



Jul 1st, 12:00 AM

Ten Iterative Steps In Development and Evaluation of Environmental Models

A. J. Jakeman

R. A. Letcher

J. P. Norton

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Jakeman, A. J.; Letcher, R. A.; and Norton, J. P., "Ten Iterative Steps In Development and Evaluation of Environmental Models" (2006). *International Congress on Environmental Modelling and Software*. 77. <https://scholarsarchive.byu.edu/iemssconference/2006/all/77>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.



ELSEVIER

SCIENCE @ DIRECT®

Environmental Modelling & Software xx (2006) 1–13

Environmental
Modelling & Softwarewww.elsevier.com/locate/envsoft

Ten iterative steps in development and evaluation of environmental models

A.J. Jakeman^{a,b,*}, R.A. Letcher^{a,c}, J.P. Norton^{a,c}

^a *Integrated Catchment Assessment and Management Centre, Building 48A, The Australian National University, Canberra, ACT 0200, Australia*

^b *Centre for Resource and Environmental Studies, The Australian National University, Canberra, ACT 0200, Australia*

^c *Department of Mathematics, The Australian National University, Canberra, ACT 0200, Australia*

Abstract

Models are increasingly being relied upon to inform and support natural resource management. They are incorporating an ever broader range of disciplines and now often confront people without strong quantitative or model-building backgrounds. These trends imply a need for wider awareness of what constitutes good model-development practice, including reporting of models to users and sceptical review of models by users. To this end the paper outlines ten basic steps of good, disciplined model practice. The aim is to develop purposeful, credible models from data and prior knowledge, in consort with end-users, with every stage open to critical review and revision. Best practice entails identifying clearly the clients and objectives of the modelling exercise; documenting the nature (quantity, quality, limitations) of the data used to construct and test the model; providing a strong rationale for the choice of model family and features (encompassing review of alternative approaches); justifying the techniques used to calibrate the model; serious analysis, testing and discussion of model performance; and making a resultant statement of model assumptions, utility, accuracy, limitations, and scope for improvement. In natural resource management applications, these steps will be a learning process, even a partnership, between model developers, clients and other interested parties.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Model testing; Verification; Uncertainty; Sensitivity; Integrated assessment; System identification

1. Motivation

The pursuit of good practice in model development and application deserves thorough and sustained attention, whatever the field. Good practice increases the credibility and impact of the information and insight that modelling aims to generate. It is crucial for model acceptance and is a necessity for long-term, systematic accrual of a good knowledge base for both science and decision-making. The complexity and uncertainty inherent in management for better sustainability outcomes make the pursuit of good practice especially important, in spite of limited time and resources. Natural resource management confronts a complex set of issues, usually with

environmental, social and economic trade-offs. These trade-offs are characterised by interactions at many scales and often by scarcity of good observed data. Thus natural resource managers commonly have to trade uncertain outcomes to achieve equitable results for various social groups, across spatial and temporal scales and across disciplinary boundaries. This must be achieved on the basis of information that varies in relevance, completeness and quality.

The complexity of these situations has led to model-based approaches for examining their components and interactions, and for predicting management outcomes. There is wide agreement on the potential of models for revealing the implications of assumptions, estimating the impact of interactions, changes and uncertainties on outcomes, and enhancing communication between researchers from different backgrounds and between researchers and the broader community.

Managers and interest groups can also potentially benefit from use of a model to define the scope of a problem, to

* Corresponding author. Integrated Catchment Assessment and Management Centre, Building 48A, The Australian National University, Canberra, ACT 0200, Australia.

E-mail address: tony.jakeman@anu.edu.au (A.J. Jakeman).

115 make assumptions explicit, to examine what is known and
 116 what is not, and to explore possible outcomes beyond the ob-
 117 vious ones. If models are accessible enough, they can act as
 118 a medium for wider participation in environmental manage-
 119 ment. However, the pressing need to use models in managing
 120 Footnote for first page of position paper:

121 Position papers aim to synthesise some key aspect of the
 122 knowledge platform for environmental modelling and software
 123 issues. The review process is twofold – a normal external re-
 124 view process followed by extensive review by EMS Board
 125 members. See the Editorial in this issue.

126 Complex situations, rather than in sharply defined areas of
 127 research, has resulted in people with little modelling or quan-
 128 titative background having to rely on models, while not being
 129 in a position to judge their quality or appropriateness. [Caminiti](#)
 130 (2004) provides a resource manager’s perspective on the diffi-
 131 culties of choosing the best modelling approach for catchment
 132 management, concluding that “[m]odellers can help by trying
 133 to understand the needs and expectations of the resource man-
 134 ager, who may not have the technical knowledge or language
 135 to express them.” Managers may also not initially understand
 136 their own needs fully, so modelling must be an iterative learn-
 137 ing process between modeller and manager.

138 The uses of models by managers and interest groups, as
 139 well as by modellers, bring dangers. It is easy for a poorly in-
 140 formed non-modeller to remain unaware of limitations, uncer-
 141 tainties, omissions and subjective choices in models. The risk
 142 is then that too much is read into the outputs and/or predictions
 143 of the model. There is also a danger that a model is used for
 144 purposes different from those intended, making invalid con-
 145 clusions very likely. Taking a longer-term perspective, such in-
 146 advertent abuses detract from and distort the understanding on
 147 which science and decision-making are built.

148 The only way to mitigate these risks is to generate wider
 149 awareness of what the whole modelling process entails, what
 150 choices are made, what constitutes good practice for testing
 151 and applying models, how the results of using models should
 152 be viewed, and what sorts of questions users should be asking
 153 of modellers. This amounts to specifying good model practice,
 154 in terms of development, reporting and critical review of
 155 models.

156 As a move in that direction, this paper outlines ten steps in
 157 model development, then discusses minimum standards for
 158 model development and reporting. The wide range of model
 159 types and potential applications makes such an enterprise
 160 prone to both over-generalisation and failure to cover all cases.
 161 So the intention is to name the main steps and give examples
 162 of what each includes, without attempting the impossible task
 163 of compiling a comprehensive checklist or map of the model-
 164 development process. Such checklists have been developed
 165 within certain modelling communities where particular para-
 166 digms are dominant. Thus the Good Modelling Practice Hand-
 167 book ([STOWA/RIZA, 1999](#)), financed by the Dutch
 168 government and executed by Wageningen University, has
 169 a well developed checklist for deterministic, numerical
 170 models. The guidelines for modelling groundwater flow devel-
 171 oped by the [Murray-Darling Basin Commission \(2000\)](#) in

172 Australia provide another example. Our purpose, by contrast,
 173 is to point to considerations and practices that apply in a broad
 174 range of natural resource modelling situations.

175 It is hoped that this paper will prompt modellers to codify
 176 their practices and to be more creative in their examination of
 177 alternatives and rigorous in their model testing. It is intended
 178 to provide a synoptic view for model builders and model users,
 179 applying to both integrated models and models within distinct
 180 disciplines. It does not deal with the surrounding issue of the
 181 appropriate development and use of environmental decision
 182 support systems (e.g. [Denzer, 2005](#)), which in addition involve
 183 issues of user interfacing, software usability and software and
 184 data integration. The paper discusses good practice in con-
 185 struction, testing and use of models, not in their imbedding
 186 and use in decision support systems or with software interfaces
 187 more widely.

188 As already indicated, the idea of guidelines for good model
 189 practice is not new. [Parker et al. \(2002\)](#) call for the develop-
 190 ment of guidelines for situations where formal analysis and
 191 testing of a model may be difficult or unfeasible. They state
 192 that “the essential, contemporary questions one would like
 193 to have answered when seeking to evaluate a model (are):
 194

- 195 i) Has the model been constructed of approved materials i.e.,
 196 approved constituent hypotheses (in scientific terms)?
- 197 ii) Does its behaviour approximate well that observed in re-
 198 spect of the real thing?
- 199 iii) Does it work i.e. does it fulfil its designated task, or serve
 200 its intended purpose?”

