Int. J. Water, Vol. 3, No. 4, 2007

System dynamics and auto-calibration framework for NSM model: Murrumbidgee River

Amgad Elmahdi*

CSIRO, land and water, Adelaide,
PMB 2, Glen Osmond SA 5062, Australia
Fax: +61-8-83038750
E-mail: Amgad.Elmahdi@csiro.au
*Corresponding author

Hector Malano and Teri Etchells

Department of Civil & Environmental Engineering,
The University of Melbourne,
Victoria 3010, Australia
E-mail: h.malano@civenv.unimelb.edu.au
E-mail: t.etchells@civenv.unimelb.edu.au

Abstract: Water sharing management is the major problem for water resources and irrigation management decision makers. However, irrigation systems are very complex and interconnected, posing significant difficulties in managing irrigation economically and environmentally. Therefore, it is imperative that innovative modelling approaches are employed to deal with the feedback loops inherent in these systems. Through the application of a system dynamics approach, a Network Simulation Model (NSM) was developed. The purpose of the NSM is to measure and identify the change in economic and environmental outputs of various allocations and demand scenarios. The aim of this study is to examine the use of two methods of auto calibration (single objective and multiobjective) over a variety of climatic and hydrological conditions. These methods have been compared and applied to three periods of calibration and validation using seven performance criteria. Results indicate that multiobjective method yields better identifiable parameters and an improved model structure

Keywords: calibration; parameter estimation; performance criteria; system dynamics; irrigation system.

Reference to this paper should be made as follows: Elmahdi, A., Malano, H. and Etchells, T. (2007) 'System dynamics and auto-calibration framework for NSM model: Murrumbidgee River', *Int. J. Water*, Vol. 3, No. 4, pp.381–396.

Biographical notes: Amgad Elmahdi is a Research Scientist with CSIRO land and water, Adelaide lab. His PhD research may change the way water resources are managed – a topic of increasing importance in Australia and throughout the world about underground dams – the new water-storage solution. He has more than 11 years' experience in various aspects of hydrology and water management and holds three Masters (Water Ecological Studies, Environmental Conservation and Land and Water Management).

In 2003, his latest thesis was named Best Master Thesis in Water Resources Management by the Egyptian Ministry of Water Resources and Irrigation.

Hector Malano has conducted research on various aspects of water resources at three scales: (i) on-farm modelling of surface irrigation systems, (ii) modelling of irrigation distribution networks and (iii) water allocation between competing uses at the catchment level. More recently, he has conducted research on modelling to total water and solute cycles in catchments with significant irrigation systems. He has authored and co-authored over 100 scientific papers on these topics. More recently, he concluded a three-year term as Vice-President of the International Commission on Irrigation and Drainage. He has consulted for several international organisations including the World Bank, AusAID and Food and Agriculture Organisation of the United Nations.

Teri Etchells is a Research Fellow with the Department of Civil and Environmental Engineering at The University of Melbourne. She has a significant track record of research focusing on water allocation, accounting and markets, with a particular emphasis on managing uncertainty. Prior to completing her PhD in water markets and exchange rates, she spent three years as a Strategy Consultant with The Boston Consulting Group and several years as a Cadet Engineer with Melbourne Water Corporation. She graduated top of her class from Monash University in 1997 with degrees in Engineering (first class honours) and Commerce.

1 Introduction

A Network Simulation Model (NSM) (Elmahdi et al., 2004) has been developed using a system dynamics approach as an appropriate tool to analyse irrigation demand management strategies. This model is designed to operate on a monthly basis to assist managers in analysing the system's behaviour under various management scenarios. It also demonstrates the effect of climate variability between cropping seasons. The model gives priority to environmental demand as the main constraint at end-of-system flow and then attempts to satisfy the biophysical crop demand. The major drive for understanding the link between irrigation demand and environmental demand is to measure and identify the change in economic output and environmental impacts of various allocations and demands from irrigation on improved seasonality of flows. The NSM incorporates a wide range of complexities likely to be encountered in water resource management: surface and ground water sources, water trading between sources, system constraint such as maximum ground water pumping rates, environmental flows and channel capacity.

Many studies point out a clear difference between model verification, calibration and validation processes (Mihram, 1972; Schlesinger et al., 1974; Stedinger and Taylor, 1982). According to Schlesinger et al. (1974) verification is defined as "The certification that the model is implemented on the computer in a manner that truly depicts the idealized model". Stedinger and Taylor (1982) described the idealised model as a group or series of mathematical equations and logical relations which are used to represent the real system. In contrast, calibration is the process of selecting the 'best' parameters to populate a model, assuming that the model is correct. Finally, Schlesinger et al. (1974)

describe the model validation as "quality of match of simulated and real data with some interpretation of the appropriateness of the data for validation purposes". Also, Stedinger and Taylor (1982) stated that validation tests used to indicate a simulation model reasonably represent a real system.

Many years ago, the calibration processes were implemented using a trial and error approach to estimate model parameters. However, Madsen et al. (2002) assert this approach can be very inefficient, depending on how many free parameters the model has and their interrelations. Furthermore, Madsen et al. (2002) claimed that manual calibration may not follow a logical process because of the complexity in identifying and quantifying model performance. Thus, a great deal of research has been devoted to developing autocalibration tools that are able to overcome the complexity and time issues. In the last decade, the model calibration process has received many advantages by using computer system technology such as flexibility and better fit between model measurements and real system data. This robust fit could be achieved by using a good optimisation tool which has become available such as PEST (parameter estimation optimisation software), a non-linear parameter estimation and optimisation package. Moreover, one of the main advantages gained from these optimisation tools is its capability to understand the uncertainty of the optimisation outcomes, which helps identify the extent of uncertainty in the estimated parameters.

