19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Are environmental models transparent and reproducible enough?

M.G. de Vos[a], S.J.C. Janssen[b], L.G.J. van Bussel[c] , J. Kromdijk[d], J. van Vliet[a,] J.L. Top[e]
a Netherlands Environmental Assessment Agency (PBL),
b Alterra, Wageningen UR, Wageningen, the Netherlands
c Plant Production systems Group, Wageningen University, Wageningen, The Netherlands
d Greenhouse horticulture, Wageningen UR, Wageningen, the Netherlands
e VU University Amsterdam, Amsterdam, The Netherlands
Email: *P.O. Box 303, 3720 AH Bilthoven, the Netherlands [martine.devos@pbl.nl]*

Abstract: Environmental computer models are considered essential tools in supporting environmental decision making, but their main value is that they allow a better understanding of our complex environment. Environmental models are in fact applications of shared theories on how real-world systems are functioning. Just like the underlying scientific theories they need to be evaluated and discussed among peers. To allow proper assessment of the quality and suitability of environmental models peers should be able to trace their results and insights through the model structure to the underlying choices and assumptions. This ideal of model evaluation should take place in the peer-review process before their output and the analysis of their output are published in scientific journals, but is hardly ever realized in reality. It is hypothesized that, despite the numerous attempts to promote good modeling practice and extended peer review, the reproducibility and transparency of environmental models is limited in actual terms. To test the validity of this hypothesis we have reviewed publications, documentation and software of four environmental models. We analyzed to what extent this material provided insight in the model structure and the modeling process and to what extent model findings could be traced back to the underlying choices and assumptions.

All four models and their applications have been described in dozens of articles in peer reviewed journals. This indicates that these models and there results and insights are trusted and used. They can be understood as well-established models grounded in a scientific theory. However, in our study we found that for at least three of the models reviewers lack information to evaluate their quality or suitability. Neither can they ensure the reproducibility and transparency of the model results and insights, for a number of reasons. First, information on model design is scattered over several sources, which are to a limited extent freely and easily available to reviewers. Second, written model documentation does not provide a sufficient description of the modeled system. Third, written model documentation does not provide a sufficient description of the calculation of model results. Finally, results presented in scientific publications do not necessarily correspond with parameter values or equations in the model source code.

Our findings suggest that environmental models lack essential quality characteristics in terms of transparency and reproducibility. This raises the concern that they are being used in applications without respecting and discussing their underlying choices and assumptions. We identify three structural causes for this lack of transparency and reproducibility: (1) the size and complexity of environmental models, (2) a lack of incentives for environmental modelers to be to transparent in the modeling process and (3) the use of computers and the focus on computation and simulation instead of the descriptive side of modeling. We submit that openness in the modeling process can only be achieved with a general change of attitude. On the one hand model developers must become explicit and open about their choices and assumptions. On the other hand peers, stakeholders and journals must request openness and challenge these choices and assumptions. In an operational sense, computers and networks can be turned to their advantage by having them disseminate high-quality model descriptions using shared vocabularies.

Keywords: *Transparency, reproducibility, peer review, model evaluation, model quality, Good Modelling Practice*

1. INTRODUCTION

Environmental computer models are considered essential tools in supporting environmental decision making. They facilitate exploring the consequences of alternative policies or management scenarios (Jakeman et al., 2006; Schmolke et al., 2010b; van der Sluijs, 2002) which may form the basis for policy decisions that can have significant societal impact. But the main value of environmental models is that they allow a better understanding of our complex environment (e.g. (Jakeman et al., 2006; Oreskes et al., 1994; Rykiel Jr, 1996; Schmolke et al., 2010b)). Environmental models are simplified representations of reality (Rosenblueth and Wiener, 1945), they are in fact applications of shared theories on how real-world systems are functioning. Just like the underlying scientific theories they need to be evaluated and discussed among peers (Refsgaard and Henriksen, 2004). In that way we gain knowledge on our environment. We submit that unfortunately current practice of environmental modeling does not support such debate.

In the process of developing environmental models, modelers inevitably make choices and assumptions. They have to decide which processes and concepts to include and which to simplify or neglect (Jakeman et al., 2006; van der Sluijs, 2002). To allow proper assessment of the quality and suitability of these models peers should be able to trace model results and insights through the model structure to the underlying choices and assumptions. Reproducibility of model findings as well as transparency of both the model and the modeling process are crucial aspects in model evaluation (Jakeman et al., 2006; Risbey et al., 1996).

