The model validation debate and implications on decision making processes

Master thesis by Despina Tsopoglou-Gkina [3776352]

Master: Environmental Biology

Supervisors: Dr. Stefan Dekker, Daily supervisor: Dr. Patrick Bogaart

Utrecht University

4 March 2013

Summary

In the procedure of trying to understand complex environmental processes, models have been widely used as tools in scientific research. According to their potential purposes, models have multiple applications, mainly linked to scientific research (scientific models) and to management advice provision (engineering models), with those aiming to form predictions gaining ground and forming the basis for public policy decision making in local, regional, national and international levels. Therefore, there is a need of an objective form of model assessment in order to evaluate the reliability of such models. There has been a lot of debate on the model validation issue, especially around the radical view of Oreskes *et al.* (1994) stating that is impossible. These discussions, although constructing, are at times hampering the processes for which models are made, such as decision and policy making. The implications on these procedures are increasingly being studied, and new methods are continuously being suggested in order to address and tackle them.

Contents

1. Introduction

In the procedure of trying to understand complex environmental processes, models have been widely used as tools (Wainwright and Mulligan 2004). The limited knowledge humans are able to acquire on both time and space of various phenomena (Oreskes *et al.* 1994), is a major incentive for modelling their observable effects and outcomes. Although observation is always considered as a key element to scientific research, models are necessary in order to tackle problems like predictions and answers to theoretical problems that no traditional experiments can give. As Caswell (1988) stated, “models are to theoretical problems as experiments are to empirical problems”. They are always a simplification of the complex reality presenting only the features of the system (Khaiter and Erechtchoukova 2007) corresponding to the models’ application nature.

According to their potential purposes, models have multiple applications, mainly linked to scientific research (scientific models) and to management advice provision (engineering models) (Rykiel 1996; van Voorn *et al.* under review). First, they can be used to test new hypotheses and support or confirm former proposed ones. For instance, ecological processes models can be targeted towards analysing environmental questions and developing of the ecological theory (Reynolds and Ford 1999). Moreover, models can verify inconsistencies in other models, as well as in performing sensitivity analyses in order to further explore understudied properties of the system under consideration. Furthermore, they can reveal patterns in observation data, or even synthesize new data when connecting elements of a system to reproduce its functions. Such models’ application is usually locally executed so as to characterise a specific study area with its unique characteristics (Beven 2002; Aral 2010). Last but not least, an increasingly appealing application of models is predicting the future behaviour of aspects of studied systems (Khaiter and Erechtchoukova 2007; Oreskes *et al.* 1994). The latter application, in spite of its weak reliance on the results, is largely used in management purposes for over a decade now. This kind of application-driven models, have been gaining ground in forecasting purposes and have formed the basis for public policy decision making in local, regional, national and international levels (Oreskes *et al.* 1994; Rykiel 1996). Attention has to be paid, however, due to the aforementioned weakness of models’ predictability, which derives from the uncertain nature of scientific knowledge. The facts that more than one model set-up can have similar output (non-uniqueness of models), that temporal and spatial data vary, and, ultimately, that model assessment is subjective – given that all models are approximations of real systems – contribute to this uncertainty (Oreskes and Belitz 2001). Consequently, there is a need of an objective form of model assessment with respect to its scientific purpose or application, to cover for any deficiencies.

As far as the ecological process models are concerned and according to Reynolds and Ford (1999), four major sources of models’ insufficiencies can be described. They can be found in the process of forming the hypotheses and their mathematical representation, during the fitting procedure of the model, and in the selection of the assessment criteria. In addition to these, the non-uniqueness of the models leads to their inadequate validation (Oreskes *et al.* 1994; Reynolds and Ford 1999). Different terms have been used to describe various types of model assessment, often leading to confusion due to lack of common definitions and implementations. Therefore, besides validation, verification, calibration, and confirmation are types of model assessment. Verification is linked with truth demonstration by the model, which sets the base for decision-making through its reliability. Validation, on the other hand, is relevant to the depiction of consistencies within a system or between systems, which is different from the accuracy with which the system represents the truth of natural phenomena. Moreover, calibration is the process of tuning the model in order to match predicted and obtained output of the model. Calibration can be strongly connected with the verification phase, as additional tuning can be realised during the latter. Last but not least, when the model output is confirmed by empirical observations, it is a way of evaluating the probability of the model to reproduce observed data, but not necessarily demonstrating its veracity (Oreskes *et al.* 1994).

