

Sensitivity Analysis of a Network-Based, Catchment-Scale Water-Quality Model

J.P. Norton^{1,2,3}, L.T.H. Newham¹ and F.T. Andrews¹

¹ *Integrated Catchment Assessment and Management Centre, School of Resources Environment and Society, The Australian National University, Canberra, ACT 0200, Australia*

² *Dept. of Mathematics, The Australian National University, Canberra, ACT 0200, Australia*

³ *School of Engineering, The University of Birmingham, Birmingham B15 2TT, UK*

Abstract: Careful consideration of the uncertainties and sensitivities associated with model outputs is essential when critical decisions are made on the basis of such results. Consideration of uncertainty is particularly important in the context of natural resource management, where models are often used to tackle complex and conflicting issues across multiple scales, as is the case in evaluating management options to reduce surface water pollution. This paper describes an analysis of uncertainty in the catchment-scale integrated hydrologic, economic, stream sediment and nutrient export model known as CatchMODS. The paper briefly describes the linked components of CatchMODS and its application in the Ben Chifley Dam catchment, Australia. An initial investigation to investigate some of the most important sources of output uncertainty is described. First-order sensitivities to selected model parameters are found analytically by linearising parts of the model and used, together with knowledge of where non-linearity has most effect, to point to conditions to be investigated further. The extent of non-linear effects is also checked by comparing the analytical results with the results of parameter-perturbation tests. Results from the analysis are used to prioritise continuing model development and data-collection activities. The results are also to be incorporated into a decision-analysis framework to evaluate management options to reduce surface water pollution. The decision-analysis framework and incorporation of uncertainty analysis into it are outlined.

Keywords: Sensitivity analysis; Water quality modelling; CatchMODS model; Decision analysis framework.

1. INTRODUCTION

As natural-resource managers increase their reliance on the outputs of complex environmental models, there is an increasing need for better understanding of model behaviour, particularly the impacts of uncertainties. Such is the case where decisions are based on the outputs of hydrologic and water-quality models.

This paper describes sensitivity analysis (SA) on the catchment-scale integrated hydrologic, stream sediment and nutrient export model known as CatchMODS. The aims are first to improve understanding of the behaviour of CatchMODS and second to examine SA techniques appropriate for such models. The results of the analysis are to be further considered in a decision-analysis framework [Myšiak et al 2004] for evaluating the efficacy of management options in controlling diffuse-source pollution.

Section 2 gives a brief description of CatchMODS. The next section briefly describes the SA techniques used. The results of the SA are presented in Section 4, initially by algebra-based SA of the two-parameter non-linear nitrogen routing submodel in detail. It shows that a range of useful information is obtainable in this way with very few model runs. Experimental results from perturbations of the parameters are then used to check the extent of non-linear effects on sensitivity. Next the ease of algebraic SA for a linear dynamical model is illustrated by analysis of the linear part of the hydrological submodel. Finally, the implications of the SA results for multi-criteria decision analysis are discussed.

2. CatchMODS MODEL

The Catchment-scale Management of Diffuse Sources (CatchMODS) model simulates current conditions and the effects of land and water management activities on diffuse-source pollutant

loads. CatchMODS links several components: a regionalised hydrologic model based on the IHACRES rainfall-runoff model [Jakeman *et al.* 1990, Croke and Jakeman 2003], a suspended-sediment model developed from the SedNet model [Prosser *et al.* 2001], and simple empirical total phosphorus and total nitrogen models. The model also incorporates a simple cost-accounting component to enable the tradeoffs between environmental remediation costs (fixed and continuing) and environmental benefits (pollutant load reductions) to be explored. To provide a catchment-scale perspective, CatchMODS has a node-link spatial structure, with upstream subcatchments (typically 20-50km² in area) and river reaches (typically 7-12km long) providing input to downstream elements, so that pollutants can be routed through the stream network. Outputs are available for each subcatchment and the downstream end of each reach.

The data dependencies in CatchMODS, shown in Figure 1, are relatively simple, representing the influence of the drivers of the physical processes.

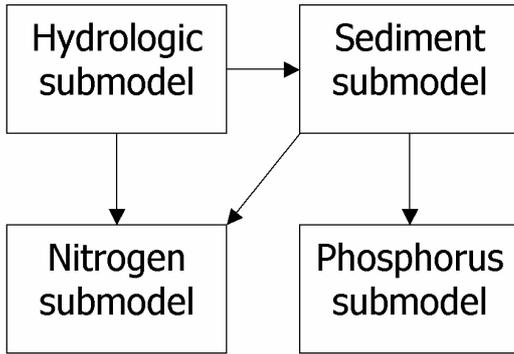


Figure 1. Data dependencies in CatchMODS.

