Jul 13th, 9:30 AM - 9:50 AM

Modelling for reproducibility, transparency, and continuous learning: Revisiting Jakeman's 10 steps

Thorsten Arnold
Grey Bruce Centre for Agroecology, horsten.arnold@gbcae.com

International Environmental Modelling and Software Society (iEMSs)
8th International Congress on Environmental Modelling and Software
Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2016-proceedings

Modelling for reproducibility, transparency, and continuous learning: Revisiting Jakeman's 10 steps

Thorsten Arnold
Grey Bruce Centre for Agroecology, thorsten.arnold@gbcae.com

Abstract: Numerical modelling has become the standard tool to evaluate vast amounts of data, both for the public management of natural resources by government agencies and to understand environmental interactions within academia. While an ever-increasing amount of software tools is being developed and approaches are formulated how modellers should interact with stakeholders, the modelling community has not yet addressed criteria how these tools and approaches are applied cost-effectively. Across the board, public agencies and academia continue to experience budget and time overruns that are systematic and need to be addressed if environmental modelling shall provide the benefits that modellers claim. This paper discusses public sector requirements such as reproducibility and transparency, lays out current approaches and practical challenges, and suggests a framework for organizing the modelling process that is derived both from public sector agencies and from integrated environmental modelling in academia. This paper identifies a multi-step layout of modelling studies (e.g. following Jakeman 2006) as core shortcoming, because project managers are mislead into translating these steps directly into project proposals, Gant charts, and budgeting. Authors always acknowledge that in practice, projects will have to revisit earlier steps in order to correct or refine assumptions, which basic project planning has not been accounting for adequately. With linear, multi-step project planning approach to a task that is indeed iterative or circular, extra costs arise from access to knowledge, inadequate design of software tools and intellectual property rights to these tools that often don't foster iterative work, and inadequate workflows. This paper suggest an alternative, circular framing to modelling projects and lays out requirements for the design of software tools, intellectual property rights, and the roles of knowledge and staff.

***Keywords**:* Model Management, integrated modelling, environmental modelling, cost-effectiveness, project planning.

1 THE NATURE OF THE BEAST

There is little dispute that numerical models are the best suited tool for evaluating vast amounts of observations on environmental systems and forecasting into the future. Fundamentally, quantitative modelling takes the spatial information of GIS software, plus the temporal information of time series tools, into a continuous temporal-spatial description of all variables relevant to describe an environmental system. While observation data can only quantify small patches in this data continuum, quantitative models can interpolate between the patchy data points in a manner that coheres to physical principles and science-based mechanisms, allowing modellers to extrapolate into the future while assessing the uncertainty of their forecasts. Not unlike geospatial interpolation, models use a selection of basic scientific principles, their task-dependent simplifications into mathematical equations, assumptions about the system state and its external forcing based on expert judgement, and numerical routines that have quantifiable inaccuracies. However, the conceptual complexity of modelling is much greater than of standard GIS-based technology: aleatoric uncertainty (e.g. chaos) sets principle limits to deterministic analysis and epistemic uncertainty is inevitable in any simplified system (Walker et al., 2003, 2013, Brugnach et al., 2008). But modelling is also procedurally complex: stakeholders often do not understand the technical language used in modelling and are overwhelmed by the amount of technical detail, such that modellers easily step out of their role as facilitators in a knowledge generation process and take ownership in political decisions, thus imposing opinion in the disguise of a "science-based" model (Pahl-Wostl et al, 2007). Furthermore, knowledge about a system is dispersed over the brains of disciplinary trained professionals who work across specialized departments with inconsistent mandates, timelines, and success criteria for individuals (Arnold 2013). Finally, many public agencies rely on external partners such as consulting firms or academics to fulfil technical and conceptual tasks of the modelling process. These partners have yet another set of objectives, timelines and needs.

Compared to GIS technology, which is a well-established tool in public agencies, modelling remains poorly utilized and institutionalized. Regular budget overruns and difficulty to maintain timelines, are common occurrences, as are studies that are shelved and underutilized. Given the urgency of managing natural resources in respond to global change, the perceived potential of modelling as a tool, and the procedural complexities and limitations listed above, what approaches can public agencies take to use models in ways that are better projectable and more cost-effective? Unfortunately, the scientific literature provides almost no guidance here, and even gray literature is only emerging (e.g. ASTM, Rassam et al., 2011).

