2008

Good Modelling Practice

N. Crout
University of Nottingham

T. Kokkonen
Helsinki University of Technology

A.J. Jakeman
The Australian National University

J.P. Norton
The Australian National University

L.T.H. Newham
The Australian National University

See next page for additional authors

Crout, N.; Kokkonen, T.; Jakeman, A.J.; Norton, J.P.; Newham, L.T.H.; Anderson, R.; Assaf, H.; Croke, B.F.W.; Gaber, N.; Gibbons, J.; Holzworth, D.; Mysiak, J.; Reichl, J.; Seppelt, R.; Wagener, T.; and Whitfield, P., "Good Modelling Practice" (2008). *U.S. Environmental Protection Agency Papers*. 73.
https://digitalcommons.unl.edu/usepapapers/73

Authors
N. Crout, T. Kokkonen, A.J. Jakeman, J.P. Norton, L.T.H. Newham, R. Anderson, H. Assaf, B.F.W. Croke, N. Gaber, J. Gibbons, D. Holzworth, J. Mysiak, J. Reichl, R. Seppelt, T. Wagener, and P. Whitfield

CHAPTER TWO

Good Modelling Practice

N. Crout[a], T. Kokkonen[b], A.J. Jakeman[c], J.P. Norton[d], L.T.H. Newham[c], R. Anderson[e], H. Assaf[f], B.F.W. Croke[g], N. Gaber[h], J. Gibbons[i], D. Holzworth[j], J. Mysiak[k], J. Reichl[l], R. Seppelt[m], T. Wagener[n], *and* P. Whitfield[o]

Contents

[a] School of Biosciences, University of Nottingham, Nottingham, NG7 2RD, UK
[b] Department of Civil Environmental Engineering, Helsinki University of Technology, PO Box 5300, FI-02015 TKK, Finland
[c] Integrated Catchment Assessment & Management Centre, Fenner School of Environment & Society, The Australian National University, Bldg. 48A, Linnaeus Way, Canberra, ACT 0200, Australia
[d] Integrated Catchment Assessment & Management Centre, Fenner School of Environment & Society, and Mathematical Sciences Institute, The Australian National University, Bldg. 48A, Linnaeus Way, Canberra ACT 0200, Australia
[e] Nicholas School of the Environment and Earth Sciences, Environmental Sciences and Policy Division, Box 90328, Duke University, Durham, NC 27708, USA
[f] Department of Civil & Environmental Engineering, American University of Beirut, PO Box 11-0236 Riad El-Solh, Beirut, Lebanon
[g] Integrated Catchment Assessment & Management Centre, Fenner School of Environment & Society, The Australian National University, Bldg. 48A, Linnaeus Way, Canberra, ACT 0200, Australia
[h] US Environmental Protection Agency, 1200 Pennsylvania Avenue, N.W., 8105R, Washington, DC 20460, USA
[i] School of the Environment & Natural Resources, University of Wales, Bangor, LL57 2UW, United Kingdom
[j] CSIRO Sustainable Ecosystems, PO Box 102 Toowoomba, Qld 4350, Australia
[k] Fondazione Eni Enrico Mattei, Palazzo Querini Stampalia, Campo S. Maria Formosa, Castello 5252, 30122 Venezia, Italy
[l] Department of Civil and Environmental Engineering, University of Melbourne, Victoria 3010, Australia
[m] Centre for Environmental Research, Martin-Luther University, Halle-Wittenberg, 04301 Leipzig, Germany
[n] Department of Civil and Environmental Engineering, 226B Sackett Bldg., Pennsylvania State University, University Park, PA 16802, USA
[o] Meteorological Service of Canada, Environment Canada, 401 Burrard Street, Vancouver, BC, Canada

Environmental Modelling, Software and Decision Support
1574-101X.

2.1. Introduction

Best-practice guidelines for modelling have been developed by a number of organisations to promote better understanding of model development and application, facilitate tests of model quality and provide a framework for documenting and communicating modelling activities among modellers and decision makers. Good practice within a Data Mining paradigm is presented in Chapter 12.

Refsgaard and Henriksen (2004) reviewed a number of modelling guidelines and proposed a framework for quality assurance, including the development of consistent terminology. Current practice was found to vary widely by domain as well as among countries, revealing varying levels of scientific maturity in the disciplines and the modelling market.

The key elements of existing guidelines cover technical issues of development, implementation and use of models, primarily domain-specific, as well as issues involving interaction between the modeller and end-user, the content of which may be more general. Key elements of existing technical guidelines include definition of the purpose of the modelling; collection and processing of data; establishment of a conceptual model; computer implementation; model set-up; establishment of performance criteria; calibration; validation; uncertainty assessments; simulation with the model for a specific purpose; and reporting.

Another approach to developing comprehensive guidelines for environmental modelling was taken by the US Environmental Protection Agency's Council for Regulatory Environmental Modeling. Given inherent uncertainty in the approximation of reality by models, the EPA view was that the most important issue facing model developers and users is determining when a model can be appropriately used to inform a decision. This led to the Draft Guidance for Environmental Models, which focuses on three major steps in the modelling process and proposes the following best practices for each.

