Assessing model calibration adequacy via global optimisation

JG Ndiritu[1]* and TM Daniell[2]
[1] *Civil Engineering Department, University of Durban-Westville, Private Bag X54001, Durban 4000, South Africa*
[2] *Civil and Environmental Engineering Department, University of Adelaide, SA 5005, Australia*

Abstract

An assessment of the application of varying levels of optimisation on model simulation performance and parameter identification was done using the genetic algorithm (GA) and a 10-parameter version of the MODHYDROLOG rainfall-runoff model. Four levels of optimisation were obtained through the use of two GA formulations, the traditional and an improved GA, and by varying the optimisation parameters with each formulation. Sixteen years of data from a 27 km^2 Australian catchment was used. With each level, ten randomly initialised optimisation runs were made. The differences in simulation performance quantified by the coefficient of efficiency, bias, absolute deviation and a residual mass curve coefficient were not considerable although the performance improved as the level of optimisation effort increased. Superior parameter identification, and consequently a better detection of parameter correlations was achieved with the higher optimisation levels. Based on the objective function values, the highest level of optimisation practically located the global optimum in all the ten runs. The second level achieved this in nine of the ten runs while the lower two levels did not locate the global optimum in any of the ten runs. It is proposed that the systematic verification of the adequacy of optimisation should be an integral part of model calibration exercises. The form of verification should depend on the specific problem at hand.

Introduction

The rainfall-runoff process is the component of the hydrological cycle involving the time-space conversion of precipitation into runoff on land. The process is very complex considering the large number of factors involved and their variability. Rainfall-runoff models idealise the rainfall-runoff process to varying degrees and in various forms. A plethora of rainfall-runoff models exist with Wheater et al. (1993) putting the number into the hundreds. Some of the applications of rainfall-runoff modelling include:

- Creating, extending or filling in missing runoff data for water resources assessments
- Flood peak estimation and flood forecasting
- Investigating the hydrological and water quality impacts of land use and climate change on catchments
- Studying and modelling specific processes in the catchment such as sedimentation.

A common mode of classification based on model complexity, groups rainfall-runoff models into: empirical, conceptual and process models (Grayson and Chiew, 1994, Wheater et al., 1993). Empirical models mimic the rainfall-runoff process to a slight degree while conceptual models mimic the processes more closely by using interconnected storages and simple equations to represent the water movement among them.

Empirical and conceptual rainfall-runoff models are not designed to use parameters that are directly measurable in the field and their parameters are usually obtained by calibration. Process models, most of which are distributed, were initially designed and intended to represent the physical processes closely enough to enable the use of measurable parameters only. Inadequacy of data and model imperfections have, however, been found to limit the application of process models in this 'ideal' manner and process model applications invariably include some form of calibration (Refsgaard and Knudsen, 1996; Refsgaard, 1997; Western et al., 1997; Demetriou and Panthakey, 1997).

Calibration is the determination of a parameter set that gives a simulated hydrological or hydrochemical series that adequately matches the observed series. It is fundamentally an iterative process involving:

i) The simulation/s using (a) parameter set/s from the search space to obtain the model performance/s.
ii) The determination of (a) parameter set/s that is/are likely to perform better than that/those used in the previous simulation/s; and simulation/s using the new parameter set/s.
iii) The repetition of step (ii) until a satisfactory performance is obtained or until further improvements are negligible.

During calibration, the performance is quantified by an objective function. Some commonly used coefficients to quantify the quality of the simulated series are the coefficient of determination and the bias. Graphical plots such as hydrograph and scatter plots are also applied often. The three steps of the process could be undertaken manually or automatically using an optimisation method. Effective automatic methods are preferable to manual methods as they give a better chance of obtaining superior parameter sets. However, where a modeller is adequately experienced with a given model, manual calibration could suffice. This publication deals with automatic calibration methods.

While research on model calibration has been active for decades, most of the studies have focused on the location of the global optimum and/or the efficiency at which this is achieved (Johnston and Pilgrim, 1976; Duan et al., 1992; Bates, 1994; Tanakamaru and Burges, 1996; Kuczera, 1997). Gan and Biftu (1996) included the simulation performance in validation providing a notable exception to this trend.

* To whom all correspondence should be addressed.
☎(031) 204-4749; fax (031) 204-4291; e-mail jndiritu@dixie.udw.ac.za
Received 10 December 1998; accepted in revised form 23 February 1999.

In practical model calibrations, although global optimum location is desirable, other issues such as the quality of the model simulation could be more important. It is often found that a unique global optimum parameter set does not exist because of parameter correlations. The detection of parameter correlations therefore becomes an integral component of parameter identification. An adequate level of optimisation should therefore detect parameter correlations precisely. This could be particularly important where modifications to the structure of the rainfall-runoff model are being considered or in model regionalisation studies. If significant parameter correlations exist, any parameter set that gives an objective function value very close to the lower bound can be considered a global optimum parameter set. The lower bound of the objective function is the lowest value obtained from an adequate number of randomly initialised optimisations.

This study was aimed at assessing the effect of the level of optimisation on:

- Model simulation performance
- Parameter identification
- The detection of parameter correlations
- The capability to locate the global optimum.

The traditional split sample calibration-validation approach was used. Four different optimisation levels were obtained by using two formulations of the genetic algorithm (GA) method: the traditional and an improved GA. With each formulation, different optimisation parameters were used to achieve different levels of optimisation. The MODHYDROLOG, a conceptual model applied extensively in Australia (Chiew and McMahon, 1994) was applied in the analysis using historical 16 year-long rainfall, potential evapotranspiration and runoff series' from a South Australian catchment.