This paper first summarizes current and emerging conceptual and technical approaches to modelling from a workflow perspective. Then, Jakeman's 10 steps of a model are listed and shortcomings are pointed out from a procedural perspective. A cyclical framework to modelling is proposed and preconditions and opportunities are discussed.

2 ENVIRONMENTAL MODELLING AS AN ORGANIZATIONAL PROCEDURE

2.1 Some definitions

An *environmental model site application* is one or more computer models that are set up to simulate a particular location and system using observation data. *Modelling* is the process of developing and analyzing a model site application. A *model code* is a software package (or several couple ones) that perform core tasks of a computer simulation, generally combining input files, scientific laws and assumptions in order to create output files. In addition, modelling requires (1) *observation data, (2) preprocessing*, which are more manipulations of observation data in order to create input files, and (3) *output post-processing*, which transforms the output of one or many model runs into the actual results with policy implications. Both pre- and post-processing of data utilizes additional scientific laws and principles and requires data translations that change information content through technical or conceptual assumptions. Finally, model uncertainty is the combined impact of all assumptions, omissions and errors in this process on study results.

Model site applications are increasingly documented as scientific workflows, which are sequences of operations that transforms observation data into model results (e.g. Billah et al., 2016 Leonard and Duffy 2014 and 2016, D'Agostino et al., 2015). A workflow can be described in text form, e.g. as part of model documentation. Automated forms are either a script or a workflow management system that offers a menu-based user interface. Automated workflows are essential for designing and distributing multi-run modelling experiments on computer clusters or the cloud, for example for sensitivity and uncertainty analysis. Other advantages of automated workflows are manifold: assumptions can be varied systematically, changes to system conceptualization or methodological choices can be implemented and tested separately, updating of a model with new information is relatively simple and rapid, reproduction of result for quality assurance is provided, and general transparency of the modelling process is enhanced. However, initial set-up of automated workflows also requires skill and time, and commercial modelling software providers increasingly offer guided user interfaces without necessarily providing options to convert these “manual” clicks into a workflow script.

Workflow solutions can be characterized by their respective resource needs for (1) initial set-up, (2) executing a single or n repetitions of the workflow, and (3) implementing changes (assumptions, data, methods, or any other operation component). Resources include capacity (conceptual system knowledge, modelling knowledge, IT skills), staff time, computation facilities, and fees for using licensed software.

2.3 Multi-step descriptions of a modelling process and procedural challenges

A milestone publication by Jakeman (2006) sub-categorizes the modelling process in ten steps. Many environmental disciplines and fields of engineering applications rely on Jakeman's 10-step approach or a comparable list of modelling steps (e.g. Crout et al., 2008, Robson et al. 2008, Blocken et al. 2012):

1. Definition of the purposes for modelling
2. Specification of the modelling context: scope and resources
3. Conceptualisation of the system, specification of data and other prior knowledge
4. Selection of model features and families
5. Choice of how model structure and parameter values are to be found
6. Choice of estimation performance criteria and technique
7. Identification of model structure and parameters
8. Conditional verification including diagnostic checking
9. Quantification of uncertainty
10. Model evaluation or testing (other models, algorithms, comparisons with alternatives), a formulation where Jakeman acknowledges that the commonly used term 'validation' is "rarely possible (or perhaps even appropriate)" for environmental models.

Authors who propose this or similar lists of modelling steps consistently point out that the outcomes of later steps (7 and upwards in Jakeman's case) may require revising and refining earlier steps, including the conceptualization of the system and its boundaries, the level of simplification and choice of scale, the conceptual assumptions and methodological choices within the process. From a project planning perspective, such a stepwise approach to modelling poses fundamental challenges, especially because more than one person are involved in this process, even multiple organizations:

1. What happens if later modelling steps require revising earlier steps?
2. What are resource implications of decisions in Step 3-6 for later steps 7-10?
3. Are those project partners involved in decisions in Step 3-6 aware of resource implications for those partners who implement later steps 7-10?