Model Development: present a clear statement and description (in words, functional expressions, diagrams, and graphs, as necessary) of each element of the conceptual model and the science behind it; when possible, test competing conceptual models/hypotheses; use sensitivity analysis early and often; determine the optimal level of model complexity by making appropriate tradeoffs among competing objectives; where possible, model parameters should be characterised using direct measurements of sample populations and all input data should meet data quality acceptance criteria.

Model Evaluation: peer review of models, development of a quality assurance project plan including measures to assess input data quality, model corroboration and sensitivity and uncertainty analysis. In this guidance, corroboration is defined as a qualitative and/or quantitative evaluation of the accuracy and relevant capabilities of a model. Given the iterative nature of the model evaluation process, it follows that these qualitative and quantitative assessment techniques may be effectively applied throughout model development, testing and application.

Model Application: it is considered that model-based decision making is strengthened when the underlying science is transparent via: (1) comprehensive documentation of all aspects of a modelling project; and (2) effective communication between modellers, analysts, and decision makers. This transparency encourages a clear rationale for using a model in a specific regulatory purpose. Proper documentation enables decision makers and other users of models to understand the process by which a model was developed, its intended area of application, and the limitations of its applicability. One of the major objectives of documentation should be to reduce the uncertainty with respect to areas of application.

2.2. Key Components of Good Modelling Practice

From the work outlined above we can identify some general components of best modelling practice: (1) definition of purpose; (2) model evaluation, however that should be defined; and (3) transparency of the model and its outputs. Aspects of each of these components are described below.

2.2.1 Model purpose

What is the model for? Without defining the model's purpose its degree of success cannot be judged and its structural complexity cannot be advantageously tuned. The entire process of model development and evaluation will be driven by the underlying model purpose; the more explicit the statement of this purpose, the better.

In general, models can be used to (i) measure and represent; (ii) describe structure, behaviour and pattern; (iii) reconstruct past or predict future behaviour; (iv) generate and test theories and hypotheses; (v) display, encode, transfer, evaluate and interpret knowledge; (vi) guide development and assessment of policies; and (vii) facilitate collective learning and settlement of disputes (Morton, 1990; Beven, 2002; Jakeman et al., 2006). Practical uses of models may be blurred or overlapping, but this does not change the implications of the intended purpose for model development. Further, wide ranging examples are discussed in this volume in the papers by McIntosh et al. (Chapter 3), Brugnach et al. (Chapter 4), and Maier et al. (Chapter 5).

Bankes (1993) cautions against confusion between the purposes of consolidative and exploratory models. A consolidative model sums up facts known to be correct in a single package, used as a surrogate for the actual system. The system behaviour is predicted reliably enough to derive, for example, likely consequences for management interventions. If however the available knowledge and inherent uncertainties preclude building a surrogate for the system, a model functions as an experiment to explore the implications or varying assumptions and hypotheses. Exploratory models, that is models in which not all components of the system can be established independently or are known to be 'correct' (Sarewitz and Pielke, 2000; Pielke, 2003), require a different development methodology. Instead of providing unreliable prediction, they can help to (i) discover unexpected results of various assumptions, (ii) generate hypotheses; and (iii) identify limiting, worst cases under various assumptions (Bankes, 1993).

To a large extent all models aim at explanation (Jakeman et al., 2006), but models which are good at explaining a system's causal mechanisms, behaviour or patterns are not always built to predict. An example involves models in the earth sciences (Oreskes, 2000) which aim to understand and anticipate contingencies in the natural world (e.g. earthquakes, landslides, volcano eruptions). On the other hand, some prediction models perform poorly in explanation (discovering causal relationships), as a consequence either of "black-box" model structures or assumptions being known to be grossly simplified. There is a danger of over-generalising about how to assess models. For instance, in his controversial statement Friedman (1953) argues that the only quality of a model is whether it yields predictions that are good enough for the purpose in hand and that are better than predictions from alternative models. This ignores purposes not served by prediction alone, and in any case the second criterion is plainly not needed if the first is met. Important and significant hypotheses are frequently inaccurate, descriptive representations of reality, but that does not necessarily disqualify them from usefulness.

Another case in which different purposes are frequently confused is prediction for science versus prediction for policy making (Pielke, 2003). Although both are driven by a similar aim, that is to anticipate outcomes and consequences, their use and motivation (how and why to predict) are different. Fundamental research is typically curiosity-driven, often unpredictable in its course and outcomes, concerned with testing of scientific hypotheses. Researchers are interested in discovering salient features at the frontier of knowledge. As a consequence, scientific studies may be framed (prejudiced) or yield results which are too narrow, not transferable and of limited use for practical policy making. Policy makers, on the other hand, deal with wider contexts, conflicts and large uncertainties. Models are expected to yield not only reliable, but also socially robust knowledge. The misunderstanding of these differences is wrongly attributed to policy makers not being able to understand the scientific models or scientists oversimplifying the complexity of policy issues. Such misunderstandings often manifest themselves when 'science'-driven models are developed and adapted for application as 'policy' models. The importance of the original model purpose to the subsequent model development process needs to be more widely recognised and understood.

2.2.2 Model evaluation

The evaluation of models should be a central part of the model development process, not an afterthought. Even today it is often the case that primary model development consumes more time and resources than model evaluation. Traditionally model evaluation has involved some measures of predictive performance and perhaps an uncertainty analysis. Although important, these should only be a starting point and increasingly effort is being devoted to evaluating the model assumptions and formulation within iterative processes of development (Wagener, 2003). These approaches should be more appropriate for evaluating whether a model is suitable for its purpose than a simple evaluation of its predictive capability.