The traditional genetic algorithm

The GA method is a population-based optimisation method that uses the concepts of natural selection and natural evolution as hypothesised by Darwin in his theory of evolution. The GA approach was initiated by Holland (1975) and has been used and studied extensively since then. Goldberg (1989), Davis (1991) and Bäck (1996) give comprehensive reviews of the GA. Rainfall-runoff model calibration is an example of the global optimisation problem that takes the form of Eq. (1). The decision variables x_i ($i=1$ to n) for the calibration problem are the model parameters. f is the objective function, which quantifies the extent of agreement between the simulated series and the observed series.

$$\text{minimise} \quad f(x_1, x_2, \ldots, x_i, \ldots, x_n) \qquad (1)$$

subject to:

$pmin_i \leq x_i \leq pmax_i, \quad i=1,2, \ldots, n$

$g(x_j) \leq 0$

$g(x_j)$ are the inequality constraints on some or all the design variables

$[pmin_i - pmax_i]$ is the feasible range of the decision variables.

Following is a description of the traditional GA applied to the general form of the global optimisation problem (Eq. (1)). The main steps of the traditional GA are: coding, simulation and selection, crossover, mutation and replacement.

Coding

A binary substring of bit length l is used to represent each decision variable. Figure 1 illustrates a possible coding for a 6-parameter problem using an l value of 5. The coding is commonly referred to as a chromosome.

x_1	x_2	x_3	x_4	x_5	x_6
00010	10101	11101	01010	10100	01110

Figure 1
A chromosome for a parameter set

The lower limit of the search range $pmin_i$ is represented by the decoded integer 0 and the upper limit $pmax_i$ by the decoded integer 2^l-1. The actual value of the decision variable x_i is determined by linear interpolation of its decoded integer value. As an example, the decoded value of parameter x_2 of Fig. 1 is 2^0x1 + 2^1 x 0 + 2^2 x 1 + 2^3 x 0 + 2^4 x 1 = 21 and $x_2=pmin_i + [21/(2^5\text{-}1)] \times (pmax_i\text{-}pmin_i) = pmin_i + (21/31)\ (pmax_i\text{-}pmin_i)$. A population of p chromosomes is generated randomly in the search space where p is the population size.

Computation of fitness and selection

The chromosomes are decoded, the x_i values obtained and simulations done to obtain the objective function *(f)* values for each individual of the population. Fitnesses are then computed using a function that gives higher fitness for individuals with lower objective function values. Equation (11) in Section 5 gives an example of such a function. The chromosomes (parents) to use for the generation of new chromosomes (children) of the next generation are selected in proportion to the fitness. Each individual is assigned to a range of a magnitude directly proportional to its fitness. All the ranges are proportioned such that they all together span from zero to one. Two random numbers lying between zero and one are then generated. The two individuals assigned to the ranges within which the two random numbers lie are selected as a pair of parents. This is repeated 0.5 x c x p times where c is the probability of crossover. The probability of crossover is the proportion of individuals of the current population that will be replaced by children to form the new population.

Crossover and mutation and replacement

Taking n as the number of parameters, l x n gives the bit length of each chromosome. A point is selected randomly in the range l x n, and the codes of the pairs of parents are exchanged. This is repeated for all the 0.5 x c x p pairs of parents. In the illustration of Fig. 2, parents 1 and 2 are individuals of a 6-parameter problem in which a 5 bit binary coding has been used. With a crossover point of say 15 obtained randomly within the range 1 to 29, the chromosomes of the two children are as shown.

Chromosome for parent 1: 000111000101010 ↓ 101000110010101
Chromosome for parent 2: *010100011100001* ↓ *101010100000110*
Chromosome for child 1: 000111000101010 ↓ *101010100000110*
Chromosome for child 2: *010100011100001* ↓ 101000110010101

Figure 2
An illustration of crossover

To implement mutation, a small proportion of the bits of the children are changed (1 becomes 0 and 0 becomes 1). The proportion of the bits to change is referred to as the probability of mutation. If m is the probability of mutation, then $m \times n \times l \times p$ randomly selected bits are changed.

The $0.5 \times c \times p$ pairs of children are then used to replace $c \times p$ randomly selected individuals of the initial population. This gives the new population.

Selection gives fitter individuals a better chance of creating children. Crossover helps to obtain more useful combinations of the useful 'traits' within the fitter individuals and mutation helps to generate new 'genetic material' leading the search to new regions of the search space. Fitter individuals are thus obtained and the population on the whole also becomes fitter as optimisation proceeds. The steps **"Computation of fitness and selection"** and **"Crossover and mutation and replacement"** are repeated until a designated termination criterion is met. This could be set at the maximum number of simulations or generations, the minimum improvement of the best performance in successive generations, or the known global optimum.

The traditional GA is direct and population-based and is therefore expected to work effectively on rough response surfaces and on problems with multiple optima. However, studies with the traditional GA have shown that it is ineffective in obtaining the global optimum (Wang et al., 1995, Tanakamaru and Burges, 1996, Kuczera, 1997). This is attributed to two features. Firstly, the inability to fine-tune as the traditional GA searches only in a fixed grid of the search space. Secondly, no explicit approach for leading the search into several regions of attraction exists. The search therefore converges into a single region of attraction, which may contain a local and not the global optimum. Ndiritu (1998) developed approaches to overcome the two limitations. These consist of three procedures: Fine-tuning, hill-climbing and the use of independent subpopulation searches coupled with shuffling. The improved GA was found to effectively locate the global optima of a rainfall-runoff model calibration and the Griewank function, a hard theoretical problem (Ndiritu and Daniell, 1997).

The improved genetic algorithm

Fine-tuning

Fine-tuning is achieved through the gradual reduction and shifting of the search range. After every specified number of generations denoted as *s1*, the locations of the parameters of the best performing individuals in a specified number of successive previous generations denoted as *s2* are checked. If the values are within a small portion of the search space, then the search range is reduced and centralised about the parameter value of the best individual of the current generation. This gives a finer grid and a more concentrated search. Equations (2) to (5) give the fine-tuning strategy.