201 [Risbey et al. \(1996\)](#) call for the establishment of quality-
 202 control measures in the development of Integrated Assessment
 203 (IA) models for climate change, and suggest several features
 204 that must be considered:
 205

- 206 • a clear statement of assumptions and their implications;
- 207 • a review of ‘anchored’ or commonly accepted results and
 208 the assumptions that created them;
- 209 • transparent testing and reporting of the adequacy of the
 210 whole model, not only each of the component parts;
- 211 • inclusion of the broadest possible range of diverse per-
 212 spectives in IA development;
- 213 • supply of instructions to model end-users on the appropri-
 214 ate and inappropriate use of results and insights from the
 215 analysis;
- 216 • ‘A place for dirty laundry’, that is, for open discussion of
 217 problems experienced in constructing complex integrative
 218 modelling, in order for solutions to these problems to be
 219 found, and to facilitate the appropriate level of trust in
 220 model results.

221 [Ravetz \(1997\)](#), considering integrated models, argues for
 222 validation (or evaluation) of the process of development rather
 223 than the product, stating that in such circumstances “the inher-
 224 ently more difficult path of testing of the process may actually
 225 be more practical”. Ravetz finds that in general “the quality of
 226 a model is assured only by the quality of its production”.
 227
 228

229 However, he does not define the essential components or steps
230 in model development that would make up such a quality-as-
231 surance process, nor does he discuss how far the quality of
232 production can be assessed without assessing the quality of
233 the product.

234 Caminiti (2004) outlines a number of potential pitfalls in
235 using models for management, and proposes steps that re-
236 source managers should take to avoid them.

237 Refsgaard et al. (2005) address the issue of quality assur-
238 ance (QA), defined as protocols and guidelines to support
239 the proper application of models. They argue that “Model
240 credibility can be enhanced by a proper modeller-manager di-
241 alogue, rigorous validation tests against independent data, un-
242 certainty assessments, and peer reviews of a model at various
243 stages throughout its development.”

244 In promoting responsible and effective use of model infor-
245 mation in policy processes, Van der Sluijs et al. (2005) discuss
246 four case-study experiences with the NUSAP system for un-
247 certainty assessment. This system, due to Funtowicz and Rav-
248 etz (1990), offers analysis and diagnosis of uncertainty in the
249 knowledge base of complex policy problems. Van der Sluijs
250 et al. (2005) show that extending the scheme beyond main-
251 stream technical methods of sensitivity and uncertainty analy-
252 sis, by complementing it with qualitative approaches, further
253 promotes reflection and collective learning. Thus they cover
254 societal aspects such as differences in framing of the problem,
255 inadequacy of institutional arrangements at the science-policy
256 interface, and controversy.

257 These authors argue that good practice in the development
258 of integrated models is made all the more necessary by the in-
259 herent difficulties in validating them. As implied in the open-
260 ing paragraph, many disciplinary modelling studies lack
261 elements of good model practice, such as a clear statement
262 of modelling objectives, adequate setting out of model as-
263 sumptions and their implications, and reporting of model re-
264 sults, including validation/evaluation. Cross-disciplinary
265 models for influencing management should be tested against
266 additional criteria such as fitness for purpose, flexibility to re-
267 spond to changing management needs, and transparency so
268 that stakeholders can see how the results were derived.

270 2. Improving the modelling process

272 2.1. Introduction

274 Broad areas where better modelling practice can improve
275 models and their adoption are suggested below, before more
276 detailed discussion of ten steps in model development.

277 Wider and more strategic application of good models, com-
278 parison of models and associated long-term data acquisition
279 can assist not only in exploiting existing knowledge but also
280 in accruing new knowledge. An example is the current Predic-
281 tion in Ungauged Basins program of the International Associ-
282 ation of Hydrological Sciences. It has several groups, one the
283 Top-Down Working Group (<http://www.stars.net.au/tdwg/>).
284 The groups are tackling questions of how to predict streamflow
285 in ungauged catchments through systematic studies, typically

286 involving comparison of traditional and novel models and da-
287 taset benchmarking across a range of hydroclimatologies. The
288 Top-Down Working Group expects to improve understanding
289 of the drivers of catchment processes and how they relate to
290 fluxes from river basins. Its success will depend on attention
291 to the areas outlined below.

293 2.2. Proper definition of scope and objectives of the 294 model

296 In making a case for modelling to help managers respond to
297 a problem in natural resources, it is all too easy:

- 298 • to extend the scope beyond what is needed to answer the
- 299 questions at hand;
- 300 • to promise more than can be delivered in the time
- 301 available;
- 302 • to ignore or underestimate the difficulties and the limita-
- 303 tions in data and techniques;
- 304 • to oversimplify or overelaborate;
- 305 • to push a particular approach not well suited to the job;
- 306 • to rely too much on existing, familiar but less-than-ideal
- 307 models, and conversely;
- 308 • to overlook existing knowledge and previous experience;
- 309 • to take too little note of the need for consultation and
- 310 cooperation;
- 311 • to commit to a time scale preventing unforeseen factors
- 312 from being adequately dealt with, and, most crucially;
- 313 • to obfuscate the objectives, knowingly or inadvertently.
- 314
- 315

316 How often does one see objectives explicitly stated *and*
317 *iterated upon*? Refinement of an objective can lead to a simpler
318 task, as some factors are found to be unimportant, others crit-
319 ical, and the available information becomes clearer. Assess-
320 ment of uncertainty plays a crucial role in such refinement;
321 better a useful answer to a simple question than too uncertain
322 an answer to a more ambitious question.

324 2.3. Stakeholder participation in model development

325 Stakeholders comprise all those with an interest. For natural
326 resources, this is especially the managers and the various sec-
327 toral interests. Stakeholder participation is a key requirement
328 of good model development, particularly when models are to
329 address management questions. Aside from equity and justice,
330 there are two main reasons for increased stakeholder participa-
331 tion in model development. The first is to improve the model-
332 ler’s understanding, allowing a broader and more balanced
333 view of the management issue to be incorporated in the model.
334 The second is to improve adoption of results from the assess-
335 ment, increasing the likelihood of better outcomes, as model
336 development becomes an opportunity for stakeholders to learn
337 about interactions in their system and likely consequences
338 of their decisions. Both reasons work iteratively. That is, contin-
339 ued involvement is necessary because neither the modeller nor
340 the manager usually has a clear and comprehensive idea at the
341 outset of what the model must do.

343 Stakeholder participation in the past has often been limited
 344 to researchers wishing to exploit the results of the modelling
 345 exercise. A better approach, increasingly employed, is to in-
 346 volve all stakeholders throughout model development in a part-
 347 nership, actively seeking their feedback on assumptions and
 348 issues and exploiting the model results through feedback and
 349 agreed adoption. This approach is expensive in effort, time
 350 and resources, but the aim of modelling is often to achieve
 351 management change, and the learning process for modellers,
 352 managers and other stakeholders inherent in this approach is
 353 essential to achieving change. Examples of such participation
 354 in model development can be found in [Fath and Beck \(2005\)](#),
 355 [Hare et al. \(2003\)](#) and [Letcher and Jakeman \(2003\)](#). [Beck](#)
 356 [\(2005\)](#) “examines the implications of the ongoing shift –
 357 from the technocracy of the past century to the democracy
 358 of stakeholder participation in the present century – for the
 359 more widespread use of information and technologies in man-
 360 aging water quality in urban environments.” An excellent
 361 overview of participation as part of integrated assessment
 362 can be found in [Mostert \(in press\)](#).

364 2.4. Conceptualising the system

366 Consideration and justification of options in defining the
 367 system warrant attention by modellers and their clients.
 368 What to include and what not to incorporate in a modelling ac-
 369 tivity should be addressed explicitly at the outset and itera-
 370 tively revisited as far as resources allow. The system being
 371 modelled should be defined clearly, including its boundaries
 372 (e.g. physical, socioeconomic and institutional). Boundary
 373 conditions can then be modelled as constraints or as input sce-
 374 narios, whose values can be perturbed in line with stipulated
 375 assumptions.

377 2.5. Embracing alternative model families and structures

379 Comparisons between alternative model families and struc-
 380 tures are sometimes advocated (as above), but seldom per-
 381 formed systematically against specified criteria or, indeed, at
 382 all in environmental modelling. Failure to carry out compari-
 383 sons is understandable, given that most modellers have strong
 384 preferences for particular model structures and model-
 385 development approaches. Such preferences may be built on
 386 experience and constrained by resource limitations or lack of
 387 open-mindedness. In an ideal world, a modelling project
 388 would be let out to two or more groups to encourage rigorous
 389 comparison. In the real world, with limited resources, sponsors
 390 of modelling could have a strong influence by demanding
 391 comparisons, if they took the view that a limited but thorough
 392 exercise is preferable to a more ambitious but less well tested
 393 one.

