Sensitivity-Assessment Needs of Complex Simulation Models for Integrated Catchment Management

J.P. Norton[a], **R.Nathan**[b], **G. Podger**[c,d] **and R. Vertessy**[c,e]

a Integrated Catchment Assessment & Management, ANU, Canberra, Australia: j.p.norton@bham.ac.uk

b Sinclair Knight Merz, Malvern, Victoria, Australia c CRC for Catchment Hydrology

d NSW Dept. of Land & Water Conservation, Sydney, Australia

e CSIRO Land & Water, Canberra, Australia

Abstract: The Australian CRC for Catchment Hydrology has developed a software environment for integrating models of various aspects of catchments. Effective use of the resulting complex models requires knowledge of the sensitivity of their outcomes to variations in parameters or inputs, and hence to the underlying data and assumptions. Sensitivity assessment (SA) for such models makes demands beyond the capability of conventional techniques. A new SA approach devised in iCAM, ANU translates requirements on model outputs into constraints on input or parameter values. This allows assessment of the parameter ranges yielding outputs leading to the same management decision. SA tools employing this approach must take account of the needs of modellers, model users, catchment managers and other stakeholders. A review of the SA needs of integrated simulation models for catchment management has been carried out within the CRC's programme, with the new approach in mind. Its motivation, scope and conclusions are summarised.

Keywords: *Hydrology; Integrated catchment management; Modelling; Sensitivity assessment;*

1. INTRODUCTION

The Australian Cooperative Research Centre (CRC) for Catchment Hydrology has developed a software environment (TIME: The Invisible Modelling Environment) allowing integration of models of catchment attributes such as rainfall-runoff processes, irrigation, land use, ecology and economics. Integrated models to assist environmental management, e.g. of water use in a catchment, are also appearing elsewhere. Their effective use requires examination of sensitivity of the modelled outcomes to uncertainty or variations in parameters, inputs and assumptions. Sensitivity assessment (SA) for such models makes demands beyond the capability of conventional techniques. A new approach devised in the Integrated Catchment Assessment and Management group at ANU treats SA as translation of a set of requirements on model outputs (e.g. that they fall within given ranges) into bounds on the input and/or parameters. This can indicate the range of parameter values yielding outputs leading to the same management decision. The adequacy of parameter estimates can be gauged by comparing their estimated uncertainty with those ranges. The most critical parameter combinations are those with the narrowest bounds, while combinations with large ranges are candidates for elimination.

Computing the parameter bounds for a large model will be expensive, so we must focus on the simplest tools able to meet the SA needs of modellers, model users, catchment managers and other stakeholders. The success of SA also depends on being able to accommodate model-output requirements arising in practice and to present conclusions in acceptable forms. SA needs for integrated simulation models have been reviewed as part of the CRC's programme.

Section 2 discusses the uses of SA and outlines the proposed approach, which underlies some of the review questions. Reasons for preferring this approach to others are summarised. The review questions and answers are given in Section 3, which also lists the conclusions.

2. USES OF SENSITIVITY ASSESSMENT AND PROPOSED NEW APPROACH

2.1. Potential uses of sensitivity assessment

SA is often equated with probabilistic uncertainty analysis, relating distributions of model inputs or parameters and the resulting output distributions (Saltelli *et al*., 2000). However, SA may be entirely deterministic if the model equations and inputs are. A working definition of SA is "*examination of the relations between variations of the inputs and/or parameters of a model and the resulting variations of selected outputs*". Only if inputs or outputs must be characterised by ensemble properties (*e.g*. climatic variables) need SA be probabilistic. Otherwise, responses to deterministic change are individually meaningful. The review viewed SA as deterministic, but not excluding Monte Carlo trials or random search to explore sensitivities.

The above definition of SA says nothing about uncertainty, although input-to-output mapping of uncertainty is one motive for SA. The view taken here is that SA and uncertainty specification are distinct, although the nature of each affects how the other must be done in uncertainty analysis.

An important motive for SA is to focus attention on critical parts of the model. Even with physical insight, understanding of submodels in an integrated model does not ensure insight into the overall behaviour. Local responses may have little overall significance or be crucial, according to the properties of the rest of the model. Environmental processes are often low-pass, so the output from a series of submodels may be insensitive to error or uncertainty in rapid components of a variable further up the chain, while accumulation of error or uncertainty in slower components may cause large output variation. Feedback may strongly modify model behaviour. In stable systems, it typically reduces the sensitivity of overall behaviour to variations or disturbances in the forward path, so high input-output sensitivity in a submodel may have little overall effect. By contrast, a small change in a submodel or in the strength or speed of feedback may destabilize a marginally stable system such as a stressed ecosystem. Systematic SA for the overall model is thus essential even if its submodels are well understood, to take interactions between submodels into account.