CatchMODS has been applied in the Ben Chifley Dam catchment in New South Wales, Australia, as part of a project to improve the management of diffuse-source pollutants. A description of that application and greater detail on the model can be found in Newham *et al.* [2004].

Several features of CatchMODS make it a useful example for investigating SA techniques for environmental models: its application is network-based, allowing cascade (routing) effects to be investigated; it incorporates components with a range of complexity; the data dependencies between submodels are not complicated and submodels can be largely assessed individually; and SA of CatchMODS is part of a process of iterative model development, with CatchMODS incorporating many of the modifications suggested by Newham *et al.* [2003] following analysis of the SedNet model.

3. SENSITIVITY/UNCERTAINTY ANALYSIS

3.1 Experiment and analysis for SA

It is now widely accepted that the outputs of environmental prediction models to aid decision-making should be accompanied by quantitative assessment of their uncertainty. This ideal may not be realisable, not least because of difficulty in quantifying the contributing uncertainties. Input uncertainties are likely to include unpredicted disturbances and slow, poorly identified trends. Estimation of parameter uncertainty is usually either subjective or dependent on restrictive and perhaps unjustified probabilistic assumptions. Systematic modelling error, although assessable to some extent from the historical fit of the model to observations, is likely to be inhomogeneous, making extrapolation dubious. The next best thing to an uncertainty analysis is a sensitivity assessment, which can show which input features and model parameters influence the output behaviour most strongly and require most careful attention.

SA usually treats the model as a “black box”, investigated by Monte Carlo trials or systematic perturbation of parameter values. The latter relies on calculating (approximately) some of the derivatives in the Taylor series

$$\delta y_i = \sum_{j=1}^p \frac{\partial y_i}{\partial \theta_j} \delta \theta_j + \frac{1}{2} \sum_{j=1}^p \sum_{k=1}^p \frac{\partial^2 y_i}{\partial \theta_j \partial \theta_k} \delta \theta_j \delta \theta_k + \text{higher-order terms} \quad (1)$$

for the change in a scalar output y_i due to variations in parameters θ_1 to θ_p , assuming that the derivatives exist. For small enough variations in a model without sharp non-linearity, all individual cause-effect relations may be almost linear. The linear part of the variation of y_i with input or model parameter θ_j , determined by $\partial y_i / \partial \theta_j$, defines the conventional sensitivity

$$S_{\theta_j}^{y_i} \equiv \lim_{\delta \theta \rightarrow 0} \frac{\delta y_i / y_i}{\delta \theta_j / \theta_j} = \frac{\theta_j}{y_i} \frac{\partial y_i}{\partial \theta_j}, \quad (2)$$

normalised (relating *proportional* changes in y_i and θ_j) to remove dependence on the units employed. A vector \mathbf{y} of outputs or $\boldsymbol{\theta}$ of parameters merely requires the sensitivities to be found for all outputs and parameters. If the output is a time series, the sensitivity is an influence function of time. In all cases it can be found

approximately by noting the output change when the parameter undergoes a small perturbation. However, interaction between two or more parameters may affect the output, even if the output is linear in each parameter. To check for two-parameter interaction, for example through bilinear terms $p_{ijk}\theta_j\theta_k$, all second derivatives $\partial^2 y_i / \partial \theta_j \partial \theta_k$, $j \neq k$ must be found. To check the influences of terms up to total degree m in the parameters, including interaction between up to m parameters, all derivatives up to the m th must be found. Higher-order differences of results from more perturbation runs give them approximately. If m is high enough, this approach shows the effects of smooth non-linearities over specific perturbation ranges. In practice, the computing load to find all possibly significant derivatives may well be excessive. Moreover, sharp non-linearity may make Taylor-series approximation of the output variation impracticable. An alternative such as Monte Carlo (MC) trials over the parameter-uncertainty ranges will then be needed. There is a large literature on how best to arrange MC trials (Saltelli *et al.*, 2000), but they incur an inevitable risk of missing significant behaviour.

So far, the model has been treated as a “black box”, assuming very little prior knowledge. However, a simulation model is not a black box; its constituent relations are known, if complicated. This knowledge may help to guide SA in several ways: to look for significant interactions; to see what non-linearities are present and where they are sharp; to see what aspects of output behaviour are sensitive to particular parameter groups; and to focus successively on parts of the model with known connections to the rest, instead of considering all parameters at once. Catchment models, with relatively simple structure defined by the stream network (cascades and confluences), offer such opportunities.