394 A growing risk is that the wider community, decision-
 395 makers and politicians are effectively disfranchised by inabil-
 396 ity to weigh up conclusions drawn from models. Inadequate
 397 reporting and absence of discussion of alternatives can result
 398 in unsystematic, specialised representation of accrued knowl-
 399 edge, not open to challenge. This becomes profoundly

unsatisfactory when model-based conclusions are susceptible
 to gross error through lack of good practice. In some areas
 where there is a consensus on modelling issues but not solu-
 tions, a remedy may be to seek more collaborative and strate-
 gic science, funded to bring groups together internationally to
 execute comparative studies. The EU Research Frameworks
 have such aims among others and are beginning to take a wider
 perspective outside Europe, but there is a need for more flex-
 ible, rapidly responding, heterogeneous, informal yet long-
 term arrangements. Long-term, consistent collaboration is
 needed across a range of modelling communities, to generate
 systematic knowledge representation and testing, gradually de-
 veloping a widely understood and accepted methodological
 platform on which to build and test models.

2.6. More comprehensive testing of models

Environmental models can seldom be fully analysed, if
 only because of the heterogeneity of their data and the range
 of factors influencing usefulness of their outputs. In the case
 of groundwater models, [Konikow and Bredehoeft \(1992\)](#) argue
 from a philosophical and practical viewpoint that the strong
 term “validation” has no place in hydrology. They indicate
 that [Hawking \(1988\)](#) has generalised this further to state that
 “Any physical theory is always provisional, in the sense that
 it is only a hypothesis: you can never prove it.” [Oreskes](#)
[et al. \(1994\)](#) examine the philosophical basis of the terms
 “verification” and “validation” as applied to models. What
 typically passes for these terms is at best confirmation to
 some degree. The two terms imply a stark choice between ac-
 ceptance and rejection. On the contrary we recognise that
 model performance may be assessed against many criteria,
 and that often no sharp acceptance threshold exists. We urge
 discussion of performance, recommending that a wide range
 of performance indicators be examined. The problem-depend-
 ent indicators selected may include:

- satisfactory reproduction of observed behaviour;
- high enough confidence in estimates of model variables and parameters, taking into account the sensitivity of the outputs to all the parameters jointly, as well as the parameter uncertainties;
- plausibility of the model properties, e.g. values which conform with experience for biophysical and socioeconomic parameters and means or extremes of associated variables;
- absence of correlation between model residuals (output errors) and observed inputs, since correlation indicates unmodelled input-output behaviour;
- time- and space-invariance of parameter estimates, since variation indicates poorly or incompletely specified parameters (unmodelled behaviour again);
- satisfactory properties of the residuals, such as absence of significant structure over time and space, e.g. constant mean and variance;
- consistency of the model in cross-validation against different sections of the input-output records ([Janssen et al.](#),

1988) and perhaps also against perturbations of the data typical of their errors;

- along with these technical aspects, a range of model characteristics important to managers and stakeholders, including transparency and flexibility.

One could take this a step further by not only performing and reporting on model checks, but also asking for independent model auditing to provide safeguards to end-users.

2.7. Detection and reduction of overfitting

Model structures with too many parameters are still endemic. Models with too many degrees of freedom incur serious risks. Among them are: fitting to inconsistent or irrelevant “noise” components of records; severely diminished predictive power; ill defined, near-redundant parameter combinations; and obscuring of significant behaviour by the spurious variation allowed by too much freedom. Even so, model testing for redundancies and possible model reduction are seldom reported. Data paucity should limit the model complexity. For example, in modelling of flow and transport for prediction, spatial data on landscape attributes may be useful to structure and discretise a model in fine detail, but detail is unwarranted if the flux measurements available for model calibration cannot support it (Jakeman and Hornberger, 1993). A related sin is the use of a favourite model even when it is over-parameterized for the data available. Indeed there are instances in the literature of simple models with well identified parameters working better than complex models where less formal attention is paid to the parameters. One is Marsili-Libelli and Checchi (2005). They observe that “The current trend in horizontal subsurface constructed wetlands (HSSCW) modelling advocates structures of increasing complexity, which however have produced a limited improvement in the understanding of their internal functioning or in the reliable estimation of their parameters.” Their proposed use of simple model structures in combination with robust identification algorithms deserves attention in a wider domain than HSSCW modelling.

3. Ten steps

Whatever the type of modelling problem, certain common steps must be considered if the goals are credible results and knowledge acquisition, for the immediate purpose of the exercise and for the wider community and the longer term. Major steps have been elucidated, for example, by Jorgensen and Bendoricchio (2001) for ecological modelling, Seppelt (2003) for landscape ecology, Grafton et al. (2004) for economic-environmental systems and Wainwright and Mulligan (2004) for environmental modelling. Young (1993) summarizes a detailed set of steps for a “typical statistical environmental modelling procedure” and comments that it is an interpretation of the scientific method from the Popper viewpoint. The guidance offered by these authors partly complements and partly overlaps ours. We are trying to be more generic and to suggest guidelines for a wide range of model

types. It would be futile to try to categorise families of models comprehensively, but the list below serves to illustrate the breadth of choice. In the main we also avoid reference to real-life examples. Model families and their features include:

- empirical, data-based, statistical models, with structures chosen primarily for their versatility and assuming little in advance, e.g. data-mined clusters, parametric or non-parametric time series models, regressions and their generalisations such as autoregressive moving-average exogenous models, power laws, neural nets;
- stochastic, general-form but highly structured models which can incorporate prior knowledge, e.g. state-space models and hidden Markov models;
- specific theory-based or process-based models (often termed deterministic), as often used in environmental physics and economics, e.g. specific types of partial or ordinary differential or difference equations;
- conceptual models based on assumed structural similarities to the system, e.g. Bayesian (decision) networks, compartmental models, cellular automata;
- agent-based models allowing locally structured emergent behaviour, as distinct from models representing regular behaviour that is averaged or summed over large parts of the system;
- rule-based models, e.g. expert systems, decision trees;
- a spectrum of models which represent dynamics (time-spread responses to the inputs at any given instant) in differing degrees of detail. This spectrum spans instantaneous (static, non-dynamical), discrete-event and discrete-state models (e.g. Petri nets, Markov transition matrices), lumped dynamical (finite-state-dimensional, ordinary differential equation), distributed (partial differential equation) and delay-differential infinite-state-dimensional models;
- a corresponding spectrum of spatial treatments, comprising non-spatial, ‘region-based’ or ‘polygon-based’ spatial, and more finely (in principle continuously) spatially distributed models (e.g. finite-element/grid-based discretisations of partial differential equations).

Many authors also find it useful to distinguish between white box (theory-based), black box (empirical) and grey box (theory-influenced empirical) models (e.g. Seppelt, 2003). The steps we shall delineate are appropriate whether the exercise employs traditional models, e.g. the dynamical-statistical families of models considered by Jung (1999), Norton (1986), Söderström and Stoica (1989), and Young (1984); the empirical, deterministic or conceptual families covered by Jakeman et al. (1993); more recent artificial-intelligence or “knowledge-based” model types (e.g. Davis, 1995; Forsyth, 1984; Kidd, 1987; Schmoldt and Rauscher, 1996); or a mixture. Most of the essential features of development practice outlined in this section are shared by all these types of model. In addition we broaden the context to include the specification of objectives, choice of approach for finding model structures, involvement of interest groups, and choice of parameter estimation methods and algorithms. Although

571 examples will be given, the focus throughout is mainly on
572 what questions must be addressed, not what alternatives exist.

573 The steps sketched in Fig. 1 and listed below are largely it-
574 erative, involving trial and error. If there is pressure to use an
575 already developed model for all or part of the exercise, atten-
576 tion to all steps remains warranted. That is, the steps proposed
577 are not just of relevance for developing a new model. Depend-
578 ing on the purpose, some steps may involve end-users as well
579 as modellers. The steps are not always clearly separable. For
580 instance, it is a matter of taste where the line is drawn between
581 model-structure selection and parameter estimation, as model
582 structures are partly defined by structural parameters.