As a consequence of these shortcomings, modelling studies are not designed for the iterations of revisiting earlier steps:

- Many *workflow operations* are performed "manually" by clicking, copy-pasting, and menu-based call of routines. When revisiting any operation in the workflow, all subsequent manual operation require staff time for implementation and verification, access to specialized knowledge, at best the same person who performed it the first time. If project managers originally project few or no revisiting, then software design is skewed toward manual workflow execution but ultimately constraints revisits, updating, and uncertainty analysis.
- *Access to all relevant knowledge* is required when revisiting earlier modelling steps, especially when changing system conceptualization or removing technical errors. If considering revisiting and reuse of models, managers would favour workflow solutions that minimize the need for IT skills and other technical knowledge, transparently document the knowledge that went into the model, and ensure that project partners are available to refine a model later.
- *Licensing and user fees* often protects intellectual property of software. Short-term rental of intellectual property is often the most cost effective approach to accessing software, especially if project management assumes a linear modelling process that is performed one time only. However, software license costs can become a barrier to revising earlier steps, and also can constrain uncertainty analysis that runs the software on multiple processors on clusters or the cloud.

Without addressing the fundamental planning questions, it is no surprise that almost all modelling processes exceed allocated budgets and timelines and/or results fall behind initial expectations. In the context of public decision making, which generally falls into annual budgeting cycles and regulated timelines of planning procedures, such procedural shortcomings limit the applicability of models as policy tools.

Examples are manifold: at the planning and funding stage, project managers follow the ten-step approach and feel obliged to play down potential costs for revisiting and refining model assumptions. Also, decisions taken during the steps 3 to 6 may further increase overall resource needs and impact the timeline of deliverables. In the end, core elements of steps 9 and 10 (uncertainty quantification, comparison of model against alternatives) are hardly ever accomplished for public decision making because resources that were initially allocated have run out. Without adequate uncertainty analysis, however, the scientific weight of model results remains weak and undermine the modelling purpose of elevating natural resources management to more objectivity. Furthermore, model site applications are seldom designed for reusability or revisiting earlier steps. Many applications are thus lost and never updated.

3 A CYCLIC PERSPECTIVE ON MODELLING PROJECTS

3.1 A different perspective on modelling tasks

Instead of following the set of technical procedures of a modelling study, an alternative perspective is proposed that organizes a modelling process along the type of knowledge involved, and the character of decisions in each step. Environmental model site application is proposed to be organized along five phases:

1. ***Defining modelling goals****:* either through regulations or by considering multiple perspectives from a diverse group of stakeholders. This phase also specifies available resources (staff time, budget) and potential future uses of a site model.
2. ***The conceptualization phase of a system of interest*** responds to the question: What is the shared understanding about a system and its boundaries with respect to the policy goal? What are the driving forces on this system, and what are the response mechanisms? Which system elements and feedbacks need to be represented and which ones are not?
3. ***The technical implementation phase*** is concerned with how a system conceptualization is transformed into computer code in a manner that is consistent with scientific knowledge and good modelling practices, and how the computer code is then evaluated in order to create results. The technical implementation phase responds to a different type of the fundamental question: What tools, theories, and methods are available to create a computer simulation from the conceptualization?
4. ***Quality control (QA/QC)*** is a mechanism that independently corroborates goal setting (Are we asking the right questions?), system conceptualization (Is the system conceptualization adequate in order to achieve the modelling goal?) and its methodological and technical implementation (Is the study based on sound methods, implemented correctly, and does it follow good modelling practices?).
5. ***Communication with decision makers*** feeds modelling results back into the policy cycle.

Technical implementation should not require any additional assumptions on how the simulated system functions. Tasks are mainly performed by modellers and may involve software developers and professionals qualified to make methodological choices. QA/QC affirms adequate use of methods through independent academic or consultant review, which neither requires location-specific knowledge about environmental conditions nor awareness of the political context. Errors are eradicated by using diagnostic tests and standard datasets, or by relying on well-established software.

System conceptualization deals with questions whether the model captures the relevant elements of a system, and whether modelling results support recommendations. It encompasses Jakeman's step 3 (hypothesis formulation) and additionally a test whether, upon implementation and execution of the model, these hypothesis are supported by results. Ultimately, conceptualization assesses whether the recommendations derived by a model have relevance for the stated modelling purpose. Tasks are mainly performed by local experts and practitioners with knowledge of the study system (e.g. engineers, field scientists). Collaboration with modellers ensures consistent language and coherent concepts. QA/QC on adequate system conceptualization requires deep location-specific knowledge from multiple disciplines and extensive uncertainty analysis.

Goal setting is performed by policy makers or regulators who specify the objective of a modelling study. In practice, these individuals may have limited or no modelling knowledge. QA/QC on adequate goal setting is generally ensured through public consultation, which ensures that sufficient local knowledge is at the table to identify conceptual shortcomings. The integration of local experts and modellers can ensure that goals are realistic.