Doherty and Johnston (2003) stated out that the computer-based tools for autocalibration have been developed for many hydrological models in the last decade. In particular, many research studies have focused on development of strong optimisation tools that would be able to better estimate the model parameters (Kuczera, 1983; Sumner et al., 1997) and several studies have compare the relative performance of those tools (Kuczera, 1997; Madsen et al., 2002; Thyler et al., 1999). Another aspect which is handled by several studies is measuring the uncertainty in the parameters estimation processes (Kennedy and O'Hagan, 2001; Kuczera and Parent, 1998) and measuring the performance of models when they are used in conditions where there is a lack of data (Beven, 2000; Jakeman and Hornberger, 1993). It is recommended by environmental stakeholder groups and regulatory agencies; model validation uncertainty analysis is considered part of the model implementation and structure (NRC, 2001).

In this context, the main step in autocalibration is defining the objective function which represents the model objective. Based on this body of research, it can be concluded that the most effective automatic algorithm is the global population-evolution (is a selection from a given domain which yields either the highest value or lowest value (depending on the objective), when a specific function is applied), compared to multistart local search and pure local search methods. Mostly, the automatic calibration routines used a single objective function to compare the observed and the simulated results for example, the sum of squared errors. However, a single objective function is not necessarily appropriate to capture most of the main characteristics of the system (i.e. hydrograph) which is applied by the hydrologist to quantify the performance of the calibrated model. The main advantage of applying multiobjectives calibration is its solution overcomes the uniqueness parameters problem. Thus, in this paper, single and multiobjective automated calibration methods are applied and analysis is undertaken to determine which is the best. Specifically, a single objective measure (i.e. the Sum of Square Error (SSE)) and a multiobjective measure are used to calibrate NSM and results are compared under different time cases.

2 Parameter identification and data

Once the NSM has been structured and formulated; the second step is the estimation of a suitable parameter set, that is, the actual calibration process of the model parameters. The parameters of the model structure are adjusted until the observed system output and the model output show acceptable levels of consistency. NSM has divided the Murrumbidgee River into five reaches. Each reach has inflow and outflow measured by gauge stations (Figure 1) and uses the water balance Equation (1) to determine intermediate results.

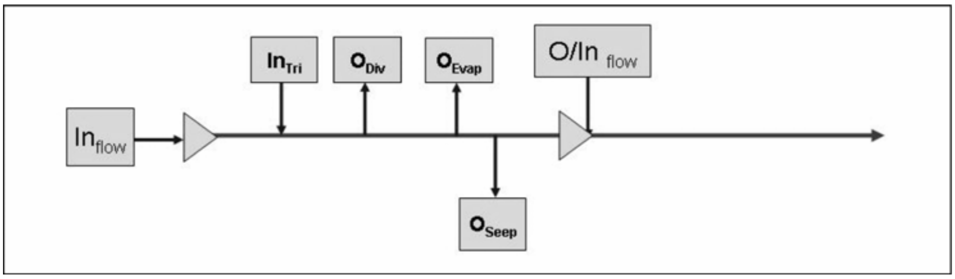
$$\frac{O}{\text{In}_{\text{Flow}}} = \left\{\text{In}_{\text{Flow}} + \text{In}_{\text{Tri}}\right\} - \left\{ O_{\text{Div}} + O_{\text{Evap}} + O_{\text{Seep}}\right\} \tag{1}$$

where $O/\text{In}_{\text{Flow}}$ is the outflow from river reach which will be inflow for the following reach, In_{Flow} is the inflow for each river's reach. For the first reach, representing the dams' release, a simple water balance has been used to calculate (dam's inflow) inflow and outflow of the two main dams (Burruinjuck and Blowering dam) of the Murrumbidgee River. The simple water balance compares the measured change in dam storage level with the measured outflow and computes the difference as total inflow (Equation (2)).

$$\begin{aligned}\text{Dam inflow} = \text{dam storage level current}_{(m)} - \text{dam storage level former}_{(m-1)} \\ +\text{outflow}\end{aligned} \tag{2}$$

If it is negative (–ve) it means that losses have exceeded gains for example, evaporation and infiltration are greater than total catchment inflow. The advantage of the water balance is that it picks up all sources and types of inflows and losses, whereas an inflow gauge would only measure inflow from a particular stream. Real dam inflows are complex distributed systems and out of the scope of this study, but a simple water balance equation captures the essential elements in this case.