584 3.1. Definition of the purposes for modelling

585
586 It is a truism that the reasons for modelling should have
587 a large influence on the selecting of a model family or families
588 (see Section 2.5) to represent the system, and on the nature and
589 level of diagnostic checking and model evaluation. However, it
590 is not necessarily easy to be clear about what the purposes are.
591 Different stakeholders will have different degrees of interest in
592 the possible purposes of a single model. For example, a re-
593 source manager is likely to be most concerned with prediction,
594 while a model developer or scientific user may place higher
595 stress on the ability of the model to show what processes domi-
596 nate behaviour of the system. That said, better understanding
597 is valuable for all parties as part of defining the problem and
598 possible solutions, and as a means of assessing how much trust

to place in the model. It is important to recognize that some
purposes, particularly increased understanding of the system
and data, may be realised well even if the final model is
poor in many respects. An inaccurate model may still throw
light on how an environmental system works.

Purposes include:

- gaining a better qualitative understanding of the system (by means including social learning by interest groups);
- knowledge elicitation and review;
- data assessment, discovering coverage, limitations, inconsistencies and gaps;
- concise summarising of data: data reduction;
- providing a focus for discussion of a problem;
- hypothesis generation and testing;
- prediction, both extrapolation from the past and “what if” exploration;
- control-system design: monitoring, diagnosis, decision-making and action-taking (in an environmental context, adaptive management);
- short-term forecasting (worth distinguishing from longer-term prediction, as it usually has a much narrower focus);
- interpolation: estimating variables which cannot be measured directly (state estimation), filling gaps in data;
- providing guidance for management and decision-making.

These motives are not mutually exclusive, of course, but the modeller has to establish the purposes and priorities within the

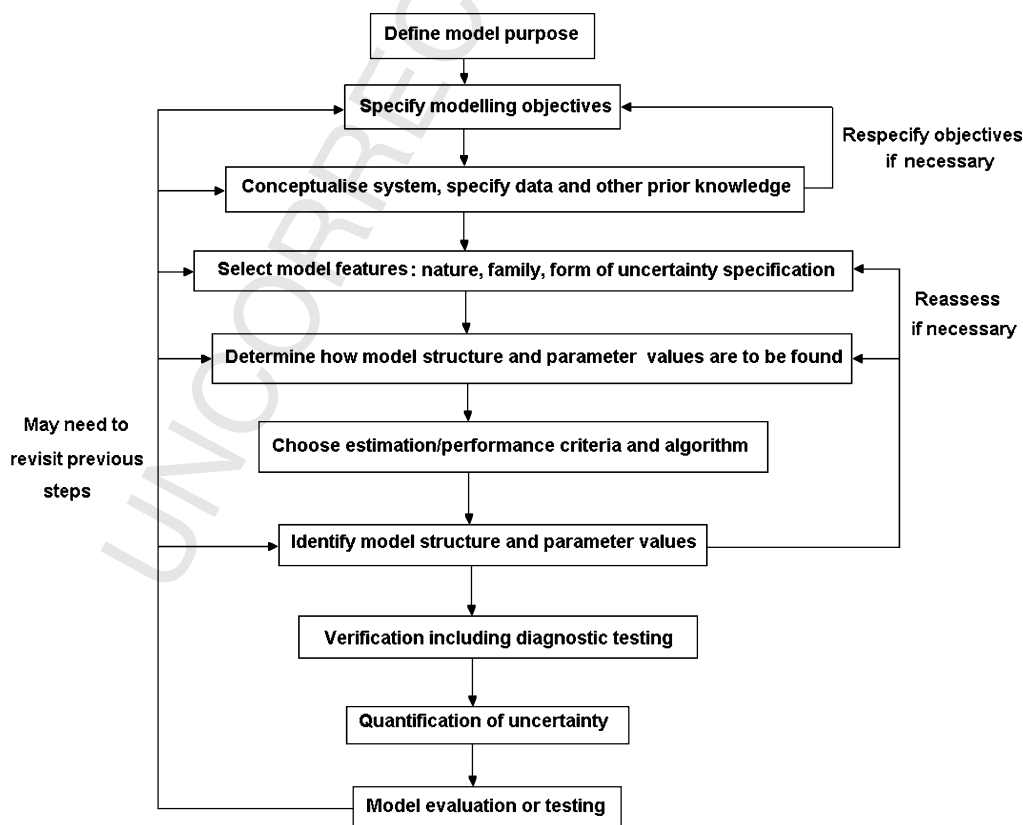


Fig. 1. Iterative relationship between model building steps.

list, because of their influence on the choices to be made at later stages. For example, economy in the degrees of freedom of a prediction model (“parsimony”) is important if the model is to register the consistent behaviour observed in the data but not the ephemeral, inconsistent “noise.” Experience confirms that it is often counterproductive to include much detail in a prediction model for a restricted purpose (Jakeman and Hornberger, 1993). Conversely, a model designed to increase insight into the processes which determine the system’s overall behaviour has to be complex enough to mimic those processes, even if only very approximately. A model intended for knowledge elicitation or hypothesis generation may have a provisional structure too elaborate to be validated by the data, but may be simplified when the knowledge or hypotheses have been tested. Reichert and Omlin (1997) point out possible difficulties in prediction using a parsimonious model with too little flexibility to accommodate changes in perception of which processes are significant. They discuss how to identify and employ non-parsimonious models for prediction. For the modelling of wastewater treatment plants, Gernaey et al. (2004) give some excellent examples of how model purpose influences model selection, data selection and model calibration.

It is worth stressing that improvement of understanding of the system is almost always a purpose of modelling, even when the users say otherwise. The quality of management decisions rests ultimately on how well the system is understood, not merely on the quality of model predictions: insight must, on average, improve decisions. Moreover, increased understanding is often the useful outcome of a modelling exercise which is, by its stated criteria, a failure.

3.2. Specification of the modelling context: scope and resources

This second step identifies:

- the specific questions and issues that the model is to address;
- the interest groups, including the clients or end-users of the model;
- the outputs required;
- the forcing variables (drivers);
- the accuracy expected or hoped for;
- temporal and spatial scope, scale and resolution (but see also Section 3.3);
- the time frame to complete the model as fixed, for example, by when it must be ready to help a decision;
- the effort and resources available for modelling and operating the model, and;
- flexibility; for example, can the model be quickly reconfigured to explore a new scenario proposed by a management group?

A crucial step here is to decide the extent of the model, i.e. where the boundary of the modelled system is. Everything outside and not crossing the boundary is ignored. Everything crossing the boundary is treated as external forcing (known

or unknown) or as outputs (observed or not). The choice of a boundary is closely tied in with the choice of how far to aggregate the behaviour inside it. Classical thermodynamics gives an object lesson in the benefits of choosing the boundary and degree of aggregation well, so as to discover simple relations between a small number of aggregated variables (e.g. energy) crossing the boundary, without having to describe processes inside the boundary in detail. In environmental management, deciding on the boundary and degree of aggregation is a critical but very difficult step. It can usually only be learnt through trial and error, since managers and stakeholders usually do not initially know the boundaries of what should be modelled.

Flexibility can be a major practical issue in matching the scope of the model to resources. For example, the time taken to introduce a new management practice proposed by an interest group might be an issue, given that, for instance, data/GIS layers need to be redrawn. A further concern is the resources to operate the model. In this example, can it be operated by people without GIS training and equipment? More generally, what specialist knowledge does a user need in order to modify a model parameter?

3.3. Conceptualisation of the system, specification of data and other prior knowledge

Conceptualisation refers to basic premises about the working of the system being modelled. It might employ aids to thinking such as an influence diagram, linguistic model, block diagram or bond graph (Gawthrop and Smith, 1996; Wellstead, 1979), showing how model drivers are linked to internal (state) variables and outputs (observed responses). Initially the conceptualisation may be rudimentary, with details postponed until the results of knowledge elicitation and data analysis can be exploited. A tentative initial conceptualisation and a visualisation such as a block diagram may be a great help in showing what else must be found out about the system.