3.2 Modelling as three embedded iterations

Within the paradigm of adaptive management in response to changing environmental circumstances and evolving social needs, policy making is understood as a responsive and self-correcting process.

Within the adaptive management process, modelling plays a core role to assess the status quo and forecastinto the future. For each iteration of adaptive management, conceptual understanding about a system is updated to reflect new information and knowledge gains. With each new conceptual hypothesis, changes need to be implemented, evaluated, and verified in the model. Then, a conceptual hypothesis can be corroborated or rejected against data, and other hypothesis are formulated conceptually, then implemented into the model which is re-evaluated and diagnostically tested. This model-based learning and policy support can be visualized as three cycles that are embedded into each other. Each operates at a different temporal scale, requires different skills and knowledge and different quality control mechanisms.

- The *policy cycle* encompasses the identification of policy goals for a system, the assessment of its current and extrapolated status (e.g. using numerical models), designing and implementing intervention policies, monitoring changes of the system and attributing some of those to policies, and the adjustment of policies and goals. Participants require in-depth understanding on governance and policy design and implementation, and some knowledge about the system to be governed.

- The *conceptual learning cycle* encompasses the system conceptualization in the form of hypothesis formation, and hypothesis testing. These two aspects frame the technical implementation of the model, and are mainly concerned with reviewing, discussing and eventually supporting intermediate modelling results. The technical phase of modelling is “wrapped” into a grey box, in order to maintain the focus on systemic questions. Participants require in-depth local knowledge and expertise about the system at hand, but superficial modelling knowledge suffices.

- The *technical modelling cycle* contains all methodological choices on how to implement the system conceptualization, the implementation, diagnostic testing and ultimately uncertainty quantification. Participants require in-depth modelling knowledge and IT skills, and training in the disciplines they are modelling.

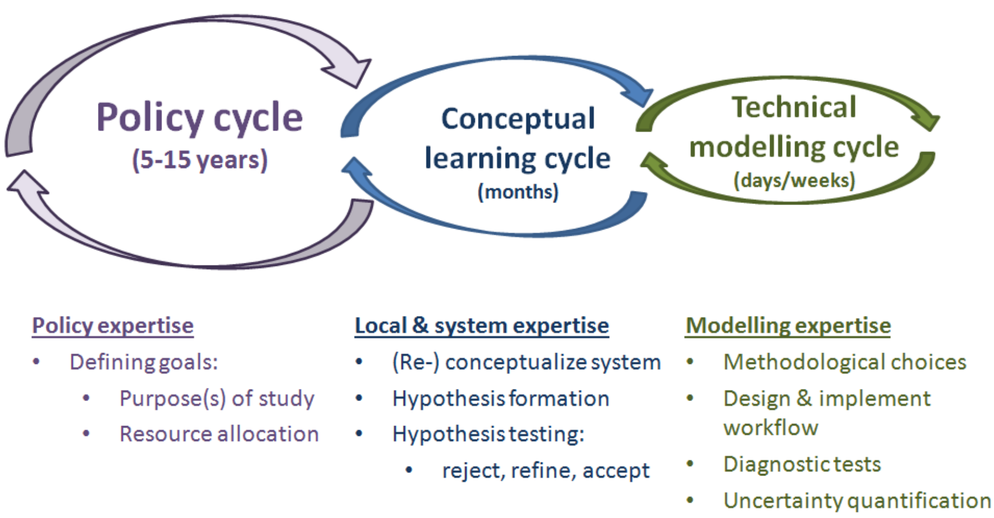

Figure 1: The three embedded circles of modelling-informed policy making

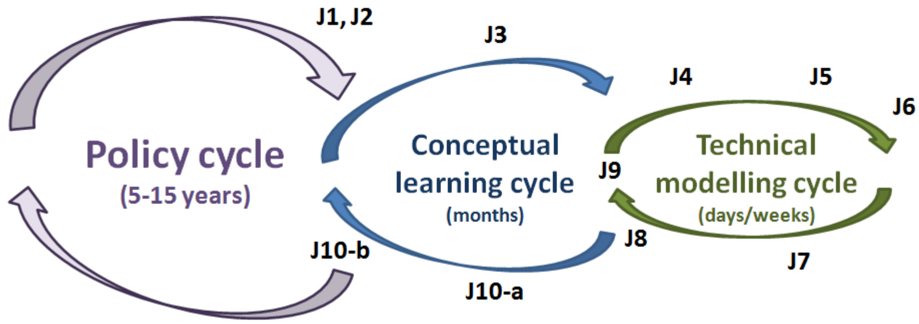

Figure 2: The three embedded circles and Jakeman's steps (J1 ~ Jakeman's step 1)

3.3 Three embedded iterations in comparison with Jakeman's steps

In general, Jakeman`s steps correspond well with the proposed phases of goal setting, conceptualization, technical implementation and quality control. The main difference is the ability to organize steps as an iterative process, and the central role of quality control.