In models with many similar sections, SA may point to a few crucial parameters or points in space and time. It may identify parameters determining dominant output modes, or points with bad combinations of output variables.

SA may reveal potential for model refinement. Large simulation models may well contain near-redundancies even if time and space intervals look suitable, since complex local behaviour may aggregate into simple behaviour overall. Removal of redundancy makes the model faster, easier to test and comprehend, and less prone to ill conditioning due to too-fine sampling or poorly determined parameters. Model reduction can be viewed as a large perturbation, requiring SA to check that its effects are acceptable. Conversely, SA may uncover a need for finer resolution.

There are also unpredictable benefits from SA, such as better insight and discovery of data inconsistency.

These uses for SA suggest questions about its working context (Section 3) and lead to features of the novel approach to SA described next.

2.2. Proposed SA approach

Conventional deterministic SA employs perturbation testing to estimate parameter-to-output derivatives (gradient and Hessian), or Monte Carlo (MC) trials. For time-series outputs, influence functions express parameter-to-output sensitivity, while distributed outputs need spatial influence functions. Second derivatives show the effects of interacting pairs of parameters and approximately quadratic non-linear dependence of output on each parameter. Computed derivatives have severe limitations: they are generally valid only for variations comparable to the perturbations; perturbation sizes are difficult to select without prior knowledge of sensitivities, risking sparse or too-narrow coverage; derivatives are generally functions of state, often high-dimensional, expensive to compute and hard to comprehend; and they require outputs to be measured on a continuous scale, whereas model outcomes may occupy discrete categories, *e.g*. acceptable/not or classified according to rankings on conflicting objectives.

The traditional alternative, MC trials to get points on the parameter-to-output map, does not rely on linearity. It shares the difficulty of choosing a range to cover, and sparseness of coverage by a given number of trials increases rapidly with the number of parameters. A less obvious problem is overparameterization, causing parameter values which yield output behaviour within the range of interest to have a large spread in some directions in parameter space but a small spread in others, *i.e*. to be almost in a subspace. Only a small proportion of trials then give output behaviour in the range of interest; most are wasted.

These limitations have led to a new approach, relying less on derivatives and MC trials. It formulates SA as translating a list of model-output requirements into requirements on the inputs and/or parameters. For example, we might ask over what range of uncertain parameter values each output changes by less than a given percentage. This amounts to specifying a set of equivalent model outcomes, the *equivalence set*, then computing the corresponding *feasible set* of parameter values. That is, SA is viewed as set-to-set inverse mapping from the equivalence set back through the model into the feasible set. For example, we might find the set of parameter values in an IHACRES rainfall-runoff model to meet given bounds on water availability above some flow threshold over a given period. Set inversion is also the basis of state bounding and parameter bounding (Walter, 1990; Norton, 1994,1995), but inversion through non-linear models has had little attention (Jaulin and Walter, 1993; Novara and Milanese, 2000).

The use of sets in SA confers great flexibility. Outcomes need not even be quantifiable. If they are numerical values of outputs, they define the equivalence set through inequality constraints. An equivalence set arises naturally whenever all outcomes meeting some management criteria are equally acceptable. Translation from equivalence set to feasible set involves either running the model repeatedly, discovering the input/parameter range by trial and error, or inverting the model to determine the feasible set directly from the output requirements. This is more efficient but harder, as outlined below. In practice, both forward runs and model inversion are likely to be needed, for economy and accuracy. Set-to-set SA cannot avoid all the difficulties of other approaches. Scales and resolution for exploring the sets must still be found, subject to load. Tracing output constraints back through the model is liable to ill conditioning, as a low-pass model has a high-pass inverse, amplifying wideband noise such as rounding error. In some (not all) cases, inverting a stable model produces an unstable model; time-reversal always does.

The following questions for the review of SA needs were framed with the set-to-set approach in mind. However, many of the conclusions apply whatever the approach.

3. QUESTIONS AND RESPONSES

To get wide coverage in a short time, the review was carried out by discussion among a number of experienced developers and users of integrated models. It is thus subjective and perhaps partial, but is believed to be reasonably representative of uses and users of environmental simulation models. For best use of space, the conclusions are included in this section, italicised, rather than reiterating them in a Conclusions section.