The model validation problem and its legitimacy have been debated since the 1960s. A chronological review of validation concepts in the ecological literature was given by Rykiel (1996), two years after the publication of the Oreskes *et al.* (1994) article, which triggered a philosophical discussion on the model validation issue. In his review, Rykiel points out that there are not available universal criteria for a unified perspective of the validation problem. Despite their common view for the necessity of purpose-driven models, ecological modellers present different opinions on model validation definition and application.

In the present paper, a brief review of views and opinions on ecological model validation from the 1960s until the early 1990s will be given, according to Rykiel (1996). After that, an outline of the Oreskes *et al*. (1994) paper "Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences" will be presented, as well as the reception of it in the environmental modelling community. Furthermore, the implications of model validation on decision and policy making will be described.

2. Pre – existing, to Oreskes *et al*. (1994) paper, views on ecological model validation (Rykiel, 1996)

The main source of disagreement among ecological modellers, concerning model validation, is the disambiguation of models nature and purpose; the question of whether a model should be used in forming scientific hypotheses or in testing one. The high variation in opinions resulted in different validation definitions and terminology.

A first entry to the discussion was made by Levins (1966). They supported the idea that models being neither hypotheses nor theories, they rather generate hypotheses which can be tested. Goodall (1972), however, rejected the need for hypothesis testing issue as being irrelevant, paying attention to the most important, according to them, question of the predictive value of a model. According to him, the match between a model and a real system should be the centre of validation process. The generated acceptable predictions, though, of a specific ecosystem model are not automatically applicable to a generalised variety of other ecosystems. Caswell (1976) used the term validation only to describe predictive models which can be validated or invalidated for their engineering performance testing. Thus, model validation is considered purpose-driven. Mankin *et al*. (1977) suggested that models' purposes can be realised without any validation, and that instead of focusing on that, we should value more the usefulness of them. They define a useful model as one that its behaviour corresponds to some real behaviour; this model is considered valid, according to the objectives set in the process of validation. Therefore, the notion of purpose-driven model evaluation is appearing as in Caswell (1976). On the other hand, Overton (1977) during the same year, stressed the relation of validation to hypothesis testing, along with its importance after the model built. As far as the validation process is regarded, they agreed with Goodall (1972) on testing the model predictions with independent data, and with Levins (1966) on model validation generating testable hypotheses in order to answer questions and deal with problems for which the model was designed (purpose-driven model validation). One year later, Holling (1978) – opposing to Levins (1966) is stating that models are hypotheses which can only be falsified. Their view is in line with Karl Popper's argument on the impossible validation of hypotheses (Konikow and Bredehoeft 1992). Shugart's (1984) point of view corresponds to that of

Overton (1977) and Goodall (1972), who indicate as validation, the procedure of testing independent observations against data used to build the model. Konikow and Bredehoeft (1992), combining the above mentioned comparison of observations with experimental measurements (history matching) and the Popperian view on models' invalidation (1959), present the reliable predictability of a model as the ultimate goal of validation. They are sceptical, however, about the use of the term "validation", because it implies that a model can produce valid, accurate and reliable predictions, which is not possible for natural systems. Lastly, Botkin (1993) agreed with Konikow and Bredehoeft (1992) to the extent of the term usage, but not to that of the validation procedure. They consider validation through history matching with independent data not consistent with its logical meaning, while validation should be deducting a logical conclusion from the arguments used to set up the model.