The outset of any modelling study is the definition of modelling goals as part of the policy cycle: the purpose (J1), the modelling context and resources (J2). Then, information about a system are gathered and causes and effects are hypothesized and formulated into a system conceptualization (J3). As part of the technical cycle, model code requirements are listed (J4), a model approach and code is selected (J5), methods for structure and parameter identification are chosen (J6), as well as criteria for evaluation performance (J7). Model structure is then implemented and parameters identified, and diagnostic checking is performed (J8). Implementation proceeds until no more errors can be identified (J7, J8). Then, uncertainty analysis is performed (J9). Because it is not possible to validate a model per se, local experts are included in the decision whether a model simulates the system reasonably well, or whether other conceptual changes are required (J10). If the latter is the case, then alternative conceptualizations are proposed (J3) and implemented and evaluated (J4-J9). Ultimately, partners responsible for system conceptualization decide whether the simulation meets the defined modelling goal, or whether this goal is not achievable with given resources (J10).

While all of Jakeman's steps fall into the three iteration cycles, this perspective does not encompass "surprises" like returning to earlier steps of the linear sequence. Instead, the iterative nature is understood as an inherent element of any modelling study and elevated to the core design principle for the study, whereas iteration are no longer shortcomings of the modelling team.

3.4 Design implications from the perspective of three embedded iterations

For project planning, this change in perspective has several implications as it redefines the roles of project partners and staff at various points in the process, as well as cost considerations across the project:

- With automation of technical tasks, the modelling process *refocuses attention on the conceptualization* of the system and hypothesis testing and away from technicalities that are difficult to comprehend outside of the modelling community.
- *Modellers play a role in both conceptualization and technical modelling*. Their main role is their implementation of the technical modelling cycle: selection of model code and choice of methods, model set-up and eventual code implementation, identification of model structure and parameters, verification and diagnostic, and ultimately designing and implementing uncertainty analysis. However, modellers also participate in the conceptual learning and hypothesis building: they must ensure consistency between expectations and technical feasibility, contribute expertise. Modellers may experiment with conceptual hypothesis

themselves in order to alleviate the workload of the conceptual team and speed up the overall modelling process.

- *Quality control mechanisms are refined* by distinguishing technical, conceptual, and goal setting phases that require different QA/QC mechanisms and knowledge. Access to knowledge for model evaluation ("validation") and quality control (J8, J10-a and J10-b) are build into the project design.
- *Knowledge management* can be organized along the three distinct cycles, which disentangles the amount of knowledge expected from each participant. The cycles also highlight that prolonged access to knowledge needs to be considered in the planning and design stage of a project, because repeated need for accessing knowledge is to be expected.
- The iterative perspective *shifts costs and benefits for automating workflows* (scripts or workflow management systems). Whereas the linear perspective leans toward manual modelling at least until uncertainty assessment is performed, the iterative approach favours automation from the beginning on. This shift fundamentally changes how modellers or modelling consultants need to design their software and deliverables.
- The shift toward automated workflows also poses an additional criterium for the *selection of model code and data processing tools*. Software that can only be operated "manually" through menu-based user interfaces is incompatible with automation. However, many software packages (e.g. ArcGIS, QGIS, MatLab, R, SQL databases, and most model codes) can be called by command line as well as through user interfaces. Indeed, user interfaces can be used to generate automation scripts.

5 CONCLUSION

It is hoped that this paper contributes to improving the design of numerical modelling studies in order to improve the management of natural resources, which are mostly hosted by public sector organizations and financed through public money.

For any managers of modelling projects, the iterative or cyclic nature of knowledge generation in modelling is nothing new (compare Scholten and Refsgaard 2010, Rassam et al., 2011), However, this iterative nature has not been translated into design guidelines for modelling projects, and public agencies are repeatedly failing to keep modelling projects within budget and timing requirements.