3.1. How are models used in practice?

The primary motives for running simulation models put differing emphases on SA. *For prediction, detailed SA is needed but only for a few scenarios, and perhaps only for outcomes showing significant inter-scenario differences.*

The review recognised a frequent need for improved insight, typically constrained to answer immediate management questions: *better insight into very specific aspects of model behaviour is a top-priority motive for SA; improving broader insight can't be ignored but has lower priority.*

Attention-focusing was also felt to be an important use of SA, requiring initial examination of inputs, outputs and internal state variables in many operating conditions. However, *useful insight may still be gained from limited SA, with input and parameter values and sensitivities known very approximately*. As an example, a relatively crude sensitivity assessment reveals that only three or four of 15 parameters are critical in a specific application of the CSIRO SedNet sedimentation model (Newham *et al*., 2003), and their physical interpretation suggests that this conclusion would remain valid more widely.

SA for points where critical output behaviour occurs, *e.g.* locations with highest salinity under irrigation, requires low sensitivity to dubious data for the pattern of relative output values, but not necessarily high accuracy at each point. The review concluded that *SA in order to identify and focus attention on critical times and places deserves high priority.* Typically, one might be interested in the relative influences of a number of parameters on output variability over an area, and more specifically at the worst spots. The effects of a few dominant parameters at those few places would then be explored in detail. Hence *SA tools should be designed for iterative use, starting with a broad investigation, gradually concentrated into a narrowly focused, more precise examination.*

Discrete, spatially distributed parameters or inputs, e.g. crop types, need care in distinguishing single- from multi-parameter SA. One qualitative parameter (crop type) is varied, but computationally there are as many parameters as subareas. Each subarea contributes numbers (e.g. interception and transpiration coefficients), defining a discrete value of a parameter vector. Thus *SA must handle changes in discrete values of input-parameter vectors at multiple locations.*

The potential of SA in helping choice between a few types and structures of model was thought to be of major interest. Set-to-set SA explicitly lists the output requirements, preventing vagueness or ambiguity in model-performance criteria.

SA of an entire model for model reduction, testing all plausible simplifications against all performance criteria, is demanding. Piecemeal testing, as in submodel development, reduces the load. However, the fact that whole-model behaviour is not always readily predicted from submodel behaviour points to *another mode of operation of SA: examining overall input-output effects of parameters in one submodel at a time*, to see the overall influence of local changes and hence find where simplification is permissible.

3.2. Who uses models?

Propective SA users for complex environmental models include modellers, experienced users (consultants, State agencies, R&D organisations, academics), catchment managers and local stakeholders (farmers, environmental interest groups, local councils). These groups differ widely in appreciation of the capabilities and limitations of models, technical background, local knowledge and breadth of concerns, and make very different demands on models and SA. It was stressed that catchment managers seldom wish to run simulation models, preferring to work through modellers and consultants. It was also clear that ultimate stakeholders in model outcomes cannot be expected to be directly involved in running models or in SA. *Running simulations and performing SA are the province of modellers, consultants and expert practitioners in bodies such as State government departments*. They are mostly well equipped to assimilate a new SA approach. However, some may be accustomed to a probabilistic approach to uncertainty analysis and reluctant to accept set-to-set SA. *The new SA approach must be well publicised as it develops, to gain familiarity among prospective users. Incorporation of distributional information into set-to-set SA* (discussed later) is also important.

The gulf between what analysis techniques for large models or data sets are on offer and what users need was emphasised; analysis problems tend to be oversimplified to allow neat solutions, ignoring practical constraints and complications. Lack of useable analysis often results in misuse or indiscriminate use of elaborate models. This underlines *an urgent need to provide more flexible, easily applied SA methods, able to inform users about the reliability of model results.*

A strongly exppressed opinion was that *exploratory SA tools, useable without extensive training, are needed, not a comprehensive, integrated package*. In an academic environment, there are additional reasons for developing software as self-contained items: software engineering resources are scarce, short contracts make continuity problematical, and informal software exchange and modification make version control, maintenance and attribution difficult.

The inertia, conservatism, caution or common sense of stakeholders was stressed, implying that *for new SA tools to be accepted, they must be expressible in non-technical terms, and must provide reliability information for the stakeholder to balance predicted benefit and risk of an action.* If the SA process is not so transparent as to engender trust, it must give managers the means to convey a fair impression of model reliability.