Two of the components of CatchMODS will be investigated in detail: the dissolved-nitrogen transport submodel and the linear module of the IHACRES rainfall-runoff model.

3.2 SA of dissolved-nitrogen transport model

The stream network is divided into stream reaches, numbered (h,i) as shown in Figure 2, where h counts down, reach by reach, from the maximum number of reaches from the catchment outlet to the headwaters and i is odd if the stream is the left-hand tributary, even if the right-hand, at the confluence at the lower end of the reach.

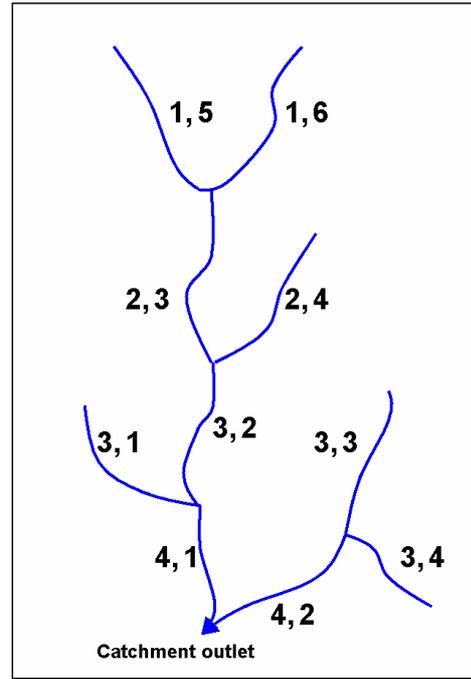


Figure 2. Example of numbering for stream reaches.

The submodel for mean annual dissolved nitrogen N at the bottom of reach (h,i) is

$$\left. \begin{aligned} N'_{hi} &= gG_{hi} + N_{h-1,2i-1} + N_{h-1,2i} \\ N_{hi} &= N'_{hi} \exp(-C_{hi}N'_{hi}/Q_{hi}) \end{aligned} \right\} \quad (3)$$

The first equation accounts for nitrogen introduced in reach (h,i) , proportional to baseflow increase G_{hi} , and from the tributaries. The second accounts for denitrification. Here C_{hi} is the channel area (reach length \times width), Q_{hi} the mean annual flow and g the parameter, assumed common to all reaches, to which the sensitivity of N at the outlet to the dam is required. Differentiating (3),

$$\left. \begin{aligned} \frac{\partial N'_{hi}}{\partial g} &= G_{hi} + \frac{\partial N_{h-1,2i-1}}{\partial g} + \frac{\partial N_{h-1,2i}}{\partial g} \\ \frac{\partial N_{hi}}{\partial g} &= \left(1 - \frac{C_{hi}N'_{hi}}{Q_{hi}}\right) \frac{\partial N'_{hi}}{\partial g} \exp\left(-\frac{C_{hi}N'_{hi}}{Q_{hi}}\right) \\ &= \left(\frac{1}{N'_{hi}} - \frac{C_{hi}}{Q_{hi}}\right) N_{hi} \frac{\partial N'_{hi}}{\partial g} \end{aligned} \right\} \quad (4)$$

so

$$\frac{\partial N_{hi}}{\partial g} = \left(\frac{1}{gG_{hi} + N_{h-1,2i-1} + N_{h-1,2i}} - \frac{C_{hi}}{Q_{hi}} \right) N_{hi} \left(G_{hi} + \frac{\partial N_{h-1,2i-1}}{\partial g} + \frac{\partial N_{h-1,2i}}{\partial g} \right) \quad (5)$$

$$S_g^{N_{hi}} = \left(\frac{1}{gG_{hi} + N_{h-1,2i-1} + N_{h-1,2i}} - \frac{C_{hi}}{Q_{hi}} \right) \cdot (gG_{hi} + N_{h-1,2i-1} S_g^{N_{h-1,2i-1}} + N_{h-1,2i} S_g^{N_{h-1,2i}}) \quad (6)$$

Here $\partial N_{hi} / \partial g$ and $S_g^{N_{hi}}$ are more complicated functions of g than appears at first sight, as all the N 's depend on g .