After every s1 generations and for all dimensions i = 1,2, ... , n,

$$r1_{i,g} = 2(\max xb_i^j - \min xb_i^j) / (x\max_i - x\min_i) \quad j = g - s_2 + 1,..,g \qquad (2)$$

$$r1_{min} \le r1_{i,g} \le r1_{max} \qquad (3)$$

$$p\max_{i,g} = xb_i^g + r1_{i,g}\left(p\max_i - p\min_i\right) \qquad (4)$$

$$p\min_{i,g} = xb_i^g - r1_{i,g}\left(p\max_i - p\min_i\right) \qquad (5)$$

where:

xb_i^i is the value of parameter x_i for the best performing individual in generation j

$\max xb_i^i$ and $\min xb_i^i$ are the maximum and minimum values of xb_i^i in the past *s2* generations (including the current generation *g*).

$r1_{min}$ and $r1_{max}$, are the fine-tuning control parameters to check premature convergence as a result of exceedingly low $r1_{i,g}$ values and enlargement of the search range in case $r1_{i,g}$ values are too high.

$pmin_i$ and $pmax_i$ are the search range limits before fine-tuning.

$pmin_{ig}$ and $pmax_{ig}$ the search range limits after fine-tuning.

After a series of trials, the following values of the fine-tuning parameters were selected: 5, 5, 0.4 and 0.5 for *s1*, *s2*, $r1_{min}$ and $r1_{max}$ respectively.

Hill-climbing

Hill-climbing consists of range-shifting towards regions of the search space that are more promising and also serves to prevent premature convergence when the fine-tuning routine is in operation. Hill-climbing is implemented after every given number of generations denoted as *s3*. The strategy is described in Eq. (6) to (8) using the same notation as for fine-tuning. The shift $sh_{i,g}(pmax_i - pmin_i)$ in Eq. (7) and (8) is the deviation from the middle of the current search range, of the mean of the values of the best individuals in the last *s4* generations (including the current one). A value of 5 was found reasonable and used for both *s3* and *s4*.

After every s3 generations and for all dimensions i = 1,2, ..., n,

$$sh_{i,g} = \frac{\left(\left[\sum_{j=g-s4+1}^{g} xb_{i,g} \Big/ s4\right] - 0.5\left(pmax_i + pmin_i\right)\right)}{\left(pmax_i - pmin_i\right)} \qquad (6)$$

$$pmax_{i,g} = p\max_i + sh_{i,g}\left(p\max_i - p\min_i\right) \qquad (7)$$

$$pmin_{i,g} = p\min_i + sh_{i,g}\left(p\max_i - p\min_i\right) \qquad (8)$$

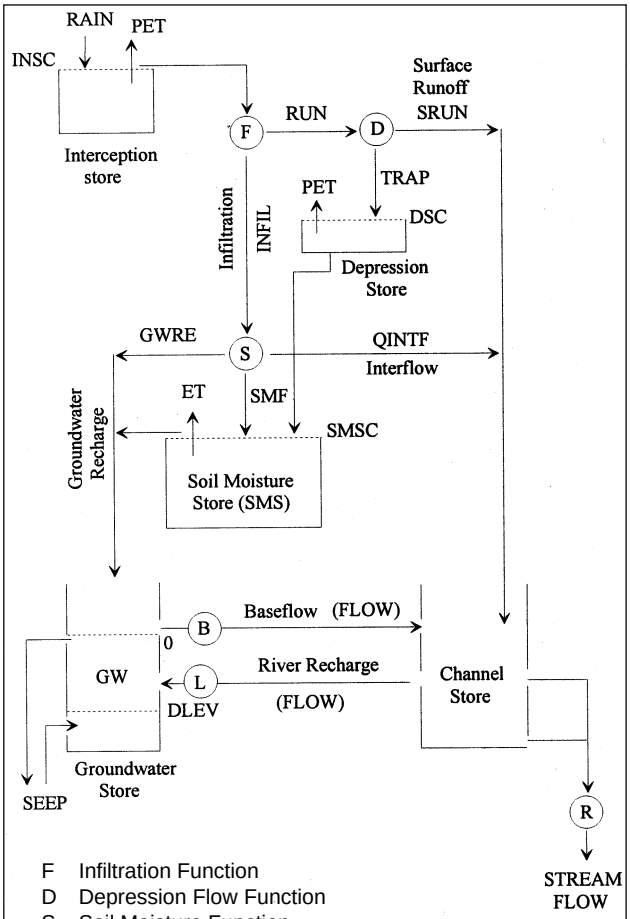
Search range-shifting in fine-tuning and hill-climbing enables the GA to search beyond the initially prescribed search space – a feature that could be particularly useful when dealing with unfamiliar problems and/or data. It is, however, appropriate that the search should be prevented from straying into unrealistic regions. Two search spaces are therefore specified. One is the initial search space $[Xmin_{1i} - Xmax_{1i}]$ which allows for the optimiser's intuition and experience but which the search can go beyond. The other is the limiting search space $[XLmin_i - XLmax_i]$, which allows for the input of the known constraints and beyond which the search is not allowed.

Independent subpopulation searches and shuffling

The total population (of p individuals) is split into n_s subpopulations of size p_s each. Each subpopulation searches independently to an optimum. The subpopulations are then shuffled as follows:

The population is ranked in order of performance to form a matrix $[ch(i),\ i=1,2, ..., p\ ]$ where $ch(1)$ is the best performing and $ch(p)$ the worst performing individual. The individuals are then allocated to the subpopulations using Eq. (9) to effect the shuffling procedure of Duan et al. (1992):

$$sch(i, j) = ch(i + p_s(j - 1)) \qquad (9)$$



F Infiltration Function
D Depression Flow Function
S Soil Moisture Function
B Baseflow Function
L River Recharge Function
R Nonlinear Routing Function

RAIN = Rainfall
PET = Potential Evapotranspiration
INFIL = Lesser of { **COEFF** exp(-**SQ**.SMS/**SMSC**), RAIN-**INSC**}
RUN = RAIN-**INSC**-INFIL
TRAP = (**DSC**-**ADS**·ARGD) exp(-**MD**·**DSC**/RUN)
SRUN = RUN-TRAP
QINTF = **SUB**·(SMS/**SMSC**)·INFIL
GWRE = **CRAK**·(SMS/**SMSC**).(INFIL-QINTF)
SMF = INFIL-QINTF-GWRE
ET = lesser of {**EM**·(SMS/**SMSC**), PET}
FLOW = **K1**.|GW|+**K2**·[1-exp(-**K3** |GW|)
SEEP = **VCOND**·(GW-**DLEV**)

Figure 3
Model structure of MODHYDROLOG (adapted from Chiew and McMahon, 1994)

where:

sch(i,j) is the *j* th (*j* = 1,2, ..., p_s) individual in the *i* th (*i* = 1,2, ..., n_s) subpopulation.