3.3. What is the likely scope of an SA query?

Specifically, what information does an SA user want; are many uncertain items likely to be examined at once; how complex is each likely to be (*e.g.* many points in space or time at once, or one at a time); are only inputs and outputs to be examined, or state also [a state variable in a simulation run being invisible unless also an output. Traditional SA does not find input-state and state-output sensitivities]; are sensitivities to initial or other boundary conditions needed?

The review stressed that the scope of SA is limited by inability to model, or even understand, behaviour in detail. An example is groundwater modelling, with limited access through bores and the system often geologically heterogeneous. Thus *aggregation is often forced on SA, and further aggregation may be necessary to present SA reults in a readily understood form*.

Other significant points made were that (i) each SA investigation is likely to concern only a few parameters and simple output metrics, but they may summarise complex spatio-temporal behaviour; (ii) small output changes are the norm in management, with little risk of large-scale changes (with exceptions, *e.g.* potential ecological disasters); (iii) thresholds affecting outputs may make input-output linearity a poor approximation even for small changes, so *SA should show how far the input-output relations, or input-state and* *state-output relations, deviate from linearity*; (iv) *small output changes predominate but there may be high uncertainty in parts of the model*; (v) degree of detail in models and SA may be severely limited by variability of the variables and by data availability, a good example being nutrient modelling; (vi) *the potential of model inversion in helping a manager determine where to intervene is a major benefit of the proposed SA approach. Such intervention is often about land use, where set-to-set SA must deal with spatial outcomes*; (vii) environmental information gets thinner as one goes back in time, and models often employ time and space intervals unjustified by the data. *SA should provide guidance on choice of spatio-temporal intervals and record lengths.* SA might borrow heuristics from parameter estimation, for models close enough to linearity to have distinguishable dynamical modes. For parameter estimation in linear models, relations between record lengths, sampling rates, time constants and signal:noise ratios are well established. In stiff systems, analysis may be split into two stages, investigating fast dynamics while treating the rest as drift, and separately the slower dynamics while treating the rest as instantaneous.

Opinions varied on how far SA need consider state. Model builders consider state, but non-modellers will not often want to delve inside a model. It is not immediately obvious that SA must consider state; running a model for a range of input/parameter values, recording outputs but not state, might seem to be enough. This is not so, as *the output depends not only on the preceding input but also on the initial, and hence current, state*. Consider a linear rainfall-runoff model feeding a finite storage. At onset of overflow, the relation between rainfall and flow from storage changes sharply with the volume stored, a state variable. Even in linear models, initial-condition response may be non-negligible. If a series of simulation runs can give the input-output relations and their sensitivities for the entire realistic range of antecedent state and any other relevant boundary conditions, state is just an invisible intermediary between inputs and outputs. Even then, it may be worth looking at state in SA, as models are often much simpler (in structure, not number of variables) in state-space form than in input-output form. In non-linear models, a further reason is the need to explore state space well enough to avoid invalid generalization.

The implications are *(i) SA must give experienced practitioners the option of expressing the model in state-space form; (ii) SA must provide choice of how far to consider state, and should allow trial-and-error choice of which state variables to worry about; (iii) for non-expert SA consumers, the interface should conceal the state unless asked to show it; (iv) SA should give guidance on what aggregation is permissible in presenting outcomes of a detailed spatio-temporal model.*

Responses about whether boundary conditions must be included as parameters in SA were: (i) initial conditions are usually known quite well and do not require SA (subject to the comments above about initial state); (ii) *other boundary conditions <u>are</u> an issue for SA*, e.g. groundwater distribution or areal rainfall pattern (which might be treated as forcing). That said, *the larger the spatio-temporal scale, the less SA need analyse the influence of detailed boundary conditions, since more averaging is acceptable.*

3.4. How is the equivalence set to be specified?

Model-output requirements are to be expressed as bounds on measures of output spatio-temporal behaviour, e.g. water available for irrigation in a given period. Bounds on

- instantaneous values of variables
- integrals of variables (*e.g.* annual income)
- variation (*e.g.* environmental flows)
- distribution (*e.g.* proportion of river-flow values above a threshold), and
- the rank in an ordering (*e.g.* of economic outcomes in alternative scenarios)

had been foreseen. *The review added bounds on*

- *the frequency of events, e.g. wetting of trees*
- *both the size of events and the intervals between them, e.g. overbank flows*
- *measures of cumulative impact*
- *weighted spatial outcomes, e.g.* salinity weighted by local seriousness, water quality weighted at extraction points, different weights for Crown or private land.

This diversity of types makes heavy demands on the mechanics of model inversion, but all are readily handled in forward runs, by simply checking outcomes against their bounds.