The evaluation phase should also include assessment of the data utilised in the modelling study. In environmental sciences one typically needs to process observed data before they can be used (e.g. correct precipitation measurements for wind effects, or derive area averages from point measurements). At least equally important in assessing the model assumptions is explicit statement of any assumptions and approximations made in compiling the data set.

Below we review some approaches to model performance measures. We broaden the discussion to the evaluation of model assumptions, and then consider the possibility of more formalised continuing model evaluation.

2.2.3 Performance measures

The role of performance indicators is often to indicate accurately the fit between a model and observations, usually from a particular viewpoint (e.g. larger individual values in the observations being more significant than smaller values). Ideally, the performance indicator(s) employed should reflect the purpose of the modelling exercise. A standard performance indicator may not always be the correct choice; for example, a study investigating low flows in rivers should not necessarily employ the same performance indicator as one investigating flood peaks.

Ideally performance indicators should take into account errors in the observations as well as in the model predictions (due to errors in inputs, model parameters and model structure). However there is a variety of widely employed goodness-of-fit indicators which do not.

In some domains particular performance measures have become generally accepted; for example, in hydrology the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970) is widely used and is referred to in the literature in a number of ways. It is often used by default and apparently without critical thought, even when it is ill-matched to the purpose of the model, or to proper comparison of models. Performance indicators can also be based on transformations of the observed and modelled values. Examples of these include cumulative probability distributions, cross-correlation functions and power spectra. Selection of such performance indicators needs as much thought as that of an indicator for untransformed series, but this does not always occur.

An alternative may be to adopt a wavelet approach, where the fit to the data is measured for a range of scales across all available time periods (e.g. Lane, 2004). This produces a 2D image representation of model performance, thus giving the user much more information at the cost of making comparisons between models more difficult.

While such statistical performance measures are frequently used to test model performance, graphical performance measures can provide valuable insight into model shortcomings not captured in simple performance statistics. Often the only graphical performance measure used is a single graph showing properties of both observed and modelled (spatial and/or temporal) series. Examining raw observed and model output series is often very informative about shortcomings in both data and model not revealed by statistics (e.g. timing errors, inhomogeneous performance, failure of matching at extremes). Few modellers seem aware of more powerful visualisation techniques.

To illustrate that visualisation techniques can be improved in many fields, consider an example from hydrology (see Figure 2.1). Plotting of observed and modelled flow time series is viewed as a fundamental step in examining adequacy of a hydrological model. The practice should be to plot observed and predicted in different line thicknesses, types or colours so that they can be clearly distinguished. Included in this plot should also be the residuals between the predicted and observed values. This time series should ideally have no structure and be simply a plot of white noise. However, for less-than-perfect models the residuals (or, for that matter, the observed and modelled series) will be instructive about a wide variety of hydrograph features. Errors in timing of peaks result in pairs of residual spikes of opposite signs, a long error sequence of residuals with the same sign indicates systematic over- or under-prediction, and in-homogeneity of the error may be easier to spot in residuals than in observed and modelled series. The problem at hand determines which part of the hydrograph is most of interest in assessing model performance. If the interest is in predicting flood peaks accurately, inadequate representation of base flow is not so important, but if the interest is in the low-flow regime, capturing the timing and magnitude of peak flows is irrelevant. In most cases it is necessary to present the hydrographs on two scales, the first [linear or logarithmic] to show the model agreement in magnitude and a second with the abscissa foreshortened to illustrate hydrograph shape better.

When studying long flow series, a shorter window should also be used in plotting the hydrographs. This is important as a compressed time scale can make timing errors undetectable by eye. Plots of the autocorrelation functions of the residuals can provide additional insight into deficiencies in model structure and allow assessment of whether they are important in a specific application.

An ideal visualisation technique will allow us to see model errors both in the timing and magnitude of the predictions. Furthermore, it will aid us in analysing which process description is most likely to give unsatisfactory model performance. Good visualisations provide valuable information for the assessment and assertions of model adequacy, in a more versatile way than simple statistical performance criteria.

2.2.4 Stating and testing model assumptions

Any model development process requires the modeller to make a series of simplifying assumptions or hypotheses (Gupta et al., 2005). This is necessary so as to describe complex natural systems using much simpler mathematical models.

These assumptions can relate to at least two aspects of model building:

(1) assumptions about the underlying conceptual model describing the modeller's understanding of the natural system;