Within this framework and background, Oreskes *et al*. (1994) paper "Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences" was added to the philosophical debate literature on model validation, presenting the extreme view of the impossibility of natural systems' validation, due to fact that they are never closed. Models are considered as representations which lead to further research, an attribute of their heuristic value. These models are by no means reliable representations of natural phenomena and their validity lies within the boundaries of as system or between systems' consistencies. According to Oreskes *et al*., the definition of validation is "the establishment of legitimacy, typically given in terms of contracts, arguments and methods". Therefore, only generic computer codes can be validated and not actual model results. The latter, depends, though, on the quantity and quality of the input parameters and the accuracy of the auxiliary hypotheses. Although the philosophical, restricted sense of the term "validation" is regarded as conflicting to common practice, numerical simulation models for large – scale or complex physical processes have been forming the basis for public policy decisions. Consequently, the validation issue remains always controversial, because it creates a division between philosophers and modellers / policy makers.

3. Reception of Oreskes *et al*. (1994) paper in the environmental community

For the paper purposes the fifty first most relevant – according to the Google Scholar citation ratings - citations of the paper "Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences" of Oreskes *et al*. (1994) were studied. Only scientific journal articles were included, whilst book excerpts and papers out of the environmental science field were excluded (Table 1).

Table 1: Fifty citations of Oreskes *et al*. (1994) paper and the environmental science field they belong to. The articles are ranked according to their relevance to the paper, using Google Scholar.

Authors	Field in environmental science
Guisan & Zimmermann 2000	Ecological modelling
Parker *et al*. 2003	Land-use and land-cover change modelling
Scheffer & Carpenter 2003	Ecosystem shifts modelling
Beissinger & Westphal 1998	Ecological modelling
Vanclay 1994	Ecological modelling
Rykiel 1996	Ecological modelling
Boyce *et al*. 2002	Ecological modelling
Beven 2006	Hydrological modelling
Araujo *et al*. 2005	Climate change modelling
Araujo & guisan 2006	Ecological modelling
Reddy *et al*. 1999	Environmental modelling
Heikkinen *et al*. 2006	Climate change modelling
Refsgaard 1997	Hydrological modelling

Demeritt 2001	Climate modelling
Finnveden *et al*. 2009	Environmental modelling
Weaver *et al*. 2001	Climate modelling
Saltelli *et al*. 2000	Modelling: sensitivity analysis
Shaw 2003	Climate modelling
Jakeman *et al*. 2006	Environmental modelling
Bugmann 2001	Ecological modelling
Kirchner *et al*. 1996	Ecological modelling
Martinez-Meyer *et al*. 2004	Climate change modelling
Araujo *et al*. 2005	Climate change modelling
Perrin *et al*. 2001	Hydrological modelling
Kobayashi & Salam 2000	Agronomy simulation modelling
Vanclay & Skovsgaard 1997	Ecological modelling
Chang & Hanna 2004	Air quality modelling
Beven 2002	Hydrological modelling
Wagener *et al*. 2003	Hydrological modelling
Beven 2002	Hydrological modelling
Schneider 1997	Climate change modelling
Erasmus *et al*. 2002	Ecological modelling
Hilborn *et al*. 1995	Ecological modelling
Saltelli 2002	Computational modelling
Refsgaard & Henriksen 2004	Hydrological modelling
Pontius *et al*. 2004	Land-use and land-cover change modelling
Van Asselt & Rotmans 2002	Integrated assessment modelling
Bradshaw & Borchers 2000	Climate modelling
Warren & Haack 2001	Geochemical modelling
Nathan *et al*. 2001	Ecological modelling
Grimm *et al*. 1996	Ecological modelling
Saltelli *et al*. 2006	Environmental modelling
Van Der Sluijs *et al*. 2005	Environmental modelling
Corwin *et al*. 1997	Environmental modelling
Van Lieshout *et al*. 2004	Climate modelling
Refsgaard *et al*. 2006	Hydrological modelling
Young 2002	Hydrological modelling
Rastetter 1996	Ecological modelling
Knutti *et al*. 2010	Climate modelling
Stanley 2003	Ecological modelling