The conceptualisation step is important even if a model is not designed from scratch because time and money (as well as the clients’ beliefs) restrict one to using a ‘canned’ model. Conceptualisation exposes the weaknesses of the canned approach and perhaps ways to mitigate them.

This third step defines the data, prior knowledge and assumptions about processes. The procedure is mainly qualitative to start with, asking what is known of the processes, what records, instrumentation and monitoring are available, and how far they are compatible with the physical and temporal scope dictated by the purposes and objectives. However, it becomes quantitative as soon as we have to decide what to include and what can be simplified or neglected. What variables are to be included, in how much detail? Once the outputs are selected, a rough assessment is needed of which drivers they are sensitive to and what internal processes influence the relations between the drivers and outputs; this will usually be partly a quantitative assessment.

The degree of aggregation and the spatio-temporal resolution (intervals and accuracy) of the outputs also have to be

799 chosen but, as for all these decisions, the choices may have to
800 be revised as experience grows. The time-step and the bounds
801 of what is to be modelled may have to be modified part way
802 through an application, perhaps more than once. This is not
803 trivial. Few models are flexible enough to respond to these
804 evolving needs, which are commonly passed off by modellers
805 as due to the client “not thinking their problem through prop-
806 erly at the beginning.”

807 The first part of this step is just to state what degree of de-
808 tail is needed in the outputs. However, the next step is to fol-
809 low up the implications: the internal resolution of the model
810 must be sufficient to produce outputs at the required resolu-
811 tion, and the time and spatial intervals throughout the model
812 must be compatible with the range of rates of change of the
813 variables. The only way to ensure that these requirements
814 are met is by a careful quantitative assessment. Such assess-
815 ment takes considerable effort and insight into the processes
816 operating in the system, so it is often given too little attention.
817 Too often sampling intervals in time and space are chosen by
818 guesswork or simply because data are available at those inter-
819 vals. Ill-chosen intervals can destroy the validity of the model,
820 but once recognized can be amended as part of the learning
821 process.

822 “Prior knowledge” can be genuinely known in advance,
823 found from experiments or analyses performed as part of model
824 development, or assumed, with reservations, on the basis of ex-
825 perience. It includes observational data and their properties (in-
826 cluding error characteristics), structural information (e.g.
827 coupling or independence, additivity of effects or interaction,
828 existence of feedbacks), the nature of processes (e.g. stationar-
829 ity, correlations, directionality of flows, conservation laws,
830 switching between modes), the extent and nature of spatio-
831 temporal forcing, and parameter values and their uncertainties.
832 Quantitative information on uncertain parameters and errors
833 may consist of point estimates and variances or covariances,
834 bounds (ranges) or, if you are lucky, probability distributions.

835 For some environmental systems one has the luxury of op-
836 timal experimental design where inputs (such as to a bioreac-
837 tor) can be manipulated to enhance the identifiability of
838 a model (e.g. Versyck et al., 1994; Walter and Pronzato,
839 1997). For most systems, however, we must at any given
840 time accept the data that are available. On the other hand,
841 modellers can play a more proactive role in designing future
842 data collection exercises. Monitoring efforts in the global
843 change community are amongst the most striking.

845 3.4. Selection of model features and families

847 Any modelling approach requires selection of model fea-
848 tures, which must conform with the system and data specifica-
849 tion arrived at above. Major features such as the types of
850 variables covered and the nature of their treatment (e.g.
851 white/black/grey box, lumped/distributed, linear/non-linear,
852 stochastic/deterministic) place the model in a particular family
853 or families. Model structure specifies the links between system
854 components and processes. Structural features include the
855 functional form of interactions, data structures or measures

used to specify links, spatial and temporal scales of processes
and their interactions, and bin sizes for AI techniques such as
data-mining. Features help to sharpen the conceptualisation
and determine what model synthesis and calibration tech-
niques are available. In simpler models, a common set of fea-
tures will apply throughout, but a more complex integrated
model may well be a hybrid, with the feature set varying
from one part to another. For example, a deterministic or sta-
tistical climate-prediction model might interface with a non-
statistical but empirical rainfall-runoff model, then with an
irrigation model consisting of predetermined rules.

Families and features often overlap, and in some cases fam-
ilies can even be transformed into each other. For instance lin-
ear, constant-coefficient, ordinary differential equations can be
transformed into, or from, Laplace or Fourier transfer func-
tions. The choice depends on the purpose, objectives, prior
knowledge and convenience.

For prediction and/or management, a key question is what
the subjects of predictive or management interest are. For ex-
ample is a qualitative idea of behaviour (e.g. direction of
change) required, or a rough indication of the extent of a re-
sponse, an extreme value, a trend, a long-term mean, a proba-
bility distribution, a spatial pattern, a time series, the
frequency or location of an event? These questions aren't
asked thoroughly enough at the beginning of model projects.
That said, the initial answers can easily change as the project
develops, especially when managers are involved, emphasiz-
ing again the need for iteration.

The selection of model family should also depend on the
level (quantity and quality) of prior information specified in
step 3.3. It must take account of what can be determined
and how far, i.e. to which accessible and inaccessible variables
the model outputs are sensitive, what aspects of their behav-
iour must be considered, and the associated spatial dimensions
and sampling intervals in space and time.

At this stage a first judgement has to be made of how prom-
inent uncertainty is likely to be. It will help to set reasonable
expectations of capability (e.g. predictive power), and to de-
cide whether and how randomness should be included in the
model formulation. It may include an estimate of how far
past observed behaviour can be extrapolated into the future
or into changed circumstances.

Selection of model features and families should be flexible,
prepared for revision according to evaluation of the reason-
ableness of initial guesses. However, in practice it is usually
difficult to change fundamental features of a model beyond
quite an early stage, for understandable but regrettable human
reasons like unwillingness to admit a poor choice or abandon
something into which much effort has already gone. A prefer-
ence for a particular model, due to familiarity, established ac-
ceptance by the technical community or availability of tools
for it, often impedes change.

The difficulty is exacerbated by uncertainty and changes of
mind about the factors which define model features and family
(part of the learning process). The problem is that expenditure
and commitment to models based on the initial judgements are
usually too powerful to allow any significant changes to be

made. The result may well be an inappropriate model. An initial exploration with a crude, cheap, disposable model would often be a better start, so long as there is enough time and flexibility of mind to allow later choices.

Model structure covers the degree of detail permitted. It may include the choice of spatial units (e.g. hydrological response units or grid cells) and corresponding variables (e.g. points where flows and precipitation are represented), the order of a differential equation representing a process, and whether or not non-linearity or time variation is included in a relation. Selection of model structure and parameter estimation jointly make up model calibration, discussed in Section 3.7. Before calibration, the methods for finding the structure and parameter values have to be selected.

3.5. Choice of how model structure and parameter values are to be found

In finding the structure, prior science-based theoretical knowledge might be enough to suggest the form of the relations between the variables in the model. This is often implicitly assumed to be so, even in complicated environmental systems where it is not. Shortage of records from a system may prevent empirical modelling from scratch and force reliance on scientific knowledge of the underlying processes. Choice of structure is made easier by such knowledge, and it is reassuring to feel that the model incorporates what is known scientifically about the parts of the system. However, empirical studies frequently find that a much simpler structure is adequate for a specified purpose. In some instances the structure may be found by trial and error among a modest number of possibilities, on the basis of credibility of model behaviour. Structural parameters, such as dynamical order or number and location of spatial subdivisions, may sometimes be treatable as extra parameters to be estimated along with the others. Parsimony (Occam's razor) is an overriding principle: avoid more complication than is necessary to fulfil the objectives.

The next choice is of how to estimate the parameter values and supply non-parametric variables and/or data (e.g. distributed boundary conditions). The parameters may be calibrated all together by optimising the fit of the model outputs to observed outputs, or piecemeal by direct measurement or inference from secondary data, or both. Coarse parameter values indicating presence or absence of a factor or the rough timing of a seasonal event, for instance, might be found by eliciting expert opinion.

The choices of how to put the model together must take account not only of what data can be obtained, but also of its informativeness. Substantial quantitative data may be needed to identify parameter values even in a model with a very simple structure. Jakeman and Hornberger (1993) show how few parameters can be identified sharply from daily streamflow data. Substantial trial and error may be required to discover how much can be adequately modelled from a given data set.