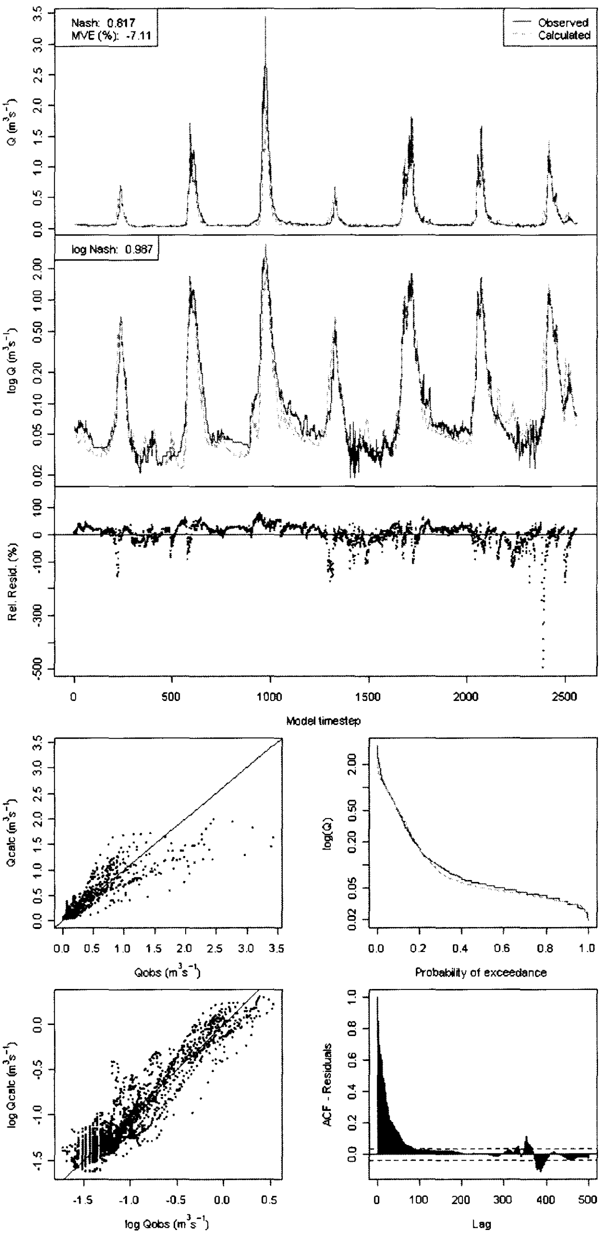

Figure 2.1 Visualisation of adequacy of model performance. [a] linear scale time series plots of observed [solid] and modelled flow time series [dashed]; [b] log scale time series plots of observed [solid] and modelled flow time series [dashed]; [c] time series plot of residuals [dotted] between observed and modelled; [d] observed vs modelled on linear scale; [e] cumulative distribution function of observed [solid] and modelled [dashed]; [f] observed vs modelled on logarithmic scale; [g] autocorrelation function of residuals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this chapter.)

(2) assumptions about how this conceptual model is translated into a model on a computer.

Assumptions of type (1) could, for example, include hypotheses about the dominant runoff-production mechanisms in a watershed, aquifer characteristics, or the behaviour of a certain plant or animal species.

Assumptions of type (2) relate to the simplifications made when translating the conceptual model into equations or rules, for example, for a specific application. They could include assuming that spatial variability below the chosen model element scale is negligible; that contaminant degradation is a first-order process; or that certain processes can be described using linear approximations without introducing too much error.

Under good modelling practice, these assumptions should be listed explicitly to describe the thought process of the modeller and to allow testing of these assumptions at a later stage. Beven (2000) provides a list of excellent examples of assumptions made in the formulation of rainfall-runoff models and in the formulation of mathematical descriptions of hydrological processes in general.

While listing the assumptions, the modeller should strive to provide brief but explicit statements to justify the assumptions. There is no reason why the justifications should not include listing subjective preferences and opinions. By stating them openly it is possible to assess or discuss them.

Listing and justifying assumptions is a very important step in model development, and it has increasingly been suggested that testing of some underlying assumptions is possible and should be included in the modelling process (e.g. Wagener et al., 2003). These suggestions mainly relate to the evaluation of the model behaviour using real data, and go beyond mere assessment of performance. The suggested additional testing of assumptions refers to answering the following questions (e.g. Wagener and Kollat, 2007):

(1) Does a model parameter or a group of parameters represent the process it is intended to represent (i.e. does it dominate the model response when this process dominates the system response)?
(2) Are regions of well performing parameter values constant in time, or do they vary with different response modes of the system? Of course there might be parameters that should vary in time; in such a case the test should be whether they vary appropriately.
(3) Is there a single set of model parameter values that is optimal in reproducing different variables (e.g. flow and water quality variables) simultaneously?

Different approaches have emerged in the literature to address these questions, for example:

- Norton (1975) and Beck (1987) argued how recursive parameter estimation treating parameters as state variables can show that model parameter values have to vary in time for high performance, thus violating the assumption of time-invariant parameters.
- Jakeman et al. (1994) show how the calibration of a rainfall-runoff model for short time periods, derived by breaking up a longer time-series, can expose

changes (in their case post-deforestation behaviour) in the underlying watershed. Some parameters have increasing optimal values reflecting regrowth of vegetation. Assuming time-invariant parameters would thus clearly be wrong in this case. The model needs augmentation by a model (with constant parameters) for the effects of regrowth (question 2).

- Gupta et al. (1998) utilised a multiobjective approach to show that a rainfall-runoff model structure was incapable of fitting different objective functions, but rather shows a tradeoff. This indicates that the assumption of a single optimal parameter set to represent all response modes is violated (question 2).
- Gupta et al. (1999) showed that a land surface scheme was incapable of simultaneously reproducing latent heat and soil moisture fluxes with a single parameter set (question 3).
- Wagener et al. (2003) used a Monte Carlo-based moving window approach to find periods of high parameter sensitivity (question 1), and to evaluate whether areas with high frequencies of well-performing parameter values moved within the parameter space through time (question 2).