In total, all papers accept Oreskes *et al*. (1994) view on validation, although it is very difficult to distinguish which of the authors actually agree with them on its strict definition. The main trend followed is the approval of the Oreskes *et al*. (1994) validation concept in a philosophical framework, which is often detoured when the implementation of the models is concerned.

a. Ecological modelling

The paper of Oreskes *et al*. (1994) made an impression even to papers published the same year as that of Vanclay (1994) who considers validation as a necessary step in the process of model evaluation, along with verification. In their field, that of forest growth modelling, validation equals quantitative tests. They conclude, nevertheless, that the term “benchmarking” should replace that of validation in accordance with Oreskes *et al*. Hilborn *et al*. (1995) illustrate the inability to extensively specify initial conditions, and consequently to base predictions on them. They deny successful model testing, driven by the inevitable result of multiple alternative hypotheses matching the data. Rykiel (1996) considers the validation debate an obstacle to models’ actual use and performance, while he states that validation criteria should be accessible to the user, who in turn will judge their adequacy. Lastly, they point out the lack of universal validation criteria and test procedures. Kirchner *et al*. (1996) pay also attention to the practical side and the purpose of modelling. Expert judgement and model building are both considered as appropriate tools when the model purpose is policy analysis. Their uncertain use is necessary for predictions and decision-making. In the same practical way, by accepting that models in population ecology for a great simplification of reality, Grimm *et al*. (1996), proceed at the usage of the models in research. This simplicity in model designing is also applied by Beissinger and Westphal

(1998). Similarly, despite the weak model evaluation by model tests used to deal with the problem of extrapolating from the specific to the general, Rastetter (1996) regards a model as an essential part of the evaluation of the responses of ecosystems to global climate and carbon dioxide change. In their opinion, models can synthesize scientific information from diverse sources in order to evaluate full ecosystem responses. The avoidance of the term "validation" is followed by Vanclay and Skovsgaard (1997) too. They propose a model evaluation with or without external data, and that a model cannot be proven correct, but instead only the inferences can be falsified. Guisan and Zimmerman (2000), are also in line with Oreskes *et al*. (1994) validation perspective, and present two approaches for model evaluation, which is their preferred term. They perform cross-validation techniques with the use of either two independent data sets or just one. A comparable approach is implemented by Bugmann (2001), as they stress the need for quantitative methods for model evaluation, and from Nathan *et al*. (2001) who compare predicted with observed data. The comparisons of present against past data is supposed to be of little predictive value by Boyce *et al*. (2002) due to the fact that these models describe dynamic systems which change in time. Another approach followed by other modellers is to fill absence data in by interpolation (Erasmus *et al*. 2002). Modellers nowadays recognise this poor predictability of models and the numerous limitations they present in their implementation, but their importance in environmental managing renders them essential (Araújo and Guisan 2006; Scheffer and Carpenter 2003; Stanley 1995).

b. Hydrological modelling

Model validation is described by Refsgaard (1997) as the process that demonstrates how accurate the model's predictions are. They implement a split-sample procedure to evaluate hydrological models, and accept a model validation only within limits of its accuracy and predictive capability. Perrin *et al*. (2001), however, prefer model simplicity over accuracy, which causes less parameter uncertainty problems. The incapability of models to describe reality, as far as hydrological models are concerns, is also accepted by Beven (Beven 2002a; Beven 2002b; Beven 2006). They avoid the philosophical debate over the model validation, and suggest an approach based on the falsification of models where it can reduce the range of hypotheses' possibilities from which one has to choose. The arguments on the possibility of model validation are regarded too philosophical by Young (2002) as well. They support the quantitative, predictive validation which uses two different sets to evaluate the model, similar to the split-sample technique of Refsgaard (1997). The type of failure of a model's structure (failure a structural model component and of the hypothesis underlying it) is interpreted by Wagener *et al*. (2003) as the key to develop an improved hypothesis and, therefore, an even more acceptable model structure, until a better one to be found. Lastly, a significant addition to the hydrological modelling literature is made by Refsgaard and Henriksen (2004) who, in agreement with the view of Oreskes *et al*. (1994), proposed modelling guidelines, terminology and guiding principles. Their definition of validation is restrained to the domains of applicability and performance accuracy of models.