Figure 1 The river reach schematic



In_{Tri} is the sum of tributary inflow for each reach and calculated by Equation (3). The first part of reach from the dam's wall to Gundagai has four tributaries and another three tributaries between Gundagai and Wagga Wagga, whereas the second reach, Wagga Wagga to Narrandera has only one tributary. The monthly flow data for these tributaries is extracted from PINNEENA 8- river operation database produced by DIPNR (2004).

$$\text{In}_{\text{Tri}} = \sum_{i}^{n} \text{In}T_i \tag{3}$$

where $\text{In}T_i$ is i tributaries gauge inflows. The O_{Div} is the total diversion from each reach which is calculated based on the cropping pattern of the irrigation areas in each river reach and O_{Evap} is the total evaporation loss from each reach calculated as in Equation (4).

$$O_{\text{Evap}} = R_{\text{W}} \times R_{\text{L}} \times \text{Evap} \times F_{\text{Evap}} \tag{4}$$

where R_{W} is the top surface reach width, R_{L} is reach length, Evap is the evaporation data and F_{Evap} is the evaporation factor. The evaporation data is calculated by compiling pan-evaporation data collected from the National Weather Service (NWS) and SILO meteorology for the land (evaporation stations). Each river reach evaporation data is calculated as average of the nearest evaporation stations. While the top reach width data calculated through cross section analysis with rating table at each inflow and outflow stations point along the Murrumbidgee River by taking the average width at inlet and outlet of the reach to represent the top reach width which depends on stream flow-stage analysis to determine the stage of the river section. While O_{Seep} is the seepage losses calculated by Equation (5).

$$O_{\text{Seep}} = R_{\text{WP}} \times R_{\text{L}} \times F_{\text{Seep}} \times N_{\text{Days}} \tag{5}$$

where R_{WP} is the reach wetted perimeter width, the main assumption is that the river channels are trapezoid. According to Ven te Chow (1959), natural channels are usually irregular and altering from trapezoid to parabola. Using PINNEENA-8-database DIPNAR (2004) to generate a cross-section analysis of the Murrumbidgee River at different inlet and outlet of each reach to extract wetted perimeter value and take the average for each month-based on flow-stage analysis. While, R_{L} is reach length; F_{Seep} is the average seepage factor mm/d and N_{Days} is the number of days.

To sum up, two parameters have been defined and selected for calibration process. These two parameters are F_{Evap} and F_{Seep} evaporation factor and seepage rate, respectively and allowed to search for the global optimal solution using Powell's optimiser (will be described in briefly later) to calculate the objective functions. Once the model parameters are chosen, the optimiser will try to find the values for those parameters that achieve the objective function that means making the model fit the data as closely as possible.

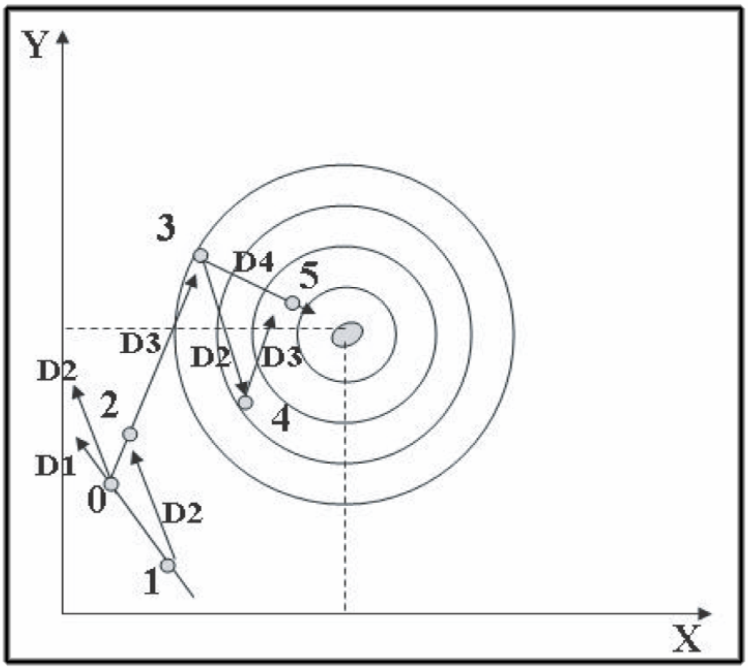
3 Calibration methods

3.1 Vensim^{TM} Powell's optimiser (system dynamics)

To do an optimisation under VENSIM$^{\text{TM}}$, a set of parameters must be specified to search over and create a payoff that determines how good an individual simulation is. The payoff for a model can be defined in terms of a comparison of model variables with actual data. Moreover, optimisation is controlled by optimisation domain or control which define the max and min range for each parameter. VENSIM$^{\text{TM}}$ optimiser is based on modified Powell theory. Powell theory is considered one of the main techniques that try to find the minimum or maximum of a function of several variables and could be defined as multidimensional unconstrained optimisation which does not require derivative evaluation (Chapra and Canale, 2002; Powell, 1964a,b). According to Chapra and Canale (2002), one of the best methods and algorithms that capitalised on the idea of pattern directions to find optimum values efficiently is named Powell's method. Basically, Powell's method depends on finding points 1 and 2 by searching in the same direction from different starting points and the line between 1 and 2 should drive to the maximum, this line called conjugate directions.

Powell's method can be initiated from any starting point (as discussed below). For example, assume the search begins from point 0 (Figure 2) and searches in two directions D1 and D2 until the maximum is reached at point 1 in direction D1. Next, the search approach through D2 direction to find point 2. After that, a new direction D3 from points 0 and 2 will form the starting point to find point 3. Then from point 3, using the previous search direction D2 to find point 4 and so on, from point 4, to find point 5 through D3 direction. At this time, points 3 and 5 have been allocated by the same direction D3 from different starting point. Powell theory has approved that the directions D4 along points 3 and 5 and D3 are conjugate directions which lead directly to the maximum (Chapra and Canale, 2002) see Figure 2.