The least and highest parameter values of each subpopulation give the initial search space [$Xmin_{1i}$ - $Xmax_{1i}$] for the next set of optimisations for the subpopulations. Each set of subpopulation optimisations is referred to as an epoch. The first subpopulation retains all the p_s individuals from the previous epoch and uses them as the initial population. The initial individuals for the other subpopulations are generated randomly.

Rainfall-runoff model and data

The MODHYDROLOG, a daily rainfall-runoff model used extensively in Australia was chosen for the simulation of monthly flows. Sixteen years of daily rainfall, potential evapotranspiration and runoff data spanning 1970 to 1985 from the 27 km^2 Scott Creek catchment located in South Australia was used. The grass-covered catchment is located 20 km south of Adelaide, and has a predominantly duplex soil type. For the study period, the mean annual rainfall, potential evapotranspiration and runoff values were 950 mm, 1 080 mm and 130 mm respectively. The data were split into two samples of eight years each: 1970-77 and 1978-85.

Figure 3 presents the basic structure of MODHYDROLOG and the equations governing water flow. Table 1 presents brief descriptions of the model parameters while Table 2 gives the parameter settings and ranges applied for this study. Nine of the 19 parameters of MODHYDROLOG were set to constant values provided by Chiew and McMahon (1994) and the other ten were optimised. Chiew and McMahon (1994) provide more information on the MODHYDROLOG. An objective function that uses the square roots of the observed and predicted values given as Eq. (10) was applied in the calibrations.

$$Minimize \qquad \sum_{i=1}^{N}\left(\sqrt{arun_i}-\sqrt{run_i}\right)^2 \qquad (10)$$

where:

N is the length of data;
$arun_i$ is the observed discharge; and
run_i is the estimated discharge.

Methodology

Four different levels of optimisation denoted as case A to case D were obtained by using the two GA formulations and varying the optimisation parameters for each as shown in Table 3. Case A and B used the traditional GA while case C and D used the improved GA. These parameters were selected on the basis of experience and trial runs. For all the optimisations, the bit length of parameter substring *l*, probability of crossover, *c*, probability of mutation, *m* used was 20, 1.0 and 0.05 respectively. Instead of the traditional 2-point crossover (Fig. 2), a 10-point crossover was applied. In this crossover, the chromosome of a child is obtained from the chromosomes of the two parents with each contributing 5 sections.

Elitism, the process that maintains the best individual of the current generation in the next one was also applied. Instead of the traditional proportionate selection (Section 2.2), tournament selection was adopted. In tournament selection, individuals are selected randomly from the population and the fittest one chosen as a parent. The number of individuals from which the parent is selected is referred to as the tournament size. A variable power equation form of fitness scaling (Eq. (11)) was used.

$$sf_1 = \left(\left(\sum_{i=1}^{p} f_1\right) \Big/ f_1\right)^{(g\,+\,3000/3000)} \qquad (11)$$

where:

sf_i is the scaled fitness
f_i *the* raw objective function value
g is the generation.

Parameter *cheperf* as applied in Eq. (12) and (13) specified the level of convergence to be achieved before the termination of an optimisation. Equation (12) was applied to populations (cases A

TABLE 1
PARAMETERS OF THE MODHYDROLOG MODEL

Parameter	Description
ADS	Fraction of total area which is depressional
CO	Routing coefficient
COEFF	Maximum infiltration loss parameter
CRAK	Constant of proportionality in the determination of groundwater recharge
DLEV	Parameter used in deep seepage equation
DSC	Depression storage capacity
EM	Maximum plant-controlled rate of evapotranspiration
INSC	Interception store capacity
K1	Constant of proportionality in linear part of aquifer-stream flow equation
K2	Constant of proportionality in exponential part of aquifer-stream flow equation
K3	Exponent in exponential part of aquifer-stream flow equation
LOCATE	Parameter to fix the origin of the seasonal cycle of COEFF, CRAK and SUB
MD	Exponent in depression flow equation
POWER	Routing exponent
SEAS	Parameter to fix the amplitude in the seasonal fluctuation of COEFF, CRAK and SUB
SMSC	Soil moisture store capacity
SQ	Exponent in infiltration capacity equation
SUB	Constant of proportionality in the calculation of interflow
VCOND	Constant of proportionality in deep seepage equation

TABLE 2
PARAMETER SETTINGS AND RANGES FOR MONTHLY SIMULATIONS WITH THE MODHYDROLOG MODEL

Parameter	Set value	Parameter	Initial parameter ranges $[Xmin_{1i} - Xmax_{1i}]$		Parameter range limits $[XLmin_i - XLmax_i]$	
ADS	0	INSC	1	2	0.5	6
MD	1	COEFF	90	190	20	400
DSC	0	SQ	1	5	0	10
LOCATE	1	SUB	0.1	0.3	0	1
SEAS	0	CRAK	0.1	0.3	0	2
POWER	0	SMSC	130	230	20	400
K2	0	EM	7	13	5	20
K3	0	CO	10	30	1	50
DLEV	-0.1	K1	0.02	0.06	0	1
		VCOND	0.05	0.15	0	0.5

TABLE 3
OPTIMISATION PARAMETER SETTINGS FOR THE FOUR LEVELS OF OPTIMISATION

Optimisation parameter	Symbol	Value for case			
		A	B	C	D
Population size	p	20	50	50	200
Subpopulation size	p_s	20	50	10	20
Number of subpopulations	n_s	1	1	5	10
Maximum number of function evaluations	Ev_{max}	5 000	5 000*	5 000*	25 000*
Number of crossover positions	n_{cross}	10	10	10	10
Tournament size	t_o	6	15	3	6
Maximum number of epochs	ep_{max}	1	1	50	50