c. Environmental modelling

Corwin *et al*. (1997) point out that the major problem in applying model simulations, like the nonpoint source pollutants models, to real problems is the fact that the uncertainties at large scales (e.g. regional scales) are high and lead to errors. Accuracy – the extent to which model-predicted values approach a corresponding set measured observations – and precision – the degree to which model-predicted values approach a linear function of measured observations – are the two operational components of model evaluation. To the inaccurate and incompletely known empirical data attribute Reddy *et al*.

(1999) the incapability of models to be validated. However, they support the idea of models' descriptive power when tested against observational data. In addition, Van der Sluijs *et al*. (2005) also are in favour of applying methods such as the Numeral Unit Spread Assessment Pedigree (NUSAP) method for multidimensional uncertainty assessment, which they trust for applying it even to complex models. Their attempt is to deal with the uncertainty produced by the degree at which modes are based on observations, informed judgement and scientists' convenience. Jakeman *et al*. (2006) appear to be flexible at the setting of the criteria by which a model performance can be assessed, implying that the philosophical basis of the validation term of Oreskes *et al*. (1994) is too strict, forcing the choice between confirmation or rejection of a model. Saltelli *et al*. (2006), on the other hand, keep a more neutral position, characterising, though, the debate on model validation as "blunt". Their focus lays on corroborating models via sensitivity analysis, to alleviate uncertainties. Finnveden *et al*. (2009) do not elaborate on the model validation issue and the uncertainty caused by the comparison of measurements or calculations with "truth", suggesting further attention and development in the area.

d. Climate and climate change modelling

For climate and, especially, climate change models which represent a very complex system with many influencing factors, empirical evidence is difficult to obtain prior to the actual experimental phase. Thus, Schneider (1997) agreed with Oreskes *et al*. (1994) that validation is not possible in advance. They still use the term, though, to define the various testing strategies used to provide subjectivity and credibility to the models' results and insights to integrated assessment models (multi-disciplinary models). The uncertainty of large-scale and complex systems like the global change models is accepted by Bradshaw and Borchers (2000) too. Although scientific confirmation helps increase confidence in public's trust towards the implementation of policies, the inability of simulating chaotic systems and producing the needed certainty levels for models' performance is a fact affecting the relation between science and policy-making. Martínez-Meyer *et al*. (2004) highlight, however, that while the ecosystem models have undergone some degree of testing and validation, the species-level models under the scope of climate change have seen little or no direct testing. Araújo *et al*. (2005a; 2005b), are also working on species' responses to climate change modelling and they are stressing the weak predictive ability of such model estimations. They, moreover, argue that it is vital that such models - although useful in providing an approximation of climate-driven range of changes - have to be applied critically, while unrealistic optimistic estimates of predictive ability have to be avoided. Heikkinen *et al*. (2006) consider Oreskes *et al*. (1994) view on model validation as extreme, but recognise that as all natural systems are not closed, many factors as e.g. the potential driving forces of species distributions are not possible to account for. Demeritt (2001) and Weaver *et al*. (2001) are influenced by the Oreskes *et al*. (1994) paper and in response, they prefer to use the term evaluation instead of validation. Both papers accept the comparison of model output with observations as the suitable evaluation process. In the same direction, Shaw (2003) suggests comparison of cloud processes models outputs with actual observations. Van Lieshout *et al*. (2004) state that modelling, despite its decreased inaccuracy and simplifications, allows for adequate prioritisation and risk estimation. Knutti *et al*. (2010), finally, reason that models adequate for a particular purpose should be satisfying, in spite of the accuracy not being perfect.