These expressions indicate that:

(i) the recursion (6) can be used to find all the sensitivities to g , starting at the top of the catchment, after a *single* run to get the nominal values of all G 's, Q 's and N 's. Generally, all first-order sensitivities to any one parameter at any operating point can be generated (exactly but for finite precision) by one simulation run and one run of (6), so long as the derivatives exist;

(ii) both (5) and (6) are inconsistent with an assumption that N is a finite-degree polynomial in g : after rationalisation, it is not possible to match coefficients of all powers of g on the two sides. This is not surprising, as the denitrification equation in (3) is of infinite degree in g ;

(iii) if the exponential in (3) is not far below unity (i.e. if denitrification is by a small percentage), (3) can be approximated by $N_{hi} \cong N'_{hi}(1 - C_{hi}N'_{hi}/Q_{hi})$, then substituting into (5) and equating highest-degree terms in g on each side, the highest (significant) degree in g in N_{hi} is found to be twice the higher of the highest degrees in $N_{h-1,2i-1}$ and $N_{h-1,2i}$. However, g is typically small and low-degree terms dominate;

(iv) in (5), the contributions of reach (h,i) and the immediately upstream tributaries to $\partial N_{hi} / \partial g$, by G_{hi} , $\partial N_{h-1,2i-1} / \partial g$ and $\partial N_{h-1,2i} / \partial g$, are additive and equally weighted, so after running (5) it is easy to see the relative importance of each source in each reach.

To illustrate, a nominal run followed by recursive solution of (5) and (6) gives the results shown in Table 1 for the lower ends of one reach and its tributaries. Finite-difference results from a 10% perturbation of g are also shown.

Table 1. Analytical and perturbation results relating to reach (5,1).

Reach	N	$\partial N / \partial g$	$\delta N / \delta g$	S_g^N
5,1	0.4905	192.3	167.3	0.1960
4,1	0.2181	94.47	91.74	0.2165
4,2	0.0185	-108.5	-95.71	-2.931

$$G_{51} = 1590, gG_{51} = 0.7949, C_{51} / \bar{Q}_{51} = 0.7207$$

Several points emerge:

- G_{51} heavily dominates $\partial N_{41} / \partial g$ and $\partial N_{42} / \partial g$ in determining $\partial N_{51} / \partial g$ in (5), and the same is true in (6), determining $S_g^{N_{51}}$
- the derivatives and sensitivities from the perturbation test differ noticeably from those from (5) and (6), because of non-linearity
- one derivative is negative, indicating that in one or more higher reaches, the effect of g on the exponent in the denitrification equation dominates its effect in increasing N' .

An important but less than obvious point revealed by comparing analytical and perturbation results is that ill conditioning can reduce the accuracy of computed normalised sensitivities. Figure 3 shows the analytical and perturbation-derived sensitivities S_g^N for all reaches, ordered by N . The discrepancies are due partly to non-linearity, but depend on N , being much larger at very low N because of rounding. Large proportional error at very low values of N is, of course, unlikely to be serious.

A recursion for $\partial^2 N_{hi} / \partial g^2$ is easily derived but algebraically complex enough not to yield easy conclusions about sensitivity.

3.3 SA of effective-rainfall / runoff model

By contrast to the submodel above, the hydrological part has several (7) parameters and is not cascaded, as runoff is found for each subcatchment and for the catchment as a whole by applying a single IHACRES model calibrated for whichever area is represented. The part of the submodel relating flow to effective rainfall is straightforward to analyse. It can be written as the temporal recursion