Environmental models are normally evaluated when their output and the analysis of their output are published in peer reviewed journals. The models are subjected to the scrutiny of experts in the same field by inspecting their behavior (Alexandrov et al., 2011). Ideally these experts assess corresponding models by tracing model findings back through the model structure to the underlying choices and assumptions. This ideal of model evaluation through the peer-review process is hardly ever realized in reality, as demonstrated by Schmolke, et al (2010b). One reason is that journal articles reporting modeling efforts focus on scientific originality rather than on model documentation (Alexandrov et al., 2011; Schmolke et al., 2010b). In addition, reviewers must base their judgments solely on the material that the authors have chosen to present and on how they present it. There is considerable variation in how this is done (Alexandrov et al., 2011). A large part of the choices and assumptions remains hidden in the model source code or in the minds of the modelers and has not been made explicit in model documentation (van der Sluijs, 2002; Villa et al., 2009). This has lead Funtowitz and Ravetz in (1990) to propose an extended peer review, in which not only the associated article but also additional materials such as models, data and scenarios are reviewed. However, reviewers do not have sufficient time or resources to conduct detailed evaluation of these models (Alexandrov et al., 2011).

Next to extended peer review, many authors advocate standardization of the modeling process to enhance transparency and reproducibility of environmental models, summarized to as Good Modeling Practice. They provide guidelines and frameworks for model development and evaluation (Gaber et al., 2008; Jakeman et al., 2006; Refsgaard and Henriksen, 2004; Rykiel Jr, 1996), model documentation (Schmolke et al., 2010b), model application (Risbey et al., 2005) and peer review of modeling projects in scientific journals (Alexandrov et al., 2011). Environmental models have an increasing influence on societal decision making on complex issues with potentially large impact (e.g. climate change, food security, biodiversity conservation, pollution). Therefore, transparency and reproducibility of environmental models are crucial quality criteria to enable the independent re-use of models. The importance of transparency and reproducibility is highlighted by the Climate Gate controversy, which seriously undermined the credibility of climate science, which is partly based on projections with computer models (Hickman, 2009).

This paper investigates the transparency and reproducibility of environmental models by evaluating a limited set of models that were re-used by researchers for their own research. It is hypothesized that the reproducibility and transparency of environmental models is limited in actual terms, even though good modeling practice and extended peer review are promoted. To test the validity of this hypothesis we have reviewed publications, documentation and software of four environmental models. We analyzed to what extent this material provided insight in the model structure and the modeling process and to what extent model findings could be traced back to the underlying choices and assumptions.

2. MATERIAL AND METHODS

We reviewed publications, documentation and software, and consulted the model developers of four environmental computer models: a photosynthesis model, a model to assess mitigation costs of climate change, a land use model and a crop growth model. All four models have been developed and run by scientists and their content and application have been described in dozens of articles in peer reviewed journals. Model results and insights are used by scientists and for two models these are specifically directed at policy makers to support decision making. One model is small, consisting of only a few equations, while the other three models are large with many equations, linked modules and different data sets, resulting in thousands of lines of source code.

In our review we adopted the perspective of a researcher who wants to familiarize her- or himself with the model, with the objective to apply it or to further develop it. In that process of model learning, we assessed the models on the four criteria, which can also be found in several guidelines on good modeling practice (Alexandrov et al., 2011; Gaber et al., 2008; Jakeman et al., 2006; Refsgaard and Henriksen, 2004; Risbey et al., 2005; Rykiel Jr, 1996; Schmolke et al., 2010b), i.e. the availability and usefulness of a description of 1) the conceptual model, 2) the calculation process, 3) the input data and results and 4) the source code.

3. RESULTS

Overall, each of the researchers spent a few months to investigate the model. The general feeling was that after this period they still did not understand all of it. In each of the aspects of documentation, conceptual model, calculation process, data and source code, limitations, errors and ambiguities were encountered. However, the researchers did not believe that the four models are poor models or that they are not suited for their job.