In order to ensure uniqueness of parameter estimates, structural identifiability analysis has been undertaken quite actively in a few environmental system types, including activated

sludge biochemical systems (Petersen et al., 2003; Checchi and Marsili-Libelli, 2005). Structural identifiability (Bellman and Åstrom, 1970) concerns what parameters can be identified, in principle, without ambiguity in the absence of measurement errors or deficiencies in model structure.

3.6. Choice of estimation performance criteria and technique

The parameter estimation criteria (hardly ever a single criterion) reflect the desired properties of the estimates. For example we might seek robustness to outliers (bad data), unbiasedness and statistical efficiency, along with acceptable prediction performance on the data set used for calibration. A great deal of effort in recent decades has gone into developing parameter-estimation algorithms with good theoretical properties (Norton, 1986; Söderström and Stoica, 1989; Ljung, 1999). Some of them make quite restrictive assumptions, not always realistic and verifiable, about the properties of the system and the imperfections in the data. Two texts that consider pertinent non-linear theory, at least from a regression analysis perspective, are Bates and Watts (1988) and Seber and Wild (1989).

In selecting an estimation algorithm, rounding errors and ill-conditioning may be a worry, especially when there is a risk that more parameters are being estimated than justified by the data. A further risk is numerical instability, which can arise through injudicious implementation of an algorithm that is stable and well-conditioned in another, exactly algebraically equivalent, implementation. An instance occurs among optimal smoothing algorithms to estimate time-varying parameters (Norton, 1975).

Well executed general-purpose parameter estimation (identification) packages and more specialised packages for hydrological and other uses have now been available for many years (e.g. Ljung, <http://www.mathworks.com/products/sysid>; <http://www.mathworks.com/products/neuralnet>). They may not be able to handle complex, integrated models with specialised structures. If, as a result, parameter-estimation software has to be written, careful testing of the model against criteria not used in the estimation is essential for at least three reasons. First, parameter-estimation algorithms are often predictor-correctors, capable of giving plausible results in the presence of coding errors. Second, parameter estimation for complex models usually involves non-convex numerical optimisation, with a risk that the global optimum is not found. Third, a model, especially one that is put together from several sub-models, may well have more parameters than necessary to prescribe its overall behaviour (over-parameterisation), and may thus not be capable of yielding well-defined estimates of all parameters. Over-parameterisation can lead to misinterpretation, numerical ill-conditioning, excessive ability to fit the "noise" (inconsistent behaviour) in records and poor prediction performance.

In summary, the parameter estimation technique should be:

- computationally as simple as possible to minimise the chance of coding error;

- 1027 • robust in the face of outliers and deviations from assump- 1084
- 1028 tions (e.g. about noise distribution); 1085
- 1029 • as close to statistically efficient as feasible (as reflected by 1086
- 1030 the amount of data required for the estimates to converge); 1087
- 1031 • numerically well-conditioned and reliable in finding the 1088
- 1032 optimum; 1089
- 1033 • able to quantify uncertainty in the results (not at all easy, 1090
- 1034 as the underlying theory is likely to be dubious when the 1091
- 1035 uncertainty is large); and 1092
- 1036 • accompanied by a test for over-parameterisation. 1093
- 1037