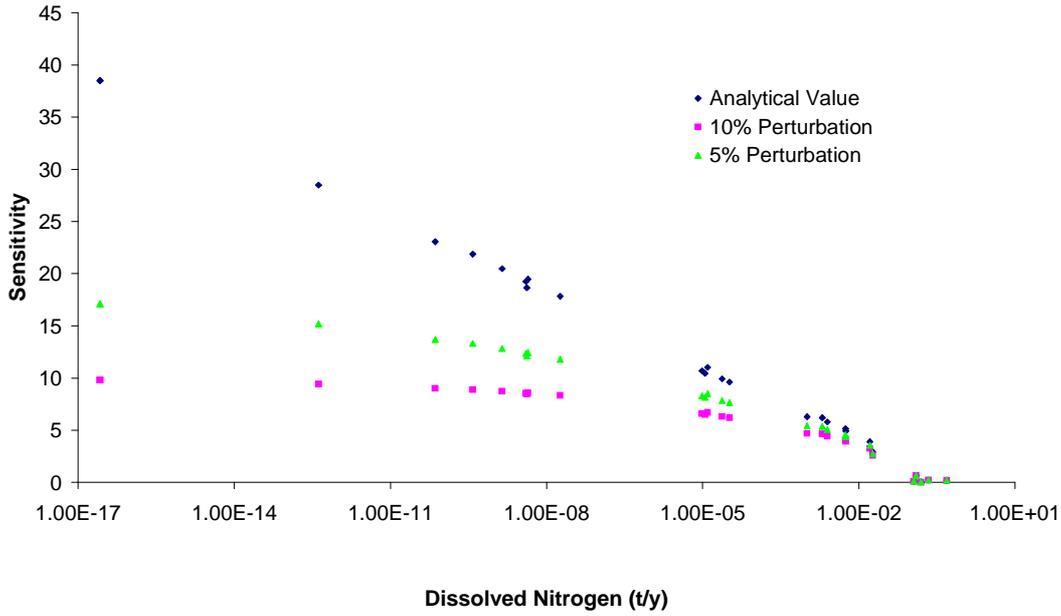


Figure 3. Comparison of analytical S_g^N and S_g^N by perturbation (5 and 10%), with N . Note log scale on horizontal axis and that absolute sensitivity values are shown.

$$Q_k = -a_1 Q_{k-1} - a_2 Q_{k-2} + b_0 E_k + b_1 E_{k-1} \quad (7)$$

where E_k is effective rainfall in day k and Q_k flow at the end of day k . With the parameter vector defined as $\theta \equiv [a_1 \ a_2 \ b_0 \ b_1]^T$, the vector influence function $\partial Q / \partial \theta$ is given by

$$\frac{\partial Q_k}{\partial \theta} = \begin{bmatrix} -Q_{k-1} - a_1 \frac{\partial Q_{k-1}}{\partial a_1} - a_2 \frac{\partial Q_{k-2}}{\partial a_1} \\ -a_1 \frac{\partial Q_{k-1}}{\partial a_2} - Q_{k-2} - a_2 \frac{\partial Q_{k-2}}{\partial a_2} \\ -a_1 \frac{\partial Q_{k-1}}{\partial b_0} - a_2 \frac{\partial Q_{k-2}}{\partial b_0} + E_k \\ -a_1 \frac{\partial Q_{k-1}}{\partial b_1} - a_2 \frac{\partial Q_{k-2}}{\partial b_1} + E_{k-1} \end{bmatrix} \quad (8)$$

and it is easy to see that the sequences $\{\partial Q / \partial a_1\}$, $\{\partial Q / \partial a_2\}$, $\{\partial Q / \partial b_1\}$ and $\{\partial Q / \partial b_0\}$ are all outputs of the same dynamical process, driven respectively by $\{-Q\}$, $\{-Q\}$ delayed by one day, $\{E\}$ and $\{E\}$ delayed by one day. The time constants of the dynamical process are the quick- and slow-flow time constants of the rainfall-runoff relation. Once the effects of differences in

initial conditions have faded, $\{\partial Q / \partial a_2\}$ is essentially $\{\partial Q / \partial a_1\}$ delayed by one day, and similarly for $\{\partial Q / \partial b_1\}$ and $\{\partial Q / \partial b_0\}$, so the parameter sensitivities of the mean flow over a year are

$$\frac{\partial \bar{Q}}{\partial a_1} = \frac{1}{365} \sum_{k=1}^{365} \frac{\partial Q_k}{\partial a_1} \cong \frac{1}{365} \sum_{k=1}^{365} \frac{\partial Q_k}{\partial a_2} = \frac{\partial \bar{Q}}{\partial a_2}; \quad (9)$$

$$\frac{\partial \bar{Q}}{\partial b_0} \cong \frac{\partial \bar{Q}}{\partial b_1}$$

It is also not difficult to see that, with the time constants very much less than a year,

$$\left. \begin{aligned} \frac{\partial \bar{Q}}{\partial a_1} &\cong -\bar{Q} - a_1 \frac{\partial \bar{Q}}{\partial a_1} - a_2 \frac{\partial \bar{Q}}{\partial a_1} \\ \text{so } \frac{\partial \bar{Q}}{\partial a_1} &\cong -\frac{\bar{Q}}{1 + a_1 + a_2} \cong \frac{\partial \bar{Q}}{\partial a_2} \\ \frac{\partial \bar{Q}}{\partial b_0} &\cong \frac{\partial \bar{Q}}{\partial b_1} \cong \frac{\bar{E}}{1 + a_1 + a_2} \end{aligned} \right\} \quad (10)$$

so these sensitivities can be found without performing a run.