3.1. Documentation

All researchers studied peer-reviewed publications, like articles and book chapters, to become familiar with the models. These publications describe the model design, in terms of primary concepts, processes and equations included in the model, as well as model results and insights. However, for the three larger models information on the model design is scattered over several publications as these do not have space for an extensive model description. They rather focus on specific parts or applications of the model or the publications describe the general functioning of the model, without being specific or concrete. Two models are described in reports that provide more comprehensive and detailed information on the model design and are available (usually from model developers) but not peer-reviewed. Furthermore, the references to model documentation used in journal articles may be confusing as some papers refer to the applied model indirectly. They refer to other papers which eventually refer to grey literature reports or book chapters (see Figure 1). Working back through the literature is not a guarantee that information on the model design can be found easily, as the figure shows. The model version used in a journal article may also be different from the version described in referred documents.

3.2. Conceptual model

In one case, peer-reviewed publications contained clear and detailed descriptions of included concepts and processes. Studying these documents was sufficient to develop a good understanding of the modeled system. In the other cases written model documentation, i.e. peer-reviewed publications and gray literature reports, described the main concepts and processes, but were not sufficient to get a clear and complete picture of the modeled system. For these three models, a comprehensive list of assumptions, concepts and processes that describes the model as a simplified representation was lacking. In these cases the researchers also studied the source code and requested personally information from the model developers. Information from both sources was complementary to the written model documentation and considered very useful.

3.3. Calculation process

For none of the four models results from scientific publications could easily be reproduced or traced through the model structure to the underlying choices and assumptions. Two researchers discovered that some results presented in scientific publications did not correspond with parameter values or equations in the model source code. Additional parameters were found in the source code, but not described in the scientific publications, in one case. In another case, the equations as documented in the scientific publication could not reproduce the behavior of the model as documented in figures, until an error was corrected in the equations. Another researcher found that the process of setting scenario parameters for model runs was arbitrary and could not be based on scientific

theories as the scenario parameters could not be measured or derived from scientific literature. At the same time these scenario parameters had a large influence on model results. Furthermore, in three cases the written model documentation only presented a selection of the equations and mathematical descriptions included in the models. Studying, altering and running the source code by trial and error proved most useful in understanding the whole process of calculation of model results.

3.4. Data

The identity and source of most of the model input data of the four case study models are described in written model documentation. For the three larger models the input data are stored in dozens of files, which had been typically manually organized and adapted. The researchers had difficulties getting an overview of the process of data management. Input files for one model do not have any structure or naming to each of the data fields, which may cause confusion and mistakes in the calculation process.

3.5. Source code

One model consists of a set of equations and parameters described in peer reviewed publications, which the researcher translated into computer code. The other three models consist of thousands of lines of source code which were available to the researchers. Only one model was available as open source, while the other models were only available upon request and in one case, only partly. The three researchers spent a few months getting familiar with the source code of their models by studying, altering and running it, but they still did not understand all of it. Two researchers discovered that the code of their models contains functions and modules that have no clear documented function in the calculation process, while still being included in model runs.