Figure 2 Powell's method



In addition, the starting point or the rules used to compute search starting points under VENSIM^{TM}, results are influenced by the domain value: the upper and lower constraints defined into the model (i.e. min< = parameter< = max). The starting point value for the parameter will be determined by Equation (6). A number *x* will be chosen automatically generated under VENSIM^{TM} by random method in the range [0, 1) including zero but not including 1. Finally, Powell's optimiser is used to solve single and multiobjective functions which are built into the NSM to calibrate and compare its performance by applying different assessment model performance criteria to determine the best case and result.

$$\text{Parameter} = \min + X(\max - \min) \tag{6}$$

3.2 Single objective function

Conventional autocalibration using a single objective function is prompt and objective, while its results reproduce the behaviour of the selected objective function. The most common single objective function in literature is the sum square error (see Equation (7)) where, O_i is the observed volume at time *i* and S_i is the simulated flow volume at time *i*, *N* is the number of time series. Frequently, hydrologists do not accept the results from single objective because of their interest in many other aspects of the system performance (Boyle et al., 2000).

$$F_1 = \sum_{i=1}^{n} \left(O_i - S_i\right)^2 \tag{7}$$

3.3 Multiobjective calibration

A multiobjective algorithm function is used to find the model population necessary to fit all aspects of the hydrograph at different points along Murrumbidgee River. In this context, four basic objective functions are considered simultaneously as follows (see Equations (8)–(11)):

1 Overall volume error

$$F_a = \left| \frac{1}{N} \left(\sum_{i}^{N} \left(O_i - S_i\right) \right) \right| \tag{8}$$

2 Overall RMSE

$$F_b = \sqrt{\frac{1}{N} \left(\sum_{i}^{N} \left(O_i - S_i\right)^2 \right)} \tag{9}$$

3 SSE

$$F_c = \sum_{i=1}^{n} \left(O_i - S_i\right)^2 \tag{10}$$

4 Relative RMSE of all flow events

$$F_d = \sqrt{\left[\frac{1}{N} \left(\sum_{i}^{N} \left(\frac{O_i - S_i}{O_i + S_i} \right)^2 \right) \right]} \tag{11}$$

All these functions are built inside the NSM as mathematical functions and are being optimised at once (i.e. minimising all of them) assuming all of these functions have equal priority and weight, to compare their results with the single objective function. Eight years dataset is used for calibration by using different three-period cases. The period for calibration is used to find the optimum value for the parameter that achieved the objective function and the period of validation is used to assess the calibration performance as follows:

Case A: two years for calibration and six years kept for validation and called (2–6 yrs).

Case B: three years for calibration and five years kept for validation and called (3–5 yrs).

Case C: five years for calibration and three years for validation and called (5–3 yrs).

4 Assessment model performance criteria

To measure the model performance for each validation test, a standard set of criteria has been defined comprising seven numerical measures. These criteria were selected in order to reflect the objectives of the modelling that are able to simulate river flows at multiple sites along the Murrumbidgee River monthly. The seven selected criteria are as follows:

1 A measure of the ability to simulate the variation in the flow hydrograph for a particular river gauging station, see Equation (12) (Nash and Sutcliffe, 1970).

$$E = 1 - \frac{\sum_i^N \left(O_i - S_i\right)^2}{\sum_i^N \left(O_i - \bar{O}\right)^2} \quad E = [0,1] \tag{12}$$

where O_i = observed flow at time step i, S_i = simulated flow at time step i and $\bar{O}_i$ = is the mean of the observed flow. Values close to 1.0 for Nash and Sutcliffe indicate the model is able to capture or accurately simulate the variance of the flows in the river. Sometimes E called the coefficient of efficiency was chosen as the likelihood measure to evaluate the accuracy of both the magnitude and timing of predicted flows (e.g. Andersen et al., 2001; McMichael et al., 2006; Tague et al., 2004; Vazquez and Feyen, 2003). Andersen et al. (2001), indicate that when the value of $E > 0.85$, the model run is good.

2 A measure of the deviation between simulated and observed flow (Henriksen et al., 2003) by Root Mean Square (RMS), see Equation (13):

$$\text{RMS} = \frac{1}{N}\sqrt{\sum_i^N \left(S_i - O_i\right)^2} \quad \text{RMS} = [0,\infty] \tag{13}$$

where O_i = observed flow at time step i, S_i = simulated flow at time step i and N = is the total number of time steps.

3 A measure of the ability to simulate the average flow for a particular river gauging station. Sometimes it called Water Balance Error (difference between average simulated and average observed flow normalised by average observed flow), see Equation (14) (Henriksen et al., 2003; Madsen et al., 2002).

$$F_{\text{Bal}} = \frac{\left|\bar{O} - \bar{S}\right|}{\bar{O}} \times 100\% \text{ error in water balance} \qquad (14)$$

where $\bar{O}_i$ = is the mean of the observed flow and $\bar{S}_i$ = is the mean of the simulated flow.