* The epoch at which maximum function evaluations reached is allowed to complete.

and B) or subpopulations (case C and D) while Eq. (13) was applied to epochs (cases C and D):

$$if \quad \sum_{i=g-4}^{g} obf_i \Big/ \sum_{i=g-9}^{g-5} obf_i \geq cheperf \qquad terminate\ the\ search \tag{12}$$

$$if \quad obf_{cep} \Big/ obf_{pep} \geq cheperf \qquad terminate\ the\ search \tag{13}$$

where:

obf_i is the objective function of the best individual in generation i

obf_{pep} and obf_{cep} are the best objective function values of the previous and the current epoch respectively.

Cases B and C were also used to evaluate the significance of the GA modifications. It was observed that a single run of case B was reaching convergence at around 800 function evaluations while case C was converging at about 5 000 evaluations. To enable a reasonable comparison of case B and case C, additional runs were made for case B until and including the run at which 5 000 evaluations were exceeded. The optimisation among all the runs giving the least objective function value was then taken as the result of an optimisation of case B. For cases C and D, the epoch in which 5 000 and 25 000 simulations were respectively attained was allowed to complete and the search then terminated. The actual number of function evaluations therefore exceeded these values by small amounts.

Split sample calibration-validation was applied. The 1970-77 data series was used for calibration and the 1978-85 series for validation. Ten runs with different random initialisations were made with each of the four cases. This allowed an evaluation of the consistency of the objective function, the consistency of parameters from different runs and an evaluation of parameter correlations. To quantify runoff simulation performance, the following four coefficients were applied:

- The coefficient of efficiency (*ce*), a measure of the variance
- The bias (*bias*), a measure of the ability to predict the volume of discharge
- The absolute deviation (*ade*), a measure of the average departure of the predictions at every time step
- A residual mass curve coefficient (*rmcc*), a measure of systematic errors in the simulations.

Equation (14) to (17) describe the coefficients:

$$ce = 1 - \left(\sum_{t=1}^{tm} (run_t - arun_t)^2 \Big/ \sum_{t=1}^{tm} (arun_t + \overline{arun})^2 \right) \tag{14}$$

$$bias = \sum_{t=1}^{tm} (run_t - arun_t) \Big/ \sum_{t=1}^{tm} arun_t \tag{15}$$

$$ade = \sum_{t=1}^{tm} \left| run_t - arun_t \right| \Big/ \sum_{t=1}^{tm} arun_t \tag{16}$$

$$rmcc = 1 - \sum_{t=1}^{tm} \left| (rmsim_t - rmact_t) / rmact_t \right| \tag{17}$$

where:

$arun_i$ is the historical discharge,

run_i the estimated discharge, and

$\overline{arun}$ the mean of the historical flows.

In Eq. (17), $rmsim_t$ and $rmact_t$ are respectively, the residual mass curve values for the simulated and actual flow at time *t*. The coefficient of efficiency (*ce*), is commonly used (Nash and Sutcliffe, 1970; Wang et al., 1995; Hughes, 1995; Yapo et al., 1996; Gan and Biftu, 1996; Refsgaard and Knudsen, 1996; Franchini et al., 1996). The *bias* has also been used by Sorooshian et al. (1993), Yapo et al. (1996), Chiew and McMahon (1994), Gan and Biftu (1996) and Refsgaard and Knudsen (1996). Aitken (1973) used a residual mass curve coefficient for the quantification of systematic errors in simulated flows. The coefficient of efficiency and the residual mass curve coefficient take on maximum values of unity in the case of a perfect fit. The bias and absolute deviation take perfect fit values of zero. The absolute deviation (*ade*) takes on a minimum value of zero but the other coefficients could take negative values.

Results and discussion

Table 4 gives the objective function values (*obf*) in mm and the corresponding number of function evaluations (*eval*) obtained. Figure 4 compares the objective function values graphically. It is observed from both Table 4 and Fig. 4 that better objective function values were obtained with higher optimisation effort. The difference between case C and D, however, seems minimal except for one of the runs of case C which gave an *obf* of 3.494 mm. Case C indeed gave the overall least objective function value (3.399 mm). Based on the objective function values, the ten runs of case D and 9 runs of case C were considered to have practically located the global optimum.

Although cases B and C used about the same number of function evaluations (approximately 5 000), better objective function values were obtained with case C indicating that the improved GA performed better.

Figure 5 shows the relative locations of parameters from the ten calibrations and serves to assess parameter identification. The scaled values were obtained as $(xb_i^o - XLmin_i)/(XLmax_i - XLmin_i)$ where xb_i^o is the optimised value of parameter *i*. These plots show a better identification with cases C and D. Case A gives the poorest parameter consistency.

Although very close objective function values were obtained for the ten runs of case D (Table 4), a unique parameter set was not obtained (Fig. 5 case D). This observation is not unusual (see for instance Sorooshian et al., 1993; Tanakamaru and Burges, 1996) and is caused by parameter interdependence. An analysis of parameter interdependence revealed high correlation coefficients between parameter SQ and COEFF and between parameter CRAK and SUB. These observations conformed with the MODHYDROLOG model structure. Parameters SQ and COEFF are both used in the infiltration rate function while CRAK and SUB determine the proportioning of the infiltrated water into interflow and groundwater recharge. Figure 6 presents the correlations for the two pairs of parameters where R^2 denotes the correlation coefficient and SE the standard error of observations. Cases C and D gave better correlations than cases A and B with case A giving the poorest correlations.