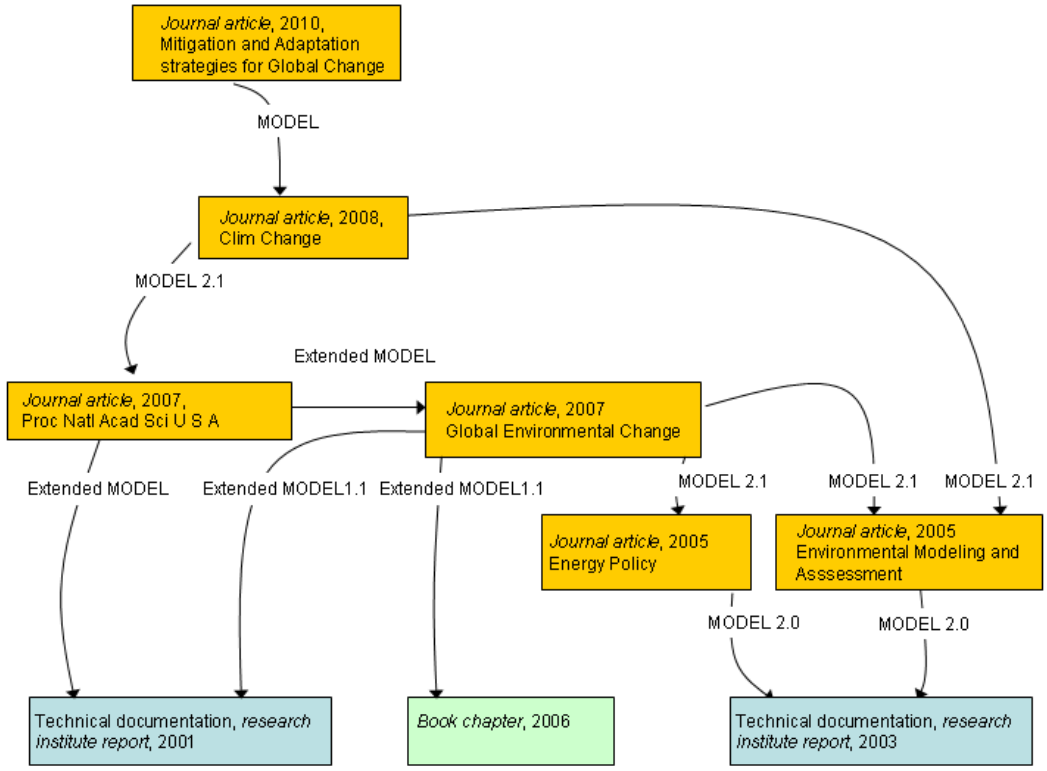

Figure 1 Chain of references for one of the case study models. Journal articles refer to the applied model indirectly by referring to other papers which eventually refer to grey literature reports or book chapters.

3.6. Reproducibility and transparency

All four models and their applications have been described in dozens of articles in peer reviewed journals. This indicates that these models and their results and insights are trusted and used. They can be understood as well-established models grounded in a scientific theory. However, in our study we found that for at least three of the models reviewers lack information to evaluate their quality and to assess their ability to support environmental decision making and give insight in the modeled system. Neither can they ensure the reproducibility and transparency of the model results and insights, for a number of reasons. First, information on model design is scattered over several sources, which are only to a limited extent freely and easily available to reviewers. Second, written model documentation does not provide a sufficient description of the modeled system. Third, written model documentation does not provide a sufficient description of the calculation of model results. Finally, results presented in scientific publications do not necessarily correspond with parameter values or equations in the model source code.

4. DISCUSSION

Our findings give the impression of environmental models lacking essential quality characteristics in terms of transparency and reproducibility. Although numerous guidelines and rules exist to enhance the transparency of environmental models (Alexandrov et al., 2011; Gaber et al., 2008; Jakeman et al., 2006; Refsgaard and Henriksen, 2004; Risbey et al., 2005; Rykiel Jr, 1996; Schmolke et al., 2010b), the modeling process and documentation of three of the case study models was not perceived as such. In our cases the environmental models effectively became black-boxes, which could only partly be uncovered with a significant investment in time and effort. Note that our objective was not to evaluate these models, nor to question their usefulness and suitability for assessment purposes, as this is typically done by experts in the peer-review process. However, in our cases we found that these experts probably lacked information to perform this evaluation properly. The concern raised through our research is that environmental models are being used in applications without explicitly respecting or at least discussing the underlying choices and assumptions, which is essential an area where assumptions may vary largely.

We believe that the models investigated are representative of environmental models in general. They are well established, published and cited in a large number of articles and regularly used in policy-oriented applications and assessment of complex societal problems. Although a more comprehensive study should provide additional insight in the overall situation, we think that already preliminary suggestions can be given on what underlying causes are and which remedies can be thought of. In particular we identify three main structural causes for lack of model transparency, which we will elaborate on in the next paragraphs:

- The size and complexity of environmental models.
- Lack of incentives for environmental modelers to display transparency in their work.
- The use of computers. .

Only a general change of attitude, including model developers, stakeholders and the scientif community as a whole, may improve this situation.

4.1. Model size and complexity

Following guidelines of good modeling practice, the developers of our case study models had made descriptions available of the concepts, processes and calculations steps included in their models. However, these descriptions only contained arbitrary selections of parts of the models. Making only part of a model explicit is a pragmatic choice because environmental models are typically big and complex. Describing all elements in the models requires a lot of time and does not fit within the constraints set by scientific publications or project budgets. The question is whether complexity is a required attribute to begin with. The complexity and size of environmental models is increasing with the advancement of computational systems, the development of integrated modeling approaches (Seppelt et al., 2009) and the modification of successful models to fit new purposes (Merali, 2010). However, the essence of modeling is abstraction and simplification by making appropriate assumptions. It is necessary to determine an adequate level of model complexity not only improve model performance (Jakeman et al., 2006; Seppelt et al., 2009), but also to enhance transparency.