4 %Bias is a measure of the bias in the validation results from the observed flow (Hogue et al., 2005).

$$\%\text{Bias} = \left[\frac{\sum_i^N \left(S_i - O_i\right)}{\sum_i^N O_i}\right] \times 100$$

where O_i = observed flow at time step i, S_i = simulated flow at time step i and N = is the total number of time steps.

5 A measure of the ability to simulate low flow conditions for a particular river gauging station (Henriksen et al., 2003; Wood, 1974).

$$F_{\text{low}} = \sum_i^N \left[\frac{\left(O_i - S_i\right))\bar{O}}{O_i^2}\right]^2$$

where O_i = observed flow at time step i, S_i = simulated flow at time step i and N = is the total number of time steps. The low flow measure could be normalised by the minimum environmental flow requirement which is 6000 Ml/month for the Murrumbidgee water sharing plan 2004.

$$F_{\text{Low-norm}} = \frac{F_{\text{Low}}}{F_{\text{min-env}}}$$

where $F_{\text{min-env}}$ = the minimum environmental flow.:

6 Correlation coefficient: correlation was calculated for all river gauge stations in both cases (single and multiobjective calibration) to demonstrate the correlation between validated flow and observed flow.

$$R = \frac{\sum_i^N \left(O_i - \bar{O}\right)\left(S_i - \bar{S}\right)}{\sqrt{\sum_i^N \left(O_i - \bar{O}\right)^2}\sqrt{\sum_i^N \left(S_i - \bar{S}\right)^2}}$$

where O_i = observed flow at time step i, S_i = simulated flow at time step i, $\bar{O}_i$ = is the mean of the observed flow and $\bar{S}_i$ = is the mean of the simulated flow.

7 Median percentage: a measure of the ability to simulate the median monthly flow for a particular river gauging station, or it is a percentage error to simulate the median monthly flow. where O_{median} is the median observed flow and S_{median} is the validated median flow

$$F_{\text{median}} = \left|\frac{O_{\text{median}} - S_{\text{median}}}{O_{\text{median}}}\right| \times 100$$

The performance/goodness of the model performance during calibration and validation periods is evaluated according to these performance criteria. These criteria are not of the fail/pass type, but assess the performance of the model under the two calibration methods and three time cases. It is hard to judge which case is the best as every criteria has its value and range. Thus, there is need to use scoring and ranking system that facilitates the evaluation procedure method of model performance. Using the ranking system used by Lorup et al. (1998) and Henriksen et al. (2003) for *E*, Fbal and correlation, other rankings are suggested for %Bias, FL and RMS/median-based on the previous ranked system, see Table 1. This ranking system could be used to calculate the overall score and acceptable case. Aggregation score system is used. The model performance results in a score of 1–5 points for each validation test for example, correlation value for one single gauge station according to the classification in the five intervals. For example, one of the five gauging stations with a correlation performance equal to excellent is given five points, and very good correlation gives four points and so on.

Table 1 Ranking and scoring system

Points	*Rank category*	*E*	F_{Bal}	*Correlation*	*%Bias*	*RMS/ median*	*FL*
5	Excellent	>0.85	<5	>0.85	<5	1–10	0.0–0.01
4	Very good	0.65–0.85	5–10	0.65–0.85	5–10	10–20	0.01–0.02
3	Good (average)	0.5–0.65	10–20	0.5–0.65	10–20	20–30	0.02–0.1
2	Poor	0.2–0.5	20–40	0.2–0.5	20–40	30–50	0.1–0.15
1	Very poor	<0.2	>40	<0.2	>40	>50	0.15–1

5 Calibration and validation results

In this section, the calibrated results are compared visually and quantitatively with the observed and simulated results. The calibration procedure can be summarised as follows for the second case (3–5 yrs):

- NSM simulate the discharge of several gauging stations along the Murrumbidgee River for three water years (June 1993–June, 1996).
- the simulated results are compared with real data (observed) for each river reach points inflow and outflow (gauge points)
- calibrate the three years' results by applying two methods single objective (minimising the SSE between the data (observed) and simulated results) and multiobjective functions simultaneously
- the same steps have been repeated with 2–6 years and 5–3 years.

Moreover, Schlesinger et al. (1974) indicate that model validation concerns the quality of match of simulated and real data with some interpretation of the appropriateness of the data for validation purposes. As mentioned above, two parameters were used for calibration (evaporation factor and seepage rate) by Powell's methods search for their optimum values within its range (minimum value and maximum value) defined based on literature review and data from the local authorities, see Table 2.

Table 2 The parameters and its range

Parameter name	*Initial value*	*Minimum range*	*Maximum range*	*Source*	*Calibration results value*
Evaporation factor	0.7	0.5	1	FAO 56 (Richard et al., 1998)	0.52
Seepage rate	6 mm/day	2 mm/day	8 mm/day	MDBC (2004–2005) and Willem et al. (2005)	2.1
Channel loss	20%	15%	30%	CIA Environmental Report 2000–2005	18.5

Briefly, applying the value of the parameter obtained from the calibration process to simulate the model for the validation period, the validation period is changed with each case such as case 3–5 yrs it means five years; the model run for validation is compared with real qualitative and quantitative data. Table 3 shows the correlation and *E*-values results at different gauge stations under the two calibration methods (SO = single objective calibration and MO = multiobjectives calibration method). It is very clear the correlation increased by moving upstream in both calibration methods without any difference. While the validation results with multiobjective method show better correlation than the single objective method, there is a slight difference between both methods at Darlington Point and Hay gauge stations. This could be attributed to high diversion which impacts the discharge flow at these stations and this is captured by multiobjective methods.