1038 In an integrated model, a second area of choice for param- 1094

1039 eter estimation at this stage is of the sections into which the 1095

1040 model is disaggregated. Disciplinary boundaries often define 1096

1041 sections, for example hydrological, policy, economic and eco- 1097

1042 logical components. Spatial sectioning, e.g. of a stream net- 1098

1043 work, is also natural. Sectioning into time segments is much 1099

1044 less common, even though many environmental phenomena 1100

1045 have time-varying characteristics which should influence 1101

1046 model applications such as prediction. 1102

1047 The last decade or so has seen a strong trend towards 1103

1048 models explicitly divided into simpler sections for parameter 1104

1049 estimation, an example being piecewise linear models. Sim- 1105

1050 pler sections make for greater flexibility and easier testing, 1106

1051 but pose a larger risk of producing a model more elaborate 1107

1052 than necessary, e.g. having internal variables with little influ- 1108

1053 ence on external behaviour or higher resolution than needed 1109

1054 to provide the required output resolution. 1110

1055 Practical convenience often dictates piecemeal identifica- 1111

1056 tion of model components, and pre-existing models are often 1112

1057 available for parts of the system (e.g. rainfall-runoff, flood, 1113

1058 groundwater and/or water quality models for hydrological sec- 1114

1059 tions), but it is wise to test the overall model to see whether 1115

1060 simplification is possible for the purposes in mind. Sensitivity 1116

1061 assessment (Saltelli et al., 2000) plays a large rôle here. 1117

1062

1063

1064 3.7. Identification of model structure and parameters 1118

1065

1066 Section 3.5 discussed choice of methods for finding model 1119

1067 structure and parameters, and Section 3.6 the criteria and tech- 1120

1068 niques. The present step addresses the iterative process of find- 1121

1069 ing a suitable model structure and parameter values. This step 1122

1070 ideally involves hypothesis testing of alternative model struc- 1123

1071 tures. The complexity of interactions proposed for the model 1124

1072 may be increased or reduced, according to the results of model 1125

1073 testing (steps 3.8–3.10). In many cases this process just con- 1126

1074 sists of seeing whether particular parameters can be dropped 1127

1075 or have to be added. 1128

1076 Formal statistical techniques for differentiating among dif- 1129

1077 ferent model structures are well developed. They provide cri- 1130

1078 teria which trade the number of parameters against the 1131

1079 improvement in model fit to observations (Veres, 1991). Be- 1132

1080 cause of their reliance on statistical assumptions, statistical 1133

1081 model-structure tests are best treated as guides, checking the 1134

1082 results of the structure recommended on other grounds such 1135

1083 as prediction performance on other data sets, credibility of

parameter estimates and consistency with prior knowledge 1084

(see Sections 3.8 and 3.10). 1085

1086 The underlying aim is to balance sensitivity to system vari- 1087

1088 ables against complexity of representation. The question is 1089

1090 whether some system descriptors, for instance dimensionality 1091

1092 and processes, can be aggregated to make the representation 1093

1094 more efficient, worrying only about what dominates the re- 1095

1096 sponse of the system at the scales of concern. Again it is im- 1097

1098 portant to avoid over-flexibility, since unrealistic behaviour, 1099

1100 ill-conditioning and poor identifiability (impossibility of find- 1101

1102 ing unique, or well enough defined, parameter estimates) are 1103

1104 severe risks from allowing more degrees of freedom than jus- 1104

1105 tified by the data. 1105

1098 3.8. Conditional verification including diagnostic 1099

1100 checking 1101

1101 Once identified, the model must be ‘conditionally’ verified 1102

1102 and tested to ensure it is sufficiently robust, i.e. insensitive to 1103

1103 possible but practically insignificant changes in the data and to 1104

1104 possible deviations of the data and system from the idealising 1105

1105 assumptions made (e.g. of Gaussian distribution of measure- 1106

1106 ment errors, or of linearity of a relation within the model). It 1107

1107 is also necessary to verify that the interactions and outcomes 1108

1108 of the model are feasible and defensible, given the objectives 1109

1109 and the prior knowledge. Of course, this eighth step should in- 1110

1110 volve as wide a range of quantitative and qualitative criteria as 1111

1111 circumstances allow. 1112

1112 Quantitative verification is traditionally attempted, but 1113

1113 rarely against a wide range of criteria. Criteria may include 1114

1114 goodness of fit (comparison of means and variances of ob- 1115

1115 served versus modelled outputs), tests on residuals or errors 1116

1116 (for heteroscedasticity, cross-correlation with model variables, 1117

1117 autocorrelation, isolated anomalously large values) and, par- 1118

1118 ticularly for relatively simple empirical models, the speed 1119

1119 and certainty with which the parameter estimates converge 1120

1120 as more input-output observations are processed. 1121

1121 Qualitative verification preferably involves knowledgeable 1122

1122 data suppliers or model users who are not modellers. Where 1123

1123 the model does not act feasibly or credibly, the assumptions, 1124

1124 including structure and data assumptions, must be re-evaluated. 1125

1125 Indeed, this stage of model development may involve reassess- 1126

1126 ment of the choices made at any previous stage. Checking of 1127

1127 a model for feasibility and credibility is given little promi- 1128

1128 nence in the literature because it is largely informal and 1129

1129 case-specific, but it is plainly essential for confidence in the 1130

1130 model’s outputs. Again this is a very important step, not 1131

1131 only to check the model’s believability, but to build the client’s 1132

1132 confidence in the model. It assumes sufficient time for this 1133

1133 checking and enough flexibility of model structure to allow 1134

1134 modifications. Often these assumptions are not met. 1135

1136 3.9. Quantification of uncertainty 1137

1137 Uncertainty must be considered in developing any model, 1138

1138 but is particularly important, and usually difficult to deal 1139

1139 with, in large, integrated models. Beven (2000) expresses the 1140

1140

concept of model equifinality, recognising that there often is a wide range of models capable of yielding similar predictions. Uncertainty in models (Walker et al., 2003) stems from incomplete system understanding (which processes to include, which processes interact); from imprecise, finite and often sparse data and measurements; and from uncertainty in the baseline inputs and conditions for model runs, including predicted inputs. In Van der Sluijs et al. (2005) uncertainties are considered from a non-technical standpoint, to include those associated with problem framing, indeterminacies and value-ladenness. Their procedure is important if these attributes dominate. A diagnostic diagram can be used to synthesize results of quantitative parameter sensitivity analysis and qualitative review of parameter strength (so-called pedigree analysis). It is a reflective approach where process is as important as technical assessments.

Some modelling approaches are able explicitly to articulate uncertainty due to data, measurements or baseline conditions, by providing estimates of uncertainty, usually in probabilistic form such as parameter covariance. Others require comprehensive testing of the model to develop this understanding. Ideally the model would be exercised over the whole credible range of every uncertain input and parameter, suitably weighted by likelihood. Such comprehensive testing is a complex task even for relatively simple integrated models, so is very rarely performed because of time and resource constraints. For example, the sensitivity of model outputs to changes in individual parameters, and perhaps two at a time, may be tested, but analysis of the effects of bigger combinations of parameter changes is usually limited to crude measures such as contribution to mean-square variation in output, under some statistical assumptions. Funds are seldom available to cover the time that this testing takes, but even some crude error estimates based on output sensitivity to the most important variables is useful. Often modellers do not provide even this level of uncertainty estimation.

The results from extensive sensitivity testing can be difficult to interpret, because of the number and complexity of cause-effect relations tested. To minimise the difficulty, clear priorities are needed for which features of which variables to examine, and which uncertainties to cover. A good deal of trial and error may be required to fix these priorities.

Few approaches explicitly consider uncertainty introduced by the system conceptualisation or model structure. Alternative structures and conceptualisations are unlikely to be examined after an early stage. The reasons include preferences of the developer, compatibility with previous practice or other bodies' choices, availability of software tools, agency policy, peer pressure and fashion within technical communities, and shortage of time and resources. It is hard to see how this sort of uncertainty can be taken into account beyond remaining alert to any compromises and doubts in such choices.

On the positive side, the issue of uncertainty is widely recognised and increasing resources are being devoted to it. For example, Hession and Storm (2000) demonstrate a method for incorporating uncertainty analysis in watershed-level modelling and summarise a lot of the literature in this applied area. A recent special issue of this journal (Jolma and Norton, 2005)

is also indicative of the attention given to uncertainty in environmental modelling. The papers there illustrate the breadth of the field and the eclectic way in which ideas, problem formulations and technical resources from many sources are being brought to bear.

Model uncertainty must be considered in the context of the purposes of the model. For example, discrepancies between actual output, model output and observed output may be important for forecasting models, where cost, benefit and risk over a substantial period must be gauged, but much less critical for decision-making or management models where the user may be satisfied to know with knowing that the predicted ranking order of impacts of alternative scenarios or management options is likely to be correct, with only a rough indication of their sizes.

3.10. Model evaluation or testing (other models, algorithms, comparisons with alternatives)

Finally the model must be evaluated in the light of its objectives. For simpler, disciplinary models, a traditional scientific attitude can be taken towards “validation” (non-falsification or provisional confirmation, strictly). That is, confirmation is considered to be demonstrated by evaluating model performance against data not used to construct the model (Ljung, 1999, ch. 16; Söderström and Stoica, 1989, ch.11). However, this style or level of confirmation is rarely possible (or perhaps even appropriate) for large, integrated models, especially when they have to extrapolate beyond the situation for which they were calibrated. If so, the criteria have to be fitness for purpose and transparency of the process by which the model is produced, rather than consistency with all available knowledge. More detailed assessment of the model ‘for the purposes for which it has been constructed’ must be considered (e.g. Ravetz, 1997).

Details of such an approach are still at an early stage of development, but should extend to: testing the sensitivity of the model to plausible changes in input parameters; where possible or desirable, changes in assumptions about model structure; as well as documentation and critical scrutiny of the process by which the model has been developed, including the assumptions invoked. A critical difference from traditional model “validation” is the openly subjective nature of such criteria.

Fitness for purpose should also include ‘softer’ criteria like ability to accommodate unexpected scenarios and to report predictions under diverse categories (by interest group, by location, by time, etc), and speed of responding to requests for modified predictions. In other words, model accuracy (the traditional modeller’s criterion) is only one of the criteria important in real applications.

In summary, the modelling process is about constructing or discovering purposeful, credible models from data and prior knowledge, in consort with end-users, with every stage open to critical review and revision. Sadly, too often in reality it is the application of a predetermined model in a highly constricted way to a problem, and to the social dimensions of which the modeller is oblivious.

4. Minimum standards and education

We conclude by noting that certain minimum standards suggest themselves in reporting on model development and performance and in progressing knowledge. Aber et al. (2003) summarise a workshop discussion on much-needed standards, such as exist for ecological data, of practice for review and publication of models in ecology. They relate to reporting on model structure, parameterisation, testing and sensitivity analysis. Hoping to cover a wide range of modelling situations, we recommend that the standards include (but may not be limited to):

- clear statement of the objectives and clients of the modelling exercise;
- documentation of the nature (identity, provenance, quantity and quality) of the data used to drive, identify and test the model;
- a strong rationale for the choice of model families and features (encompassing alternatives);
- justification of the methods and criteria employed in calibration;
- as thorough analysis and testing of model performance as resources allow and the application demands;
- a resultant statement of model utility, assumptions, accuracy, limitations, and the need and potential for improvement; and quite obviously but importantly;
- fully adequate reporting of all of the above, sufficient to allow informed criticism.

Adoption of these standards by modellers, through fuller execution and reporting of the steps outlined in this paper, would benefit both the model-building community and those relying on model-based insight and model recommendations to make decisions.

In addition to adhering to standards, the education of modellers on further aspects is warranted, for instance on how to engage with clients and stakeholders, on the need to develop more flexible models and on understanding the context in which the model will be used.

Acknowledgments

The authors are grateful to Jessica Spate, Stefano Marsili-Libelli, David Swayne, Richard Davis, Rodolfo Soncini Sessa, Ralf Seppelt, Alexey Voinov, Tom Chapman and several anonymous reviewers for their detailed comments on the manuscript. Thanks to Andrea Rizzoli for managing the manuscript review process and Giorgio Guariso for inventing the Position Paper idea and process for EMS.

References

Aber, J.D., Bernhardt, E.S., Dijkstra, F.A., Gardner, R.H., Macneale, K.H., Parton, W.J., Pickett, S.T.A., Urban, D.L., Weathers, K.C., 2003. Standards of practice for review and publication of models: summary of discussion. In: Canham, C.D., Cole, J.J., Lauenroth, W.K. (Eds.), *Models in Ecosystem Science*. Princeton University Press, New Jersey.

Bates, D., Watts, D., 1988. *Non-linear Regression Analysis and Its Applications*. John Wiley and Sons, New York.

Beck, M.B., 2005. Vulnerability of water quality in intensively developing urban watersheds. *Environmental Modelling and Software* 20, 381–400.

Bellman, R., Åstrom, K.-J., 1970. On structural identifiability. *Mathematical Biosciences* 7, 329–339.

Beven, K.J., 2000. *Rainfall-Runoff Modelling: The Primer*. John Wiley and Sons Ltd, Chichester.