3.5. How should SA results be presented?

This asks what information SA should produce,in what form, not just what the user interface is. Hard-to-meet user demands must be recognised early in SA design. It was suggested that *users need time series, cumulative distributions* (*e.g.* flow-duration curves) *and two-dimensional pictures*. Plots of residuals against input could be extended to show residuals moving with parameter values. *There is scope for presenting some higher-dimensional results, e.g.* "spider plots" showing changes in a two-variable relation as a parameter is varied in steps, or Andrews curves (Andrews, 1972) comparing parameter vectors in four or more dimensions. *Visualisation methods exist for up to about 6 dimensions but are hard to interpret and not thought useful.*

These conclusions influence what features of the feasible set are extracted. The set is an object, perhaps complicated, with as many dimensions as free parameters. In three or more, it is difficult to visualise unless approximated by a simple shape, *e.g.* a box (bounds on individual parameters). *It is proposed that feasible sets for three or more inputs or parameters be presented as two-dimensional cross sections or one-dimensional features* such as largest and smallest diameters. A set of feasible trial points forms a large multivariate data set from which principal-component analysis (PCA) can extract some one-dimensional features. PCA is too limited for SA, but the idea of *extracting extents in significant directions is applicable to feasible sets.*

Assessment of extremes (*e.g.* salinity hot spots) poses the danger of assuming that extremes translate to extremes, implying that the feasible-set boundary is determined by the equivalence–set boundary and *vice versa.* Input-output or parameter-output relations may well be non-monotonic. If tail histogram bins (top and bottom segments of cumulative distribution functions) of measured outputs are examined, combinations of non-extreme input values are often responsible for them Thus *one role of SA is to enquire into monotonicity of input/parameter-output relations, and to identify what causes extreme outputs*.

3.6. What alternatives for defining input and output behaviour should be considered?

Set-to-set SA avoids hard-to-test distributional assumptions, nor does it require all inputs, parameters and outputs to be quantifiable and continuous. These advantages are not confined to plain sets; graded deterministic entities might be used. Plain sets label values only as in or out of the set. There is no indication of the margins by which a feasible value meets the output requirements, or how far an unfeasible value is from meeting them. Fuzzy sets (Zadeh, 1965) have for each relevant set a membership degree between zero and 1, assigned to every variable or parameter value by expert knowledge. *Fuzzy sets were felt to be unappealing* because of their subjectivity and the many parameters required to specify membership functions. Rough sets (Pawlak, 1982) classify values as in the set, not in it or "don't know". *The consensus was that use of rough sets in SA should be kept open as a possibility*. They may be computable by interval analysis (Jaulin and Walter, 1993) or combining outer-bounding and inner-bounding feasible sets.

A third possibility is to specify two or more pairs of bounds on each output, defining a coarse histogram, *e.g.* wide bounds covering all credible or acceptable values and narrow ones for the great majority. Each is mapped to the parameters/input.

3.7. SA to help scaling and regionalisation?

A submodel at one spatial or temporal scale may be integrated into a model at another, so SA must consider rescaling. The appeal of physically interpretable parameters in guiding rescaling was stressed, even at the expense of parsimony. The point was made that dependence on the operating point (state) and parameter values makes the admissibility of rescaling hard to assess. *Investigation of the sensitivity of scaled results to varying state, parameters or parameterisation is therefore a significant potential use of SA.*

4. REFERENCES

Andrews, D.F., Plots of high-dimensional data, *Biometrics*, 28, 125-136, 1972.

Jaulin L. and E. Walter, Set inversion via interval analysis for nonlinear bounded-error estimation, *Automatica,*29(4), 1053-1064, 1993.

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, *Int. Congr. on Modelling & Simulation MODSIM2003*, Townsville, Qld, 2003.

Norton, J.P. (Ed.) Special issues of *Int. J. of Adaptive Control & Signal Processing*, 8(1), 1-118, 1994 and 9(1), 1-134, 1995.

Novara, C. and M. Milanese, Set membership identification of nonlinear systems, *Proc. 39th IEEE Conf. On Decision & Control*, Sydney, 2831-2836, 2000.

Pawlak, Z., Rough sets*, lnt.J. Comput. Inf. Sci.* 11, 341-356, 1982.

Saltelli, A., Chan, K. and E. M. Scott (Eds.), *Sensitivity Analysis*, Wiley, Chichester, 2000.

Zadeh, L. A., Fuzzy sets, *Inf. Control* 8, 338-353. 1965.