The average simulation performance coefficients obtained for the ten runs of each optimisation level are given in Table 5. It is evident that the simulation performance generally improved with optimisation effort. For the validations, the differences among the four levels were, however, not considerable except with the residual mass curve coefficient (*rmcc*). The calibration values are included for completeness. In calibration, the most notable observation was the considerably lower bias values for cases C and D. The higher *rmcc* values for cases A and B in calibration were not

expected but are not considered significant. A possible explanation for this is the difference in the variations quantified by the objective function and those quantified by the *rmcc*. Minimising the objective function may not always minimise the *rmcc*. Plots of the individual coefficients (*ce*, *bias*, *ade* and *rmcc*) indicated a lower variability for the higher optimisation levels.

The simulation performance results indicate that location of the global optimum may not necessarily give simulations significantly superior to those obtained with lower levels of optimisation - especially in validation. The global optimum search could at times therefore be of only secondary importance in practical modelling. Problems associated with model structure and data inadequacy, errors and lumping (in space and time) could have a greater impact on simulation performance than the applied optimisation effort (Gan and Biftu, 1996, Ye et al., 1996, Sorooshian et al., 1993). These factors and the non-representativeness of calibration data could also provide an explanation why validation simulations are invariably poorer that calibration simulations. Table 5 demonstrates the poorer validation results obtained in this study.

The results obtained here indicate that significantly better parameter identification and parameter correlations are obtained with the high optimisation effort unlike the simulation performance. The adequacy of optimisation should therefore depend on the task at hand and the quality and adequacy of available data. Without checking for this, there is always the likelihood that better modelling would have been achieved. A manual calibration approach may not be effective to check for calibration adequacy. For a modeller with extensive experience with a specific model, manual calibration could probably suffice for some applications. However, even with automatic calibration, the method should allow for varying levels of optimisation effort. To accomplish this, modifications may be required. The improvements to the traditional GA presented and tested here, are more comprehensively described by Ndiritu (1998).

TABLE 4
OBJECTIVE FUNCTION VALUES AND NUMBER OF FUNCTION EVALUATIONS FOR THE FOUR OPTIMISATION CASES (A TO D)

Run	Case A		Case B		Case C		Case D	
	eval	***obf***	***eval***	***obf***	***eval***	***obf***	**eval**	**obf**
1	320	3.620	5 300	3.548	5 040	3.425	25 040	3.418
2	260	3.827	5 050	3.568	5 030	3.402	25 040	3.426
3	360	3.664	5 400	3.628	5 010	3.395	25 080	3.407
4	380	3.985	5 350	3.599	5 030	3.415	25 040	3.403
5	320	3.847	5 350	3.583	5 020	3.494	25 020	3.418
6	240	3.684	5 250	3.552	5 030	3.422	25 020	3.405
7	300	3.677	5 350	3.534	5 030	3.407	25 040	3.409
8	400	3.742	5 550	3.622	5 030	3.428	25 040	3.423
9	320	4.497	5 300	3.605	5 020	3.399	25 020	3.418
10	280	5.442	5 500	3.548	5 020	3.415	25 000	3.410

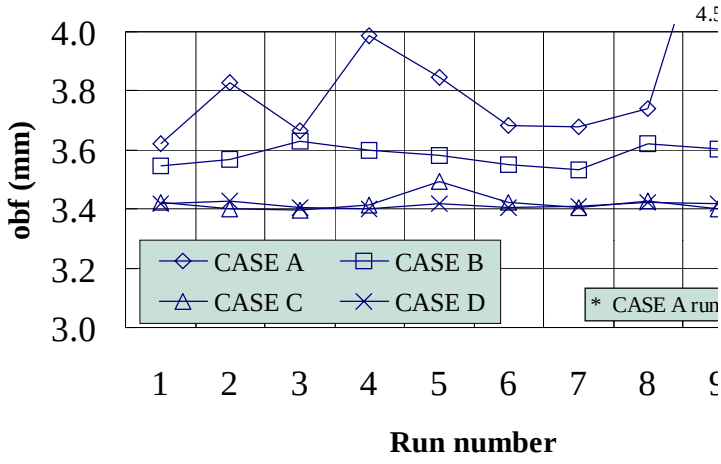


Figure 4
Objective function values for four levels of optimisation

TABLE 5
AVERAGE SIMULATION PERFORMANCES FOR FOUR LEVELS OF OPTIMISATION

Case	Validation			
	ce	***bias***	***ade***	***rmc***
A	0.8517	0.4139	0.458	0.2771
B	0.8492	0.422	0.4641	0.2793
C	0.8549	0.413	0.4529	0.3328
D	0.8556	0.4117	0.4512	0.3413
Case	**Calibration**			
	ce	***bias***	***ade***	***rmc***
A	0.9692	-0.0183	0.1399	0.9298
B	0.9757	-0.0155	0.128	0.9249
C	0.9784	-0.0092	0.1207	0.9149
D	0.9791	-0.0093	0.1201	0.9123

Conclusions and recommendations

An assessment of the adequacy of rainfall-runoff model calibration was done using four optimisation levels obtained from a global optimisation method. Higher optimisation levels were found to give better objective functions, better parameter identification, a more distinct detection of parameter interdependence and slightly better simulations in both calibration and validation. The seemingly low effect of the level of optimisation on simulation performance was considered to result from inadequacies of the model structure and data. It is also possible that the model readily fitted the data resulting in an easy calibration problem and a consequent low significance of the optimisation effort.