This paper aims at re-opening a debate on the milestone paper of Jakeman (2006), in order to address several procedural challenges that his proposed sequence of steps does not address. While pointing out the requirement of returning to an earlier step, Jakeman`s approach does not address aspects of project planning and design. This paper proposes a circular, iterative perspective that is consistent with Jakeman`s steps but assumes that these executed repeatedly. This shift in perspective raises planning considerations, such as knowledge management, software design, the role of conceptual hypothesis testing, and quality control. By re-opening a debate, modellers may refine the relationship between technical and conceptual modelling tasks, further specify deliverables for each phase, and hence redefine the engagement between public agencies and external partners (academia, private-sector consultants).

ACKNOWLEDGMENTS

Much of the content of this work was inspired by experiences in academia (Arnold, Troost and Berger, 2015), working for the public sector (Arnold 2015), and as private consultant. Inspirations were drawen from discussions with Mason Marchildon and Steve Holysh.

REFERENCES

Arnold TR. Procedural knowledge for integrated modelling: towards the modelling playground. Environmental modelling & software. 2013 Jan 31;39:135-48.

Billah MM, Goodall JL, Narayan U, Essawy BT, Lakshmi V, Rajasekar A, Moore RW. Using a data grid to automate data preparation pipelines required for regional-scale hydrologic modeling. Environmental Modelling & Software. 2016 Apr 30;78:31-9.

Blocken B, Gualtieri C. Ten iterative steps for model development and evaluation applied to Computational Fluid Dynamics for Environmental Fluid Mechanics. Environmental Modelling & Software. 2012 Jul 31;33:1-22.

Brugnach, M., C. Pahl-Wostl, K.E. Lindenschmidt, J.A.E.B. Janssen, T. Filatova, A. Mouton, G. Holtz, P. van der Keur, N. Gaber (2008). Chapter 4: Complexity and Uncertainty: Rethinking the Modelling Activity, In: A.J. Jakeman, A.A. Voinov, A.E. Rizzoli and S.H. Chen, Editor(s), Developments in Integrated Environmental Assessment, Elsevier, 2008, Volume 3, Pages 49-68,

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., Whitfield, P. (2008). Good Modelling Practice. Developments in Integrated Environmental Assessment. Vol. 3, pp. 15-31. 2008

D'Agostino D, Danovaro E, Clematis A, Roverelli L, Zereik G, Parodi A, Galizia A. Lessons learned implementing a science gateway for hydro□meteorological research. Concurrency and Computation: Practice and Experience. 2015 Jan 1.

Jakeman, A.J.,Letcher, R. A. and Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models, Environ. Model. Softw., 21, 602-614, 2006

Leonard L, Duffy CJ. Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment. Environmental Modelling & Software. 2014 Nov 30;61:174-90.

Leonard L, Duffy C. Visualization workflows for level-12 HUC scales: Towards an expert system for watershed analysis in a distributed computing environment. Environmental Modelling & Software. 2016 Apr 30;78:163-78.

Pahl-Wostl, C., 2007. The implications of complexity for integrated resources management. Environmental Modelling and Software, Volume 22, Issue 5, 2007, Pages 561-569

Rassam, D. W., Ian Jolly, and Trevor Pickett. "Guidelines for modelling groundwater–surface water interactions in eWater source, towards best practice model application, eWater." Interim version 1 (2011)

Refsgaard J.C., van der Sluijs J.P., Hojberg A.L. and Vanrolleghem P.A. (2007). Uncertainty in the environmental modelling process - a framework and guidance. Environ. Model. Softw. 22(11): 1543–1556.

Robson BJ, Hamilton DP, Webster IT, Chan T. Ten steps applied to development and evaluation of process-based biogeochemical models of estuaries. Environmental Modelling & Software. 2008 Apr 30;23(4):369-84.

Scholten Huub and Jens Christian Refsgaard 2010. Quality assurance in model-based water management: Better modelling practices. In: Vanrolleghem, Peter A. (eds.). Modelling Aspects of Water Framework Directive Implementation. Volume 1. IWA Publishing, ISBN13: 9781843392231 . Published 30/04/2010, 260 pages.

Walker, W. and Harremoes P., Rotmans J., der Sluijs J.V., Asselt M.V., Janssen P. and von Krauss M.K. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 1: 5–17

Walker WE, Haasnoot M, Kwakkel JH. Adapt or perish: a review of planning approaches for adaptation under deep uncertainty. Sustainability. 2013 Mar 4;5(3):955-79.