4. DISCUSSION AND CONCLUSIONS

This paper has described SA on the catchment-scale integrated hydrologic and water-quality model, CatchMODS. The analysis is an important step in model development and has been very useful for improving understanding of the behaviour of the model particularly with respect to cascading sensitivity effects. It has contributed to improving management outcomes by developing techniques to identify significant sources of uncertainty in model predictions i.e. where uncertainty in inputs have greatest impact on model prediction.

The results presented here also illustrate several general factors in the analysis of complex and/or cascading environmental models. Cascading makes the overall effect of even very simple non-linearities on sensitivity difficult to assess without either an algebraic analysis or considerable experimentation. Algebraic analysis plus a very modest amount of computing can yield a good deal of insight not easily obtainable by experiment. Rounding errors can give rise to significant errors in the estimation of sensitivities. In sensitivity-propagation recursions such as (5) and (6), ill-conditioning may arise (and indeed sometimes does in the Ben Chifley catchment) when the contributing terms are individually not small. As seen in the effective-rainfall/runoff submodel, sensitivity analysis is straightforward when the recursion is linear.

SA is necessary to support the decision analysis framework for the Ben Chifley Dam catchment developed in parallel with the construction of CatchMODS. The aim of the decision analysis framework is to incorporate a broader view into evaluation of the performance of various management options to reduce surface-water pollution. A preliminary description of the decision analysis framework, which tries to reconcile the ecological and economic effects of remediation actions using multicriteria decision analysis, is available (Myšiak *et al.* 2004). Our intention is to investigate further the influence of model-input uncertainties in this framework on potential management recommendations.

CatchMODS includes refinements to the SedNet sediment-transport model described in Newham *et al.* (2003). SA of both models has had a role in the iterative process of model development and testing, providing insight into the overall effects of components of the models and clarifying their relative importance and their interactions. Difficulties exist in communicating the need for, and techniques of, SA to end-users especially non-technical managers. These difficulties present possibly greater limitations than SA techniques.

As part of the continuing process of SA and continued model development, more complete SA is recommended for the CatchMODS model. This might include SA across the multiple components of the model to determine the effects of parameter interactions, using Monte-Carlo sampling techniques such as Fourier Amplitude Sensitivity Testing (Saltelli *et al.*, 2000) as necessary. More fundamental SA, using algebraic analysis where possible, is also planned to determine the effects of spatially local variations in parameter values. Such investigations may allow alternative model structures, providing adequate resolution at minimum computational cost, to be identified.

5. REFERENCES

- Croke, B.F.W. and A.J. Jakeman, A Catchment Moisture Deficit Module for the IHACRES Rainfall-Runoff Model, *Environmental Modelling and Software*, vol. 19, 1, pp. 1-5, 2004.
- Jakeman, A.J., Littlewood, I.G. and P. G. Whitehead, Computation of the Instantaneous Unit Hydrograph and Identifiable Component Flows with Application to Two Small Upland Catchments, *Journal of Hydrology*, vol. 117, pp. 275-300, 1990.
- Myšiak, J., Newham, L.T.H and R.A. Letcher, Integrated Modelling and Decision Making Analysis for Water Quality Management: Ben Chifley Dam Catchment Case Study, in *Proceedings of Integrated Water Management of Transboundary Catchments: A Contribution form Transcat*, Palazzo Zorzi, Venice, Italy, 24-26 March 2004.
- Newham, L.T.H., Letcher, R.A., Jakeman, A.J. and T. Kobayashi, A Framework for Integrated Hydrologic, Sediment and Nutrient Export Modelling for Catchment-Scale Management, *Environmental Modelling and Software*, 2004, in press.
- Newham, L.T.H., Norton, J.P., Prosser, I.P., Croke, B.F.W. and A.J. Jakeman, Sensitivity Analysis for Assessing the Behaviour of a Landscape-Based Sediment Source and Transport Model, *Environmental Modelling and Software*, vol. 18, pp. 741-751, 2003.
- Prosser, I.P., Rustomji, P., Young, W.J., Moran, C. and A. Hughes, Constructing River Basin Sediment Budgets for the National Land and Water Resources Audit, *CSIRO Technical Report 15/01*, CSIRO Land and Water, Canberra, 2001.
- Saltelli, A., Chan, K. and E. M. Scott (Eds.), *Sensitivity Analysis*, Wiley, Chichester, 2000.