Approaches like these can be taken further and the variation of parameter values in time (including their sensitivity) can be estimated more formally in an evaluation framework (e.g. Beck, 1987; Beck, 2002; Young, 1998; Wagener et al., 2003; Wagener and Kollat, 2007). Including such an explicit treatment of assumptions moves the modelling process one step towards a diagnostic analysis of how the model fails and why, thus providing the modeller with opportunity to adjust and improve the model.

There has been recent interest in methods which seek to vary the model structure and evaluate to what extent different model formulations change the predicted quantities. For example Asgharbeygi et al. (2006) introduced the idea of automatic model revision and Cox et al. (2006) suggest ways in which models can be systematically simplified. In the latter case examples are presented where this approach finds simpler models which predictively outperform the original model.

2.2.5 Ongoing model testing and evaluation

Model development and evaluation are similar to the general process of software development, albeit with scientific uncertainty as an additional consideration. Software development is a challenging task, generally prone to an exceptionally high rate of failures due to many factors including: (1) underestimation of budget and time constraints; (2) failure to adequately understand and appreciate what is expected of the system; (3) lack of technical expertise and proper development tools; and (4) the inherent uncertainty of the software development process, especially when it involves moving into new territory. These issues have continuously shaped the software development discipline since its inception in the 1940s. In the early drives to streamline the software development process, software architects initially adopted the sequential waterfall lifecycle approach, which emphasises the thorough and detailed completion of each engineering phase before signing off to the next one, analogous to the one-directional flow of waterfall.

Despite the remarkable success of the waterfall lifecycle approach in other engineering disciplines, its rigidity has contributed to the failure of many software development projects (Larman, 2002). These failures prompted a major transition in the software engineering to the more flexible incremental and iterative approach (Jacobson et al., 1999). In this approach the system is developed in cycles, with each cycle composed of all engineering phases at incrementally maturing stages. So in early cycles, emphasis will be more on specifying requirements and less on design and implementation. The early makes of the system will be implemented as prototypes for testing to provide feedback for later cycles where requirements, design and implementation could be updated. This allows early detection of problems and results in more reliable and user-acceptable systems. Of course this approach mirrors how new technology (e.g. aircraft) has always been produced.

Similarly, Jakeman et al. (2006) have proposed an iterative development scheme for constructing environmental simulation models. This idea could be further developed by borrowing from the software engineering concept of 'test first development' (e.g. http://www.agilemanifesto.org; Jeffries et al., 2001). The concept is that when a new piece of functionality is required or a defect is found, the test should be written first before coding the implementation. The suite of tests then ensures that defects stay repaired and systems (in our case models) behave as they were intended.

Huth and Holzworth (2005) describe high-level reference tests as simulations that exercise the model under extreme situations. For example, a cropping system model might have reference tests that grow a crop under very low and high water and nitrogen scenarios, observing how stably the model performs under extreme scenarios. These tests look for extremes in behaviour, providing a level of robustness.

Sensibility tests are usually required to further evaluate model usefulness. Even though a model may be well calibrated to observations, rarely do the observed data cover the range of environments and scenarios that the model will be used for. Sensibility tests fill the gap in the observed data. Simulations are created for real-world scenarios, and the outputs are shown to 'experts' who provide a qualitative response as to the credibility of the results.

There are numerous other types of model tests that help assess the usability and reliability of a model. Simply having the tests, though, is insufficient. The process of using them is also critical. When a model is actively being worked on by several model developers, a fully-tested, calibrated model quickly loses its stable, tested state. To safeguard against this, automated testing, a concept borrowed from the Extreme Programming community, can be used to automatically test, compile and run all types of tests and compare results against known 'good' values. This protects against the 'trickle' effect where a change to source code in component A has an undesired impact on an apparently unrelated component B. It also helps to keep the model tested and calibrated while development proceeds. This approach is quite simple to implement with a series of batch files or scripts. This approach has successfully been adopted by the APSIM (Agricultural Production Systems Simulator) software development team (Keating et al., 2002; Huth and Holzworth, 2005). Looking to other disciplines can often bring many benefits to the world of model development.

2.3. Model Transparency and Dissemination

A key component of good modelling practice is being transparent in defining the model's purpose, its assumptions and formulations, and its evaluation. Such transparency should aid critical peer evaluation of the model and its applications, and, potentially, its re-use in new applications if appropriate. Some relevant issues are described below.

2.3.1 Terminology

A common understanding and use of model terminology is required for communication of model development and evaluation to others. It is a key aspect of any attempt at model transparency. Terminology is used in describing (i) model structure, (ii) model parameterisation and (iii) model evaluation. Careful selection of terminology is required in all three of these areas.

While it may be unrealistic to expect a unified terminology to be adopted, a greater awareness of the origins of terms and how they are used in other fields is desirable. Much modelling theory and many applications of this theory have arisen in the statistical literature and, where possible, we would urge that modellers use this original terminology. As an example source of statistical terminology, see the glossary in Ripley (1996).

We will not recommend that best modelling practice should require adherence to any unified definitive terminology, which is close to impossible. It seems likely that model developers and model users will continue to enjoy confusing one another with terms such as 'validation,' 'verification,' 'stable,' 'dynamic,' 'state,' 'parameter,' etc. for the foreseeable future. However, best modelling practice must *require* that terminology used is fully defined in each case.