Table 3 Correlation and *E*-value for both calibration and validation results under both calibration methods

Gauge stations	MO-cal		SO-cal		MO-val		SO-val	
	Correlation	E-value	Correlation	E-value	Correlation	E-value	Correlation	E-value
Balranald	0.73	0.31	0.73	0.34	0.69	0.55	0.69	0.56
Hay	0.82	0.63	0.82	0.63	0.91	0.69	0.89	0.73
Darl Pt	0.93	0.77	0.93	0.78	0.94	0.73	0.93	0.76
Narrendera	0.94	0.69	0.94	0.69	0.95	0.72	0.95	0.72
Wagga Wagga	0.95	0.76	0.95	0.76	0.97	0.86	0.97	0.86

Note: MO-cal: multiobjective calibrated; SO-cal: single objective calibrated; MO-val: multiobjective validated and SO-val: single objective validated results.

Based on the ranking system described above and correlation results, the model could be ranked as excellent for all gauge stations except for the last gauge downstream; it is ranked very good performance 0.69 in both methods. Moreover, the *E* values result is aligned with correlation results with some minor differences. The model shows high *E* value upstream which is decreased by going further downstream. But the overall average of the model *E* value 0.71 and 0.73 validation results for multiobjective and single objective methods, respectively. According to Henriksen et al. (2003) the model under E values in both cases could be categorised as very good performance. It is clear each performance criterion shows different results and is not necessary true, high correlation and consistent visually usually lead to high performance model. Thus, this study is using different performance criteria to test and evaluate the NSM model after calibration. In general, the upper reaches show good results in most assessment criteria comparing to the lower reaches with massive irrigation diversion.

The results are consistent with Khan et al. (2004) and Pratt (2004) results which indicate there is need to validate flows and diversion statistics in the lower reaches because of measurements and reporting uncertainty. By using the ranking system in Table 1, the NSM model calibrated could be ranked between excellent for upper reaches and poor for lower reach. Table 4 shows the results for all cases under the ranking system.

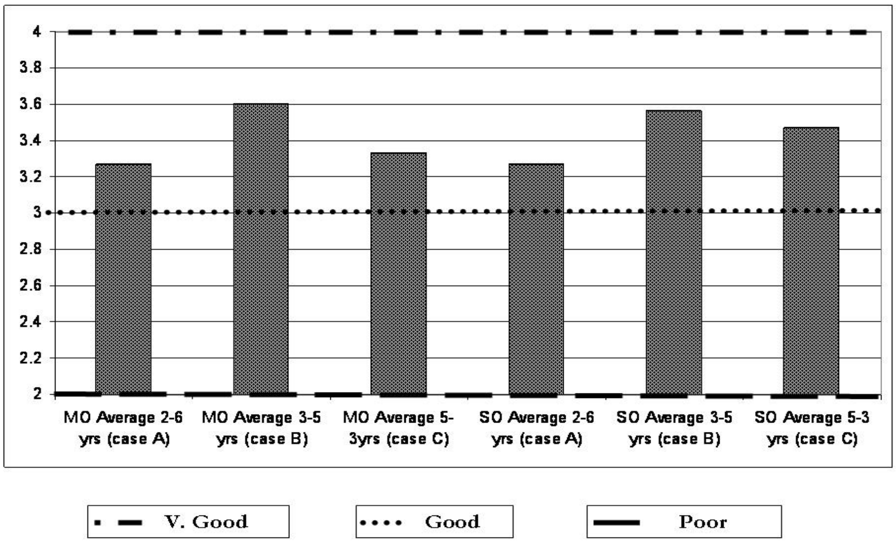
Table 4 The average ranking system for all calibration cases

Performance indicator validation	*E*	F_{bal}	*Percent bias*	*Correlation*	*FL norm*	*RMS/ median*
MO Average 2–6 yrs (case A)	3.2	2.8	2.2	4.4	2.8	4.2
MO Average 3–5 yrs (case B)	4	2.8	2.8	4.8	3.2	4
MO Average 5–3 yrs (case C)	2.4	2.4	2.4	4	4.4	4.4
SO Average 2–6 yrs (case A)	3.2	2.6	2.2	4.4	3	4.2
SO Average 3–5 yrs (case B)	4	2.6	2.6	4.8	3.4	4
SO Average 5–3 yrs (case C)	2.6	2.6	2.8	4	4.4	4.4

This ranking system could be used to calculate the overall score and acceptable method and case. Aggregation score system is used as discussed above. It is clear that the multiobjectives calibration average results for all cases show better score compared to single calibration method average results which show the same average correlation for all cases. The scoring results indicate that the model calibration results are located between good and very good rank such as correlation, E value and $\mathrm{Fl}_{\mathrm{norm}}$ except for F_{Bal} and %Bias located between poor and good, but very close to good condition. Moreover, by looking at the scoring results for the two calibration methods, it is obvious that both almost have the same value except for F_{Bal} and %Bias. Multiobjective methods have a higher and better score than single objective method.