e. Other types of modelling

Kobayashi and Salam (2000) argue that model's performance can only be discussed relatively, which is in accordance with the view of Saltelli *et al*. (2000) that models are built for a specific task and their complexity should not exceed the needed requirements. Saltelli (2002) become more specific later on by suggesting that the focus should be on global, quantitative and model free capable of testing the robustness and relevance of a model-based analysis in the presence of uncertainties. Warren and Haack (2001), in their geochemical modelling, find that the utility of models lies only in the examination of the extent of a hypothesis, and not in the establishment of the validity of a model based on the “goodness of fit”. Van Asselt and Rotmans (2002) are detouring the validation issue by incorporating multiple perspectives in Integrated Assessment modelling in order to assess the most salient uncertainties (structural and multiple model routes) derived during the model built process. Parker *et al*. (2003) place the validation process after the verification and the check of models correct function, and stress the necessity of the use of a wide range of techniques for model development and empirical assessment. Chang and Hanna (2004) in agreement with Oreskes *et al*. (1994) prefer the term evaluation over validation and state that model confirmation or evaluation are achieved by the demonstration of good agreement between several sets of observations and predictions. The scientist’s decision on the validity of a model for a particular domain of application suggest Pontius *et al*. (2004), as an inevitable condition in practice. They find it more helpful, when model improvement is concerned, to view validation as a standard procedure of science that is designed to how in what respects models perform well and poorly.

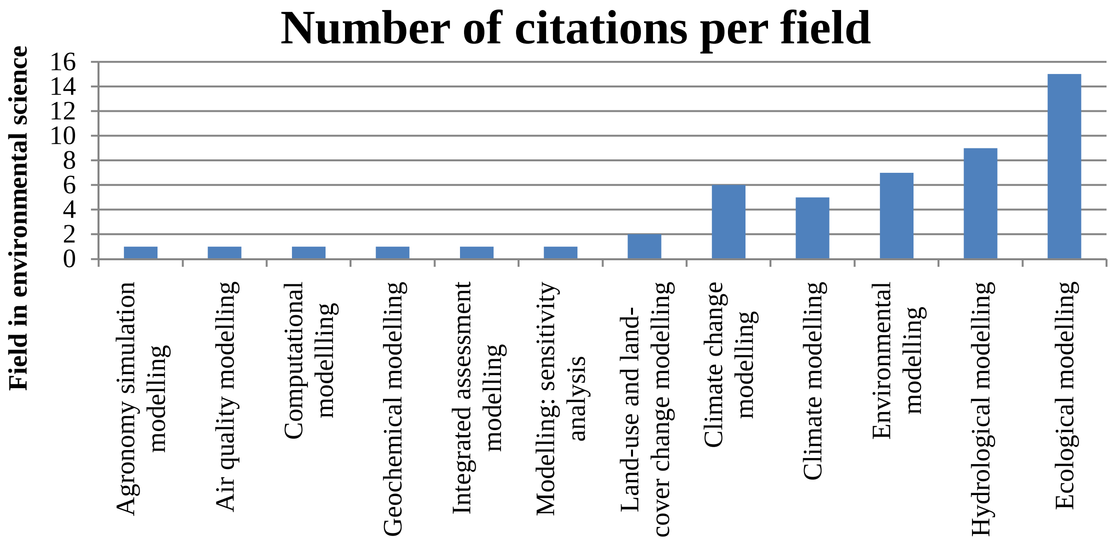

Figure 1: The amount of Oreskes *et al*. (1994) citations per environmental science field.