2.3.2 Reporting

Models should be formally reported in some way, and this should include:

- the 'mathematical' formulation and the assumptions on which it is based (ideally complete enough to allow the model's re-implementation);
- the model's parameterisation and parameter values;
- the model's implementation as appropriate, including operating instructions;
- the analysis undertaken to evaluate the model.

The principle of this is not controversial, but in many cases such documentation is incomplete. Elements of it may appear in reports to sponsors, perhaps in the peer-reviewed literature, but typically such reports are not much more than summaries. To address this, journals are increasingly starting to provide and/or require reported models to be deposited in an on-line repository of some kind. Such efforts are outlined below in more detail.

As discussed above, environmental models should undergo continuing evaluation and revision, and this in turn should be reported, with effective version control. So, for an active model, neither the model nor its documentation is ever definitive.

The main constraint to good reporting of models is the time and resources required, and while the inefficiency associated with undocumented model development is clear, this is difficult to convert into resources for model formulation.

To a large extent models can be self-documenting; for example, while perhaps not ideal the computer code (or equivalent) is at least an explicit representation of the model's formulation although not of its correctness or underlying thinking. Of course the code may not always be transparently available. Some model development packages lend themselves to developing 'self-documenting' models and such technological developments may ease the effort required for good model reporting.

2.3.3 Model dissemination

Good modelling practice should include learning from previous work, but how can this be achieved, knowing that methodologies, complexity and structure of model development vary greatly?

Various initiatives have tried to support model re-use by setting meta-data, documentation systems and meta-database standards that include sufficient information to search models and assess them for scientific questions. For example, Hill et al. (2001) published a Content Standard for Computational Models (CSCM), which led to the Register of Environmental Models (REM), an operational database providing meta information on different models for environmental processes (Benz and Knorrenschild, 1997; Benz et al., 2001; Hoch et al., 1998). This register is now available as part of the ECOBAS WWW server (Benz, n.d.), which is an information system for documenting the mathematical formulations of ecological processes. The objective of the ECOBAS WWW server is to provide easy access to available information about ecological models, including the limits of validity wherever feasible, in a standardised manner that is comparable between and transferable to different applications. ECOBAS seeks to facilitate the reuse of models by breaking up complex models into subcomponents that may be used to build new models. To facilitate this modularisation, the documentation standard ECOBAS_MIF was designed. This standard provides a set of metadata attributes that define the structure and syntax of model documentation. Using the ECOBAS_MIF, modellers can describe and advertise their model through an online entry form.

A comparable register is the EPA's CREM Models Knowledge Base (Council for Regulatory Environmental Modeling, n.d.), a web-based inventory of environmental models, which may serve as a central repository, facilitate model selection, and provide pointers to the home pages for individual models. The contents of each model record are intended to include the types of information recommended by the Draft Guidance for Environmental Models, beneficial to prospective model users. Each model's record includes three pages of information: the "General Information" page includes an overview of the model, contact information, and a link to the model's homepage; the second page, "Model Use," provides essential information for potential users, including technical requirements (hardware, operating systems, and software), directions for obtaining (downloading) the model, and basic information on using the model (model inputs, model outputs, and the User's Manual and Technical Guide); and the final page, "Model Science," includes sections on the conceptual basis of the model, scientific detail, model framework, and model evaluation studies and peer reviews.

2.4. A Definition of Good Modelling Practice

As outlined earlier a number of authors have previously issued guidelines on good modelling practice, albeit sometimes in specific domains. The list here is not very different from those previously suggested, although we perhaps have been more general, recognising that any guidelines need to accommodate a wide range of different types of application.

Good modelling practice at least includes:

- A clearly specified purpose.
- Clearly specified use of data.
- Explicitly stated assumptions and model formulation.
- Ongoing model evaluation, recognising the difference between:
  - evaluating model assumptions;
  - evaluating model implementation;
  - evaluating model performance.
- Transparent reporting.

Working to the standards of best modelling practice is the responsibility of model developers. However, even a model developed under best practice may not be fit for a given purpose. It is the responsibility of users of a model to be aware of its capabilities and to use it appropriately.

What needs to be done to move towards good modelling practice?

2.5. Progress Towards Good Modelling Practice

Is there any evidence that progress is being made towards best modelling practice? To investigate this, a crude survey was undertaken to examine whether there is any evidence that, as a community, we are undertaking more work which might be classified as model evaluation as opposed to primary model development. As we have made clear above, we do not regard model evaluation as the only important component of best modelling practice, but it was judged that its occurrence in the literature would be the easiest to test.

Title–Abstract–Keyword searches were carried out for the ten-year period 1997–2006, including articles in press through to August, 2006. The searches were performed for four contrasting journals: Environmental Modelling and Software, Ecological Modelling, Journal of Hydrology, and Mathematics and Computers in Simulation. Searches were performed, using the Scopus database with the phrases: "model"; "model" AND "sensitivity analysis"; "model" AND "parameter uncertainty"; "model" AND "model structure"; "model" AND "model testing"; "model" AND "model verification"; "model" AND "model validation."