Consequently, the multiobjective method could be recommended for model calibration and its result can be used as the best method. Furthermore, the average score for all performance criteria under the same case 3–5 years; this analysis shows multi objective calibration method is better than the single objective method (average score 3.53 and 3.48, respectively). Moreover, the average score is located between good and very good rank while multiobjective average score is close to very good and single objective average score is close to good rank. Figure 3 describes which calibration case under both calibration methods could be recommended and its results used.

Figure 3 Overall performance score for the validation results



It is clear that case 3–5 years (case B) achieved the best result in both methods under different performance criteria, which could be recommended and its results can be used. To sum up, the overall performance for the three calibration cases and two calibration methods indicate that the overall best result which could be used is that of multiobjective method under case 3–5 years, where the model performance is very good with highest overall score.

6 Conclusion

The aim of this study is to test and compare two automated calibration methods using system dynamics approach for NSM model. This model is designed to operate on a monthly basis and can assist managers in analysing the system's behaviour under various management scenarios. NSM model is calibrated and validated for eight years period by using two calibration methods-single objective and multiobjective autocalibration methods under three different cases called A, B and C (2–6 years, 3–5 years and 5–3 years) the first number indicates the number of years devoted for calibration and the second number indicates the number of years devoted for validation test.

Moreover, single objective autocalibration method was represented by sum square error while multiobjective autocalibration methods were represented by four objective functions and calibrated simultaneously (Overall volume error, Overall RMSE, SSE and Relative RMSE of all flow events). Furthermore, seven performance criteria have been applied to evaluate and assess the overall performance of the validation results and compare the two calibration methods and the three cases. A ranking and scoring system has been used to understand the overall performance of the calibration methods.

The results indicate that the multiobjective autocalibration method yields better identifiable parameters and a more well posed model structure with better performance than a single objective autocalibration method. These results are consistent with Madsen (2000) who claimed that inclusion of multi objectives in the calibration process provides better identifiable parameters. These could be attributed to the conclusion of Wagener et al. (2003) who asserted, that the single objective function ignores a great deal of information which results from the way of combination of model residuals in one objective function, which could lead to a prejudice in the model performance and reflect a specific feature. In addition, it could cause another problem with parameters identification that not necessarily affects the chosen objective function. Therefore, multiobjective functions could be successful and overcome these shortcomings. Moreover, these performance criteria recommend 3–5 years (case B) under multiobjective method to provide the best model performance result. This result is aligned with the result from MDBC report 2001 about modelling guidelines, "model calibration acceptability should be judged in relation to each of the performance measures and criteria and dividing the dataset by 20–35% for calibration period and the rest for validation period".

Acknowledgements

The NSM model system has been developed with the assistance of a grant from CRC Irrigation future (CRC-IF) and the University of Melbourne Scholarship office.

References

Andersen, J., Refsgaard, J.C. and Jensen, K.H. (2001) 'Distributed hydrological modeling of the Senegal river catchment model construction and validation', *Journal of Hydrology*, Vol. 247, pp.200–214.

Beven, K.J. (2000) 'Uniqueness of place and process representations in hydrological modelling', *Hydrology and Earth System Sciences*, Vol. 4, pp.203–213.

Boyle, D.P., Gupta, H.V. and Sorooshian, S. (2000) 'Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods', *Water Resoures Research*, Vol. 36, No. 12, pp.3663–3674.

Chapra, S.C. and Canale, R.P. (2002) *Numerical Methods for Engineers*, 4th edition, Boston: Mc-Graw Hill.

Doherty, H. and Johnston, J.M. (2003) 'Methodologies for calibration and predictive analysis of a watershed model 1', *Journal of American Water Resources Association*, Vol. 39, No. 2, pp.251–265.

Elmahdi, A., Malano, H. and Khan, S. (2004) *A System Dynamic Approach and Irrigation Demand Management Modelling*, Environmental Engineering Research Event 2004 conference 6–9 December 2004, University of Wollongong Press, ISBN: 1 74128 080 X. Available at: www.ere.org.au.

Henriksen, H., Troldborg, L., Nyegaard, P., Sonnenborg, T., Refsgaard, J. and Madsen, B. (2003) 'Methodolgy for construction, calibration and validation of a national hydrological model for Denmark', *Journal of Hydrology*, Vol. 280, pp.52–71.

Hogue, S.T., Gupta, H. and Sorooshian, S. (2005) 'A user-friendly approach to parameter estimation in hydrologic models', *Journal of Hydrology*, pp.1–16, Doi:10.1016/j.jhydrol.2005.07.009.

Jakeman, A.J. and Hornberger, G.M. (1993) 'How much complexity is warranted in a rainfall-runoff model?' *Water Resources Research*, Vol. 29, No. 8, pp.2637–2649.

Kennedy, M.C. and O'Hagan, A. (2001) 'Bayesian calibration of computer models', *Journal of the Royal Statistical Society B*, Vol. 63, No. 3, pp.425–464.

Khan, S., Beddek, R., Blackwell, J., Carroll, J., Tariq, R. and Paydar, Z. (2004) 'Whole of catchments water and salt balance to identify potential water saving options in the Murrumbidgee catchments', *Pratt Water Study Consultant Report,* Available at: (http://www.napswq.gov.au/publications/books/pratt-water/working-papers/salt-balance.html)

Kuczera, G. (1983) 'Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty', *Water Resource Research*, Vol. 19, No. 5, pp.1151–1172.