4.2. Lack of incentives

Environmental modelers do not always display transparency in the modeling process for several reasons. Firstly, they preferably do not disclose model imperfections as these could be considered as failures (Barnes, 2010). We note however that any model is imperfect. Secondly, modelers are hesitant to give away the intellectual property that is embodied in datasets or source code (Barnes, 2010; Kleiner, 2011). Finally, documenting models carefully is a labor-intensive (Barnes, 2010; Kleiner, 2011) and not very interesting job, in particular considering the on-going development of some models. Schmolke and co-authors (2010a) observe that in general there is a lack of incentives for modelers to follow good practice.

Regarding transparency and reproducibility, the incentive may come from peers, stakeholders and journals who request openness. Model developers would spend more time and effort writing clear and comprehensive model documentation when there is a demand for it, for example in cases with an extensive and well defined user group. It is not a habit of environmental scientists to publish their data or source code, nor are there many scientific journals that require them to do so (Barnes, 2010; Kleiner, 2011; Merali, 2010). However, when they are requested or even obliged to share their data and code it will make model results and insights replicable, it will make model developers more accountable and it will enhance cooperation between scientists (Barnes, 2010; Kleiner, 2011; Merali, 2010).

4.3. Use of computers

Of course the use of models and networks has had enormous impact on environmental modeling. It has accelerated development and use of models dramatically, because vast amounts of data and computations can be processed in limited time. But environmental modelers often lack the time and skills that are needed to perform rigorous testing and clear annotation of their models (Kelly, 2007; Merali, 2010). Instead, they prefer to focus on 'model validation', i.e. verifying whether the model results match their expectations or real world observations. Here we observe a striking difference between scientific and commercial programming (Merali, 2010). Training scientists by professional software developers may be useful, but does not solve the entire problem, as this will lead to focus on system architecture and computation, and cause negligence with respect to descriptive side of modeling (Merali, 2010; Villa et al., 2009). The leading principle of scientific software should be 'representing applied scientific knowledge', which requires specific software development skills and tools. For many environmental models though the original source code (of the proper model version) is the only accurate description of the model. This programming code cannot be read by most model users, if it is available to them at all. But more important is that the source code does not represent the assumptions made by the modeler, i.e. which phenomena are (not) taken into account (Villa et al., 2009). We postulate that each model requires conceptual description that is accessible and comprehensible to all stakeholders, and that is consistent with the model source code.

4.4. Change of attitude

We find that the present lack of model transparency and reproducibility can only change with a general change of attitude with respect to handling complexity, introducing incentives and conceptual modeling. This change of attitude concerns all parties involved in the development and use of environmental models, viz. modelers, peers, journals and stakeholders. A model is not a crystal ball that provides certified and unambiguous predictions, but a tool to discuss premises (assumptions) and their consequences. Of course a model should be based on scientifically established theories, but it should be clear how these theories are made operational. When approving publication of environmental models, peer reviewers should take their responsibility for the underlying assumptions and choices. The same is true for decision makers when they use environmental models in exploring the consequences of alternative policies or management scenarios. A necessary condition is that peers and decision makers are able to examine these premises and assumptions, but when they lack information it is their responsibility to request openness from model developers or journals.

To achieve this general change in attitude, we see a few promising developments. Firstly, an increasing awareness exists of the importance of methods to provide credits to modelers and data providers. This can be done through formal authorship, scientific ranking, licenses and other incentives. There is a growing awareness of the importance of open source code and data (Barnes, 2010; Boulton et al., 2011; Kleiner, 2011). Scientific disciplines like astronomy and genomics as well as software engineering in the public and commercial domain can provide useful lessons here. Secondly, the web can be used to support the openness on models. It can relate

models to journal papers through the possibility to deliver additional materials with the article, a feature that more and more journals offer. For journals, it might be time to make this delivery of additional material for models and data into an obligation. Semantic publishing of journal articles, which allows readers to access and interact with the data and conceptual knowledge of the corresponding model can then be the next step (Attwood et al., 2009; Shotton et al., 2009). Current developments in the Web of Data are perfectly suitable for publishing conceptual models based on shared vocabularies and for representing complex model systems in a clear and explicit way.