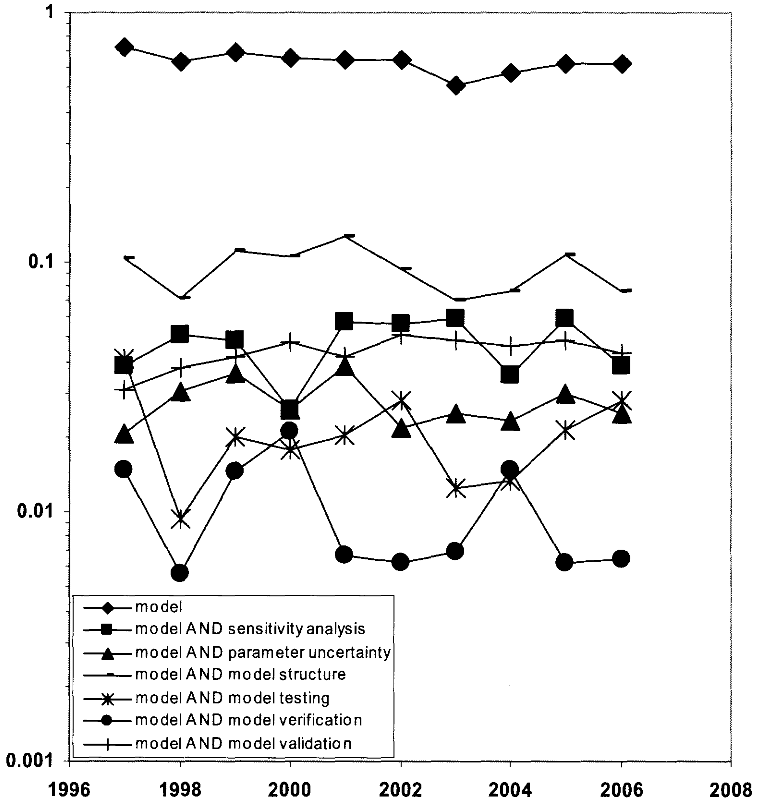

Figure 2.2 Proportion of occurrences of the term "model" and proportion of co-occurrences of the term "model" and terms indicating good modelling practice is being conducted. Denominator is the total number of journal articles. Numbers are pooled for Environmental Modelling and Software, Ecological Modelling, Journal of Hydrology, and Mathematics and Computers in Simulation, for 1997–August 2006. (Note the *y*-axis log scale.)

Co-occurrence of the term "model" and one of the other terms associated with model development (e.g. "sensitivity analysis," "parameter uncertainty," etc.) is taken to indicate that some form of model evaluation is being conducted.

Results for the four journals are pooled and presented as a fraction of the total number of journal articles for each year in the period (Figure 2.2). Trends in the time-series data were determined by simple linear regression. The fraction of journal articles containing the term "model" shows a slight downward trend. There is a small increase in the (absolute) trend of occurrences of the term "model" accompanied by occurrences of the terms "sensitivity analysis" and "model validation" and a smaller increase (relative to the proportion of occurrences of "model") of co-occurrences with the other terms in the list of Boolean searches. The most common model development aspect listed involves "model structure," followed by "sensitivity analysis" and "model validation." As a general conclusion, this analysis indicates that our interest in model evaluation work may be increasing, but only at a slow rate.

Our *implied* conclusion is that, as a community, our intentions with regard to good modelling practice are better than our deeds. While this suggests we are fairly typical examples of our species, it is not encouraging for the development of environmental models as effective tools for policy makers and planners. Of course we must emphasise that this conclusion is based on a very limited analysis of the literature, albeit an analysis which accords quite well with the professional experience of quite a large group of environmental modellers.

2.6. Recommendations

We have made some suggestions as to what constitutes 'good modelling practice.' The details are always likely to be the subject of lively debate, but the general components of this 'good modelling practice' are probably not controversial (clear purpose; adequate reporting; serious evaluation).

We have indicated some areas where current work seeks to move the process of model evaluation forward from a simple measure of performance (even a complex measure of performance) to an assessment of how performance relates to the model assumptions and formulation. Such developments are probably important; however they are academic if the community at large is not routinely as engaged with model evaluation as it is with primary model development.

We have reported a crude analysis which suggests that progress towards improving modelling practice is slow. This is despite very widespread agreement on what constitutes good practice. Why is this so?

In the research community at least, the drivers for model development and evaluation are funding and publication. If, as we think, modelling practice warrants improvement, sponsors and journals will need to take a lead in creating an environment where developing a model requires that the work be performed under some system of good modelling practice. The suggestion has been made of a 'good practice check list' in the Journal of Environmental Modelling and Software. While such a system would need to be flexibly applied, the principle is sound, and such steps should move us forward.

REFERENCES

Asgharbeygi, N., Langley, P., Bay, S., Arrigo, K., 2006. Inductive revision of quantitative process models. Ecological Modelling 194, 70–79.

Bankes, S., 1993. Exploratory modeling for policy analysis. Operational Research 41 (3), 435–449.

Beck, M.B., 1987. Water quality modeling: A review of the analysis of uncertainty. Water Resources Research 23 (8), 1393–1442.

Beck, M.B. (Ed.), 2002. Environmental Foresight and Models. Elsevier, Amsterdam.

Benz, J., n.d. WWW—Server for ecological modelling. University of Kassel and the GSF—National Research Center for Environment and Health. Internet: http://eco.wiz.uni-kassel.de/ecobas.html.