Caminiti, J.E., 2004. Catchment modelling – a resource manager’s perspective. *Environmental Modelling and Software* 19, 991–997.

Checchi, N., Marsili-Libelli, S., 2005. Reliability of parameter estimation in respirometric models. *Water Research* 39, 3686–3696.

Davis, J.R., 1995. Expert systems and environmental modelling. In: Jakeman, A.J., Beck, M.B., McAleer, M.J. (Eds.), *Modelling Change in Environmental Systems*. Wiley, Chichester, pp. 505–517.

Denzer, R., 2005. Generic integration of environmental decision support systems – state-of-the-art. *Environmental Modelling and Software* 20, 1217–1223.

Fath, B.D., Beck, M.B., 2005. Elucidating public perceptions of environmental behavior: a case study of Lake Lanier. *Environmental Modelling and Software* 20, 485–498.

Forsyth, R., 1984. The expert systems phenomenon. In: Forsyth, R. (Ed.), *Expert Systems: Principles and Case Studies*. Chapman and Hall, London, pp. 3–8.

Funtowicz, S.O., Ravetz, J.R., 1990. *Uncertainty and Quality in Science and Policy*. Kluwer, Dordrecht.

Gawthrop, P.J., Smith, L.S., 1996. *Metamodelling: Bond Graphs and Dynamic Systems*. Prentice Hall, Englewood Cliffs, NJ, USA.

Gernaey, K.V., van Loosdrecht, M.C.M., Henze, M., Lind, M., Jorgensen, S.B., 2004. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental Modelling and Software* 19, 763–783.

Grafton, R.Q., Adamowicz, W., Dupont, D., Nelson, H., Hill, R.J., Renzetti, S., 2004. *The Economics of the Environment and Natural Resources*. Blackwell Publishing Ltd, Oxford.

Hare, M., Letcher, R.A., Jakeman, A.J., 2003. Participatory modelling in natural resource management: a comparison of four case studies. *Integrated Assessment* 4 (2), 62–72.

Hawking, S.W., 1988. *A Brief History of Time: From the Big Bang to Black Holes*. Bantam Books, New York.

Hession, W.C., Storm, D.E., 2000. Watershed-level uncertainties; implications for phosphorus management and eutrophication. *Journal of Environmental Quality* 29, 1172–1179.

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall–runoff model? *Water Resources Research* 29, 2637–2649.

Jakeman, A.J., Beck, M.B., McAleer, M.J. (Eds.), 1993. *Modelling Change in Environmental Systems*. Wiley, Chichester.

Janssen, P., Stoica, P., Söderström, T., Eykhoff, P., 1988. Model structure selection for multivariable systems by cross-validation methods. *International Journal of Control* 47, 1737–1758.

Jolma, A., Norton, J.P., 2005. Methods of uncertainty treatment in environmental models. *Environmental Modelling and Software* 20, 979–980.

Jorgensen, S.E., Bendricchio, G., 2001. *Fundamentals of Ecological Modelling*. Elsevier, Amsterdam.

Kidd, A.L., 1987. Knowledge acquisition an introductory framework. In: Kidd, A.L. (Ed.), *Knowledge Acquisition for Expert Systems: A Practical Handbook*. Plenum Press, New York.

Konikow, L.F., Bredehoeft, J.D., 1992. Ground-water models cannot be validated. *Advances in Water Resources* 15, 75–83.

Letcher, R.A., Jakeman, A.J., 2003. Application of an adaptive method for integrated assessment of water allocation issues in the Namoi River Catchment, Australia. *Integrated Assessment* 4, 73–89.

Ljung, L., 1999. *System Identification – Theory for the User*, second ed. Prentice Hall, Upper Saddle River, NJ.

Marsili-Libelli, S., Checchi, N., 2005. Identification of dynamic models for horizontal subsurface constructed wetlands. *Ecological Modelling* 187, 201–218.

Mostert, E., Participation for sustainable water management. In: Giupponi, C., Jakeman, A., Karssenberg D., Hare M. (Eds.), *Sustainable Management of Water Resources: An Integrated Approach*. Edward Elgar, Cheltenham, UK, in press.

1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368

- 1369 Murray-Darling Basin Commission, 2000. Groundwater Flow Modelling
1370 Guideline. Murray-Darling Basin Commission, Canberra. Project no. 125.
1371 Norton, J.P., 1975. Optimal smoothing in the identification of linear time-
1372 varying systems. *Proceedings IEE* 122 (6), 663–668.
1373 Norton, J.P., 1986. *An Introduction to Identification*. Academic Press, London.
1374 Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation,
1375 and confirmation of numerical models in the earth sciences. *Science*
1376 263, 641–646.
1377 Parker, P., Letcher, R., Jakeman, A.J., Beck, M.B., Harris, G., Argent, R.M.,
1378 Hare, M., Pahl-Wostl, C., Voinov, A., Janssen, M., et al., 2002. Progress
1379 in integrated assessment and modeling. *Environmental Modelling and*
1380 *Software* 7 (3), 209–217.
1381 Ravetz, J.R., 1997. Integrated Environmental Assessment Forum: Developing
1382 Guidelines for “Good Practice”. Darmstadt University of Technology..
1383 ULYSSES WP-97–1, ULYSSES Project.
1384 Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholte, H., Kassahun, A.,
1385 2005. Quality assurance in model based water management – review of
1386 existing practice and outline of new approaches. *Environmental Modelling*
1387 *and Software* 20, 1201–1215.
1388 Reichert, P., Omlin, M., 1997. On the usefulness of overparameterized ecolog-
1389 ical models. *Ecological Modelling* 95, 289–299.
1390 Risbey, J., Kandlikar, M., Patwardhan, A., 1996. Assessing integrated assess-
1391 ments. *Climatic Change* 34, 369–395.
1392 Saltelli, A., Chan, K., Scott, E.M. (Eds.), 2000. *Sensitivity Analysis*. Wiley,
1393 Chichester, UK.
Schmoldt, D.L., Rauscher, H.M., 1996. *Building Knowledge-based Systems*
for Natural Resource Management. Chapman and Hall, New York.
Seber, G.A.F., Wild, C.J., 1989. *Non-linear Regression*. Wiley, New York.
Seppelt, R., 2003. *Computer-based Environmental Management*. VCH-Wiley
Verlag GmbH & Co, Weinheim, Germany.
Söderström, T., Stoica, P., 1989. *System Identification*. Prentice Hall Interna-
tional, UK.
STOWA/RIZA, 1999. *Smooth Modelling in Water Management, Good Mod-
elling Practice Handbook*. STOWA Report 99–05. Dutch Department of
Public Works, Institute for Inland Water Management and Waste Water
Treatment, ISBN 90-5773-056-1. Report 99.036.
Van der Sluijs, J.P., Craye, M., Funtowicz, S., Klopogge, P., Ravetz, J.,
Risbey, J., 2005. Combining quantitative and qualitative measures of un-
certainty in model-based environmental assessment: the NUSAP system.
Risk Analysis 25, 481–492.
Veres, S.M., 1991. *Structure Selection of Stochastic Dynamic Systems*.
Gordon & Breach, London.
Versyck, K.J., Claes, J.E., Van Impe, J.F., 1994. Optimal experimental design
for practical identification of unstructured growth models. *Mathematics*
and Computers in Simulation 46, 621–629.
Environmental Modelling: Finding Simplicity in Complexity. In:
Wainwright, J., Mulligan, M. (Eds.). Wiley, Chichester.
Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van
Asselt, M.B.A., Janssen, P., Kraymer von Krauss, M.P., 2003. Defining un-
certainty: a conceptual basis for uncertainty management in model-based
decision support. *Integrated Assessment* 4, 5–18.
Walter, E., Pronzato, L., 1997. *Identification of Parametric Models from*
Experimental Data. Springer Verlag, Berlin.
Wellstead, P.E., 1979. *Introduction to Physical System Modelling*. Academic
Press, London.
Young, P.C., 1984. *Recursive Estimation and Time Series Analysis*. Springer
Verlag, Berlin.
Young, P.C., 1993. Environmental modelling and the scientific method. In:
P.C. Young (Ed.), *Concise Encyclopaedia of Environmental Systems*. Per-
gamon, Oxford, pp. 204–206.