5. CONCLUSION

Our findings suggest that environmental models lack essential quality characteristics in terms of transparency and reproducibility. This raises the concern that they are being used in applications without respecting and discussing their underlying choices and assumptions. We identify three structural causes for this lack of transparency and reproducibility: (1) the size and complexity of environmental models, (2) a lack of incentives for environmental modelers to be to transparent in the modeling process and (3) the use of computers and the focus on computation and simulation instead of the descriptive side of modeling. We submit that openness in the modeling process can only be achieved with a general change of attitude. On the one hand model developers must become explicit and open about their choices and assumptions. On the other hand peers, stakeholders and journals must request openness and challenge these choices and assumptions. In an operational sense, computers and networks can be turned to their advantage by having them disseminate high-quality model descriptions using shared vocabularies.

REFERENCES

Alexandrov, G.A., Ames, D., Bellocchi, G., Bruen, M., Crout, N., Erechtchoukova, M., Hildebrandt, A., Hoffman, F., Jackisch, C., Khaiter, P., Mannina, G., Matsunaga, T., Purucker, S.T., Rivington, M., Samaniego, L., 2011. Technical assessment and evaluation of environmental models and software: Letter to the Editor. Environmental Modelling and Software 26(3) 328-336.

Attwood, T.K., Kell, D.B., McDermott, P., Marsh, J., Pettifer, S.R., Thorne, D., 2009. Calling international rescue: Knowledge lost in literature and data landslide! Biochemical Journal 424(3) 317-333.

Barnes, N., 2010. Publish your computer code: It is good enough. Nature 467(7317) 753.

Boulton, G., Rawlins, M., Vallance, P., Walport, M., 2011. Science as a public enterprise: the case for open data. The Lancet 377(9778) 1633-1635.

Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Kluwer Academic Press, Dordrecht, The Netherlands.

Gaber, N., Pascual, P., Stiber, N., Sunderland, E., Cope, B., Nold, A., 2008. Guidance on the Development, Evaluation and Application of Environmental Models. Council for Regulatory Environmental Modeling, U.S. Environmental Protection Agency: Washington.

Hickman, L., 2009. Climate scientist at centre of leaked email row dismisses conspiracy claims. The Guardian, 24 november [Online]. Available at http://www.guardian.co.uk/environment/2009/nov/24/climate-professor-leaked-emails-uea (accessed 14 juli 2011).

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

Kelly, D.F., 2007. A software chasm: Software engineering and scientific computing. IEEE Software 24(6) 120+118-119.

Kleiner, K., 2011. Data on demand. Nature Clim. Change 1(1) 10-12.

Merali, Z., 2010. Why scientific programming does not compute. Nature 467 775-777.

Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263(5147) 641-646.

Refsgaard, J.C., Henriksen, H.J., 2004. Modelling guidelines--terminology and guiding principles. Advances in Water Resources 27(1) 71-82.

Risbey, J., Kandlikar, M., Patwardhan, A., 1996. Assessing integrated assessments. Climatic Change 34(3-4) 369-395.

Risbey, J., van der Sluijs, J., Kloprogge, P., Ravetz, J., Funtowicz, S., Quintana, S.C., 2005. Application of a checklist for quality assistance in environmental modelling to an energy model. Environmental Modeling & Assessment 10(1) 63-79.

Rosenblueth, A., Wiener, N., 1945. The Role of Models in Science. Philosophy of Science 12(4).

Rykiel Jr, E.J., 1996. Testing ecological models: The meaning of validation. Ecological Modelling 90(3) 229-244.
Schmolke, A., Thorbek, P., Chapman, P., Grimm, V., 2010a. Ecological models and pesticide risk assessment: Current modeling practice. Environmental Toxicology and Chemistry 29(4) 1006-1012.
Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010b. Ecological models supporting environmental decision making: a strategy for the future. Trends in Ecology & Evolution 25(8) 479-486.
Seppelt, R., Müller, F., Schröder, B., Volk, M., 2009. Challenges of simulating complex environmental systems at the landscape scale: A controversial dialogue between two cups of espresso. Ecological Modelling 220(24) 3481-3489.
Shotton, D., Portwin, K., Klyne, G., Miles, A., 2009. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. PLoS Computational Biology 5(4).
van der Sluijs, J.P., 2002. A way out of the credibility crisis of models used in integrated environmental assessment. Futures 34(2) 133-146.
Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: A review of emerging semantic approaches to environmental modelling. Environmental Modelling & Software 24(5) 577-587.