream, Lakes, and Reservoirs. *International Series on Applied Systems Analysis* – **No 12**, 1997.
- Pelletier, G. J. et al. QUAL2Kw – A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration. *Environmental Modeling & Software* **21** (2006) 419–425
- Versteeg, H. K. and Malalasekera W. *An introduction to computational fluid dynamic*. Longman House, Burnt Mill, Harlow, 1995.