Kuczera, G. (1997) 'Efficient subspace probabilistic parameter optimization for catchment models', *Water Resource Research*, Vol. 33, No. 1, pp.177–185.

Kuczera, G. and Parent, E. (1998) 'Monte carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm', *Journal of Hydrology*, Vol. 211, pp.69–85.

Lorup, J.K., Refsgaard, J.C. and Mazvimavi, D. (1998) 'Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: case studies from Zimbabwe', *Journal of Hydrology*, Vol. 205, Nos. 3–4, pp.147–163.

Madsen, H. (2000) *Automatic Calibration and Uncertainty Assessment in Rainfall-Runoff Modeling, 2000 Joint Conference on Water Resources Engineering and Water Resources Planning and Management*, Hyatt regency MN, USA, 30 July–2 August.

Madsen, H., Wilson, G. and Ammentorp, H.C. (2002) 'Comparison of different automated strategies for calibration of rainfall-runoff models', *Journal of Hydrology*, Vol. 261, pp.48–59.

McMichael, C., Hope, A.S. and Loaiciga, H.A. (2006) 'Distributed hydrological modeling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation', *Journal of Hydrology*, Vol. 317, pp.307–324.

Murray-Darling Basin Commission (MDBC) (2004–2005) *Annual Report*, Available at: (http: //www.mdbc.gov. au/subs/annual_reports/ar0405/index.htm)

Mihram, G.A. (1972) 'Some practical aspects of the verification and validation of simulation models', *Operational Research Quaterly*, Vol. 23, No. 1, pp.17–29.

Murray-Darling Basin Commission MDBC (2001) 'Groundwater flow modelling guideline', November, Project No. 125, Final Guideline – Issue I, 16 January 2001.

Nash, J.E. and Sutcliffe, J.V. (1970) 'River flow forecasting through conceptual Models', *I.A Discussion of Principles, Journal of Hydrology*, Vol. 10, pp.282–290.

NRC (National Research Council) (2001) *Assessing the TMDL Approach to Water Quality Management*, Washington, DC: National Academy Press, p.109.

PINNEENA-8 (2004) 'DVD', *New South Wales Surface Water Data Archive*, ISSN 1449-6100.

Powell, M.J.D. (1964a) 'An efficient method for finding the minimum of a function of several variables without calculation derivatives', *The Computer Journal*, Vol. 7, pp.155–162.

Powell, M.J.D. (1964b) 'A method for minimizing a sum of squares of non-linear functions without calculating derivatives', *The Computer Journal*, Vol. 7, pp.303–307.

Pratt, W. (2004) *The Business of Saving Water*, The Report of the Murrumbidgee Valley Water Efficiency Project.

Richard, G.A., Luis, S.P., Dirk, R. and Martin, S. (1998) *Crop evapotranspiration – Guidelines for computing crop water requirements – FAO Irrigation and drainage paper 56*, FAO – Food and Agriculture Organization of the United Nations Rome.

Schlesinger, S.S., Buyan, J.R., Callender, E.D., Clarkson, W.K. and Perkins, F.M. (1974) 'Developing standard procedure for simulation, validation and verification', *Proceedings of the Summer Computer Simulation Conference*, Houston, TX, July, pp.927–933.

Stedinger, J.R. and Taylor, M.R. (1982) 'Synthetic stream flow generation, 1, Model verification and validation', *Water Resource Research*, Vol. 18, No. 4, pp.909–918.

Sumner, N.R., Flemming, P.M. and Bates, B.C. (1997) 'Calibration of a modified SFB model for twenty-five Australian catchments using simulated annealing', *Journal of Hydrology*, Vol. 197, pp.166–188.

Tague, C., McMichael, C., Hope, A. and Choate, J. (2004) 'Application of the RHESSys model to a California semi-arid shrub land catchment', *Journal of American Water Resources Association*, Vol. 40, pp.575–589.

Thyler, M., Kuczera, G. and Bates, B.C. (1999) 'Probabilistic optimization for conceptual rainfall-runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms', *Water Resource Research*, Vol. 35, No. 3, pp.767–773.

Vazquez, R.F. and Feyen, J. (2003) 'Effect of potential evapotranspiration estimates on effective parameters and performance of MIKESHE code applied to a medium size catchment', *Journal of Hydrology*, Vol. 270, pp.309–327.

Ven te Chow (1959) *Open Channel Hydraulics*, ISBN 07–010776–9, New York: Mcgraw-Hill.

Wagener, T., Mclntyre, N., Lees, M.J., Wheater, H.S. and Gupta, H.V. (2003) 'Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiably analysis', *Hydrological Processes*, Vol. 17, pp.455–476.

Water sharing in the Murrumbidgee Regulated River Water (2004) *DIPNR*, Published by NSW Department of Infrastructure, Planning and Natural Resources September 2004 DIPNR 04_194. Available at : www.dipnr.nsw.gov.au.

Willem, F.V., Paul, D., David, A., Kevin, K. and Mark, N. (2005) 'Channel seepage assessment with EC/EM and thermal imaging techniques', *Annual ANCID Conference Mildura 2005.*

Wood, S.R. (1974) 'A catchment simulation model developed for urbanizing catchments with particular reference to the use of automatic optimization techniques', *Proceedings of the IFPF Conference on Computer Simulation of Water Resources Systems*, Ghent, Belgium, pp.209–217.