It is proposed that checks on the adequacy of optimisation should be an integral part of rainfall-runoff modelling. Such checks

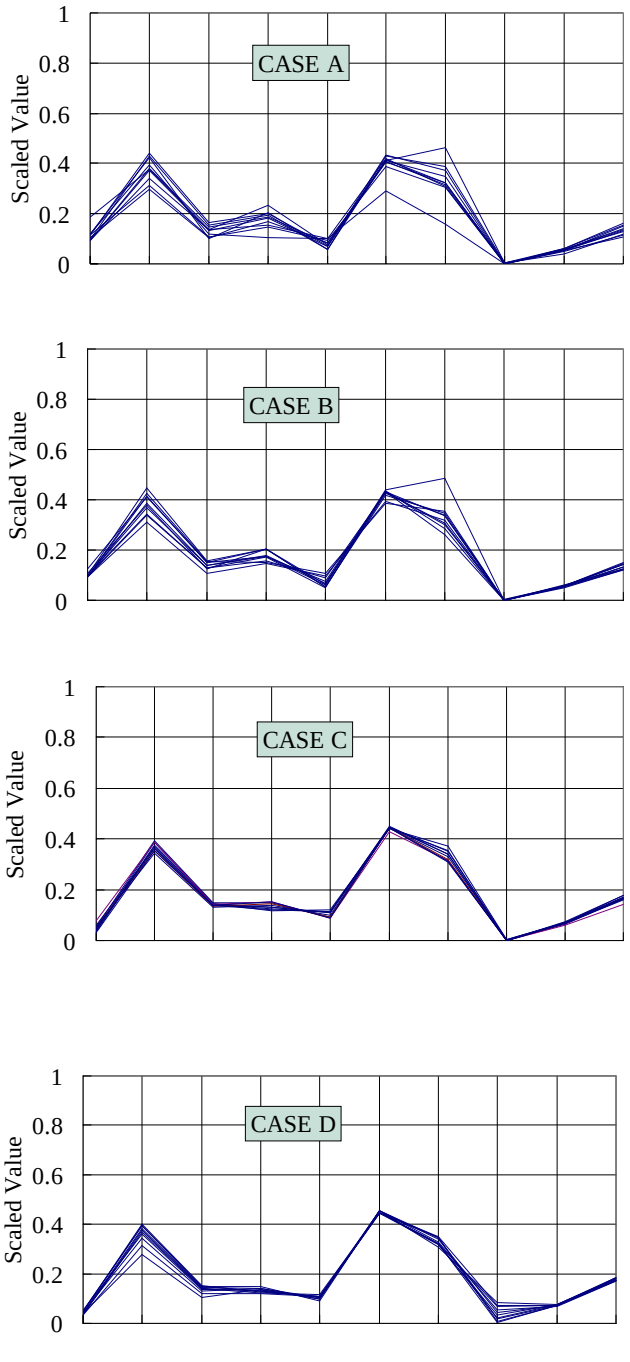


Figure 5
Parameter identification for four levels of optimisation

should be guided by the specific purpose/s of the modelling. The optimisation method should easily allow for an increase in the optimisation effort if the need arises. Multiple randomly initialised optimisations with global optimisation methods such as the improved genetic algorithm used in this study or the shuffled complex evolution (Duan et al., 1992) are reasonable choices for effective calibration. In this study, an average of about 35 min was required to implement one calibration consisting of about 5 000 simulations (case C of Table 3) using a 166 MHz 32 KB RAM Pentium. Because the simulation times of some models could be large (e.g. 15 CPU hours taken for the TOPOG_YIELD model to simulate 12 years at a daily time step on an IBM RS6000 model 320H, (Vertessy et al., 1993)), research aimed at improving the efficiency of global optimisation methods (e.g. Kuczera, 1997)) is a good choice for the future.

Acknowledgments

FHS Chiew and T A McMahon of the Centre Data for Environmental Applied Hydrology, University of Melbourne provided a hydrological model and data. This is gratefully acknowledged. The authors also express thanks to the Australian Agency for International Development (AusAID), the sponsors of one of the authors at the University of Adelaide where a considerable part of this work was conducted.

References

AITKEN AP (1973) Assessing systematic errors in rainfall-runoff models. *J. Hydrol.* **20** 121-136.

BÄCK T (1996) *Evolutionary Algorithms in Theory and Practice*. Oxford Univ. Press, New York. 314 pp.

BATES BC (1994) Calibration of the SFB model using a simulated annealing approach. *Proc. Int. Hydrol. and Water Resour. Symp.*, Adelaide, Vol.3, 1-6.

CHIEW FHS and McMAHON TA (1994) Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian Catchments. *J. Hydrol.* **153** 383-416.

DAVIS L (ed.) (1991) *Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New York.

DEMETRIOU C and PANTHAKEY J F (1997) Integrated hydrogeological model development for the Wakool irrigation district. *Proc. MODSIM 97, Int. Congr. on Modelling and Simulation*, Modelling and Simulation Soc. of Aust., Hobart, Vol. **1** 323-328.

DUAN Q, SOROOSHIAN S and GUPTA V (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* **28**(4) 1015-1031.

FRANCHINI M, WENDLING J, OBLED C and TODINI E (1996) Physical interpretation and sensitivity analysis of the TOPMODEL. *J. Hydrol.* **175** 293-338.

GAN TY and BIFTU GF (1996) Automatic calibration of rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure. *Water Resour. Res.* **32**(12) 3513-3524.

GOLDBERG DE (1989) *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison Wesley.

GRAYSON RB and CHIEW FHS (1994) An approach to model selection. *Proc. Int. Hydrol. and Water Resour. Symp.*, Adelaide, Vol.**1** 507-512.

HOLLAND JH (1975) *Adaptation in Natural and Artificial Systems*. Univ. of Michigan Press, Arbor. 183 pp.

HUGHES DA (1995) Monthly rainfall-runoff models applied to arid and semi-arid catchments for water resource estimation purposes. *Hydrol. Sci.* **40** (6) 751-769.

JOHNSTON PR and PILGRIM DH (1976) Parameter optimization for watershed models. *Water Resour. Res.* **12**(3) 477-486.

KUCZERA G (1997) Efficient subspace probabilistic parameter optimization for catchment models. *Water Resour. Res.* **33** (1) 177-185.

NASH JE and SUTCLIFFE JV (1970) River flow forecasting through conceptual models Part I - A discussion of principles. *J. Hydrol.* ***10*** 282-290.

NDIRITU JG and DANIELL TM (1997) An improved genetic algorithm for rainfall-runoff model calibration and function optimization. *Proc. MODSIM 97, Int. Congr. on Modelling and Simulation*, Modelling and Simulation Soc. of Aust., Hobart, Vol. **4** 1689-1694.