Benz, J., Knorrenschild, M., 1997. Call for a common model documentation etiquette. Ecological Modelling 97 (1–2), 141–143.

Benz, J., Hoch, R., Legovic, T., 2001. ECOBAS—Modelling and documentation. Ecological Modelling 138, 3–15.

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

Beven, K., 2002. Towards a coherent philosophy for modelling the environment. Proc. R. Soc. Land. A 458, 1–20.

Cox, G.M., Gibbons, J.M., Wood, A.T.A., Craigon, J., Crout, N.M.J., 2006. Towards the systematic simplification of mechanistic models. Ecological Modelling 198 (1–2), 240–246.

Council for Regulatory Environmental Modeling, US Environmental Protection Agency, n.d. Council for regulatory environmental modeling—Models knowledge base. Internet: http://cfpub.epa.gov/crem/knowledge_base/knowbase.cfm.

Friedman, A.M., 1953. On the Methodology of Positive Economics. University of Chicago Press, Chicago, pp. 3–43.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. Water Resources Research 34 (4), 751–763.

Gupta, H.V., Bastidas, L., Sorooshian, S., Shuttleworth, W.J., Young, Z.L., 1999. Parameter estimation of a land surface scheme using multi-criteria methods. GCIP II Special Issue. Journal of Geophysical Research – Atmosphere 104 (D16), 19491–19503.

Gupta, H.V., Beven, K.J., Wagener, T., 2005. Model calibration and uncertainty estimation. In: Anderson, M.G., McDonnell, J.J. (Eds.), Encyclopedia of Hydrological Sciences. John Wiley and Sons Ltd., Chichester, UK, pp. 1–17.

Hill, L.L., Crosier, S.J., Smith, T.R., Goodchild, M., 2001. A content standard for computational models. D-Lib Magazine 7 (6). Online available at http://www.dlib.org/dlib/june01/hill/06hill.htm.

Hoch, R., Gabele, T., Benz, J., 1998. Towards a standard for documentation of mathematical models in ecology. Ecological Modelling 113, 3–12.

Huth, N., Holzworth, D., 2005. Common sense in model testing, in: MODSIM 2005 Proceedings, International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005.

Jacobson, I., Booch, G., Rumbaugh, J., 1999. The Unified Software Development Process. Addison-Wesley, Reading, MA.

Jakeman, A.J., Post, D.A., Beck, M.B., 1994. From data and theory to environmental model: The case of rainfall runoff. Environmetrics 5 (3), 297–314.

Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21 (5), 602–614.

Jeffries, R., Anderson, A., Hendrickson, C., 2001. Extreme Programming Installed. Addison-Wesley, ISBN 0201708426.

Keating, B.A., Gaydon, D.S., Huth, N.I., Probert, M.E., Verburg, K., Smith, C.J., Bond, W.J., 2002. Use of modelling to explore the water balance of dryland farming systems in the Murray-Darling Basin, Australia. European Journal of Agronomy 18, 159–169.

Lane, S.N., 2004. Wavelet-based evaluation of rainfall-runoff models. Eos Trans. AGU 85 (17). Joint Assembly Suppl., Abstract H23F-04.

Larman, C., 2002. Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and the Unified Process, 2nd ed. Prentice-Hall, Upper Saddle River, NJ.

Morton, A., 1990. Mathematical modelling and contrastive explanation. Canadian Journal of Philosophy 16, 251–270.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I—A discussion of principles. Journal of Hydrology 10 (3), 282–290.

Norton, J.P., 1975. Optimal smoothing in the identification of linear time-varying systems. Proc. IEE 122 (6), 663–668.

Oreskes, N., 2000. Why Predict? Historical Perspectives on Prediction in Earth Sciences. Island Press, Washington, DC; Covelo, California, pp. 23–40.

Pielke, J.R.A., 2003. The role of models in prediction for decision. In: Canham, C., Lauenroth, W. (Eds.), Understanding Ecosystems: The Role of Quantitative Models in Observations, Synthesis, and Prediction. Princeton University Press, Princeton, NJ, pp. 113–137.

Refsgaard, J.C., Henriksen, H.J., 2004. Modelling guidelines-terminology and guiding principles. Advances in Water Resources 27, 71–82.

Ripley, B.D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press, UK.

Sarewitz, D., Pielke, R.A.J., 2000. Prediction in science and policy. In: Sarewitz, D., Pielke, R.A.J., Byerly, R. (Eds.), Prediction: Science, Decision Making, and the Future of Nature. Island Press, Washington, DC, pp. 11–22.

Wagener, T., 2003. Evaluation of catchment models. Hydrological Processes 17, 3375–3378.

Wagener, T., Kollat, J., 2007. Visual and numerical evaluation of hydrologic and environmental models using the Monte Carlo Analysis Toolbox (MCAT). Environmental Modelling and Software 22, 1021–1033.

Wagener, T., Wheater, H.S., Gupta, H.V., 2003. Identification and evaluation of watershed models. In: Duan, Q., Sorooshian, S., Gupta, H.V., Rousseau, A., Turcotte, R. (Eds.), Calibration of Watershed Models. AGU Monograph, pp. 29–47.

Young, P.C., 1998. Data-based mechanistic modeling of environmental, ecological, economic, and engineering systems. Environmental Modelling and Software 13, 105–122.