NDIRITU JG (1998) An Improved Genetic Algorithm for Rainfall-Runoff Model Calibration. PhD Thesis, Univ. of Adelaide, Australia.

REFSGAARD JC and KNUDSEN J (1996) Operational validation and intercomparison of different types of hydrological models. *Water Resour. Res.* **32** (7) 2189-2202.

REFSGAARD JC (1997) Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.* **198** 69-97.

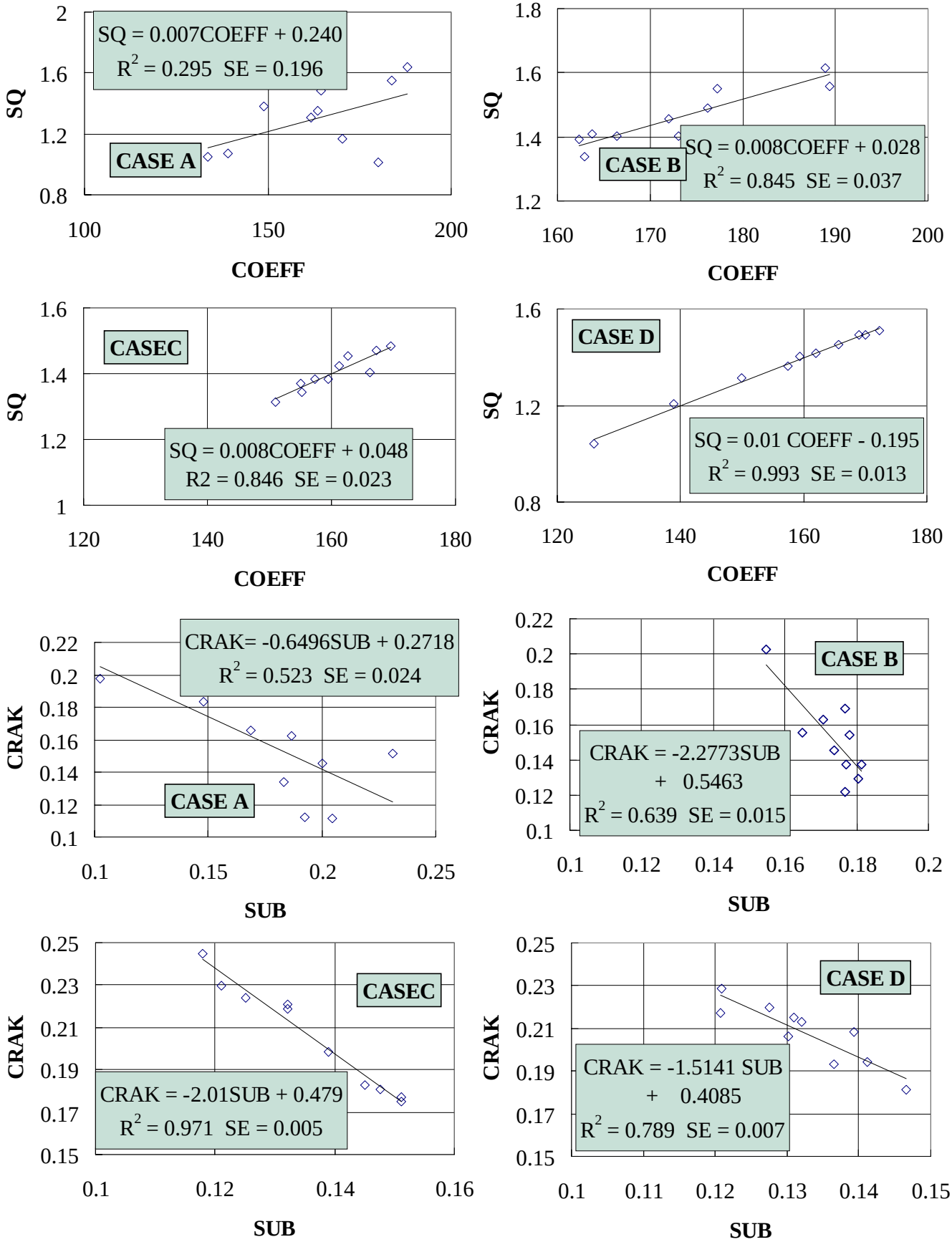


Figure 6
Parameter correlations for four levels of optimisation

SOROOSHIAN S, DUAN Q and GUPTA VK (1993) Calibration of rainfall-runoff models: application of global optimization to the Sacramento soil moisture accounting model. *Water Resour. Res.* **29** (4) 1185-1194.

TANAKAMARU H and BURGES SJ (1996) Application of global optimization to parameter estimation of the TANK model. *Proc. Int. Conf. on Water Resour. and Environ. Res.*, Kyoto, Japan, Vol. **2** 39-46.

VERTESSY RA, HATTON TJ, O'SHAUGHNESSY PJ and JAYASURIYA MDA (1993) Predicting water yield from a mountain ash catchment using a terrain analysis based catchment model. *J. Hydrol.* **150** 665-700.

WANG QJ, CHIEW FHS and MCMAHON TA (1995) Calibration of environmental models by genetic algorithms. *Proc. MODSIM 95, Int. Congr. on Modelling and Simulation*, Modelling and Simulation Soc. of Aust., Newcastle, Vol. **3** 185-190.

WESTERN AW, GREEN TM and GRAYSON RB (1997) Hydrological modelling of the Tarrawarra catchment: Use of soil moisture patterns. *Proc. MODSIM 97, Int. Congr. on Modelling and Simulation*, Modelling and Simulation Soc. of Aust., Hobart, Vol. **1** 409-416.

WHEATER HS, JAKEMAN AJ and BEVEN KJ (1993) Progress and directions in rainfall-runoff modelling. In: Jakeman AJ, Beck MB and McAleer MJ (eds.) *Modelling Change in Environmental Systems*, John Wiley & Sons, England.

YAPO PO, GUPTA HV and SOROOSHIAN S (1996) Automatic calibration of rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.* **181** 23-48.