4. Decision and policy making: uncertainty and implications of model validation

Apart from the major aim of environmental modelling being the gain of an insight into system's functions and translating them in mathematical relationships, another goal is the generation of predictions to be used for management and decision-making purposes (van Asselt and Rotmans 1996; Beven 2009). Therefore, models' credibility is a significant feature in order to base decisions upon their results (Holzbecher 1997). Credibility is defined as the scientific adequacy of the technical evidence and arguments (Cash *et al*. 2003). Thus, the incapability of defining model validation, along with the uncertainty of model accuracy in describing the underlying natural processes of the system under research, results in questioning the reliability of model-based decision-making (Oreskes and Belitz 2001). Decision and policy makers are often interested in certain, feasible, and deterministic solutions (Bradshaw and Borchers 2000), unaware of the fact yet familiar to scientists, that the magnitude and degree of complexity is continuously increasing. Consequently, scientists / modellers and decision makers have to deal with uncertainty which is also increasing, despite the accumulation of knowledge (van Asselt and Rotmans 1996). In any case, the scientist must have an established communication with the decision maker, otherwise there is little chance that the information, derived from this knowledge, will be salient, and thus useful to the decision maker (Cash *et al*. 2002). An appealing approach to tackle the problem of the different perspectives in the model-based policy and decision-making in the environmental research field is an integrated assessment approach, aiming to facilitate these processes on complex issues. Environmental decision-making is considered as one, due to the complexity of the systems considered and the competing interests of multiple stakeholders (Cummings and Cayer 1993; van Asselt and Rotmans 1996; van Asselt and Rotmans 2002; Ascough *et al*. 2008).

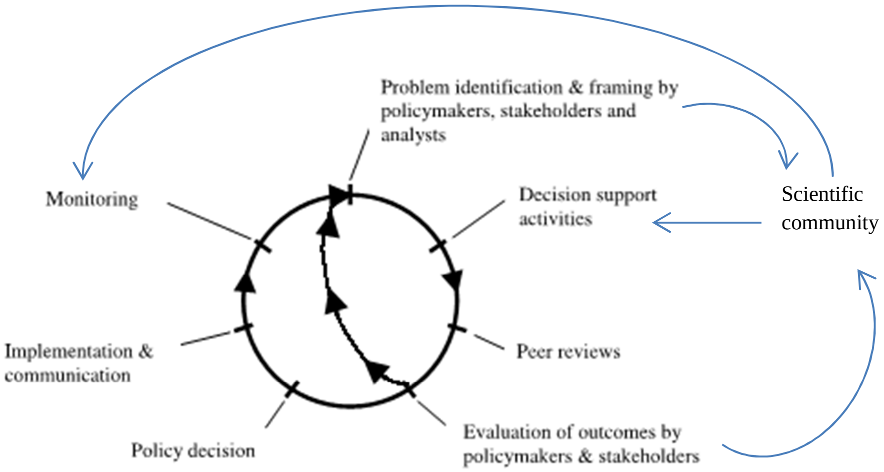

Figure 2: The policy making process viewed as a multi-stage iterative process (Walker *et al*. 2003, modified)

Uncertainty exists in all levels of the policy making process, as well as in the outcome of the simulation models used for it, or the assumptions made. For scientists it is rather an expected obstacle, whereas for the decision makers, it hampers the establishment of confidence and trust in the model and the modelling process (Bradshaw and Borchers 2000; Pappenberger and Beven 2006). In order to bridge this perception gap between actors involved in the policy making process (i.e. between scientists and the decision makers), a closer look to the process of an integrated assessment has to be taken. This process is considered as an iterative, circular process (Figure 2), where the scientific community provide decision support to decision makers by creation of a model of the system under research, and in turn, decision makers contribute to the input of scientific investigations by the experience gained from the results of the analysis, and the policy choice and implementation (van Asselt and Rotmans 1996; Walker *et al*. 2003). All actors, participating at a policy and decision making process, have different perceptions of reality, originating from their different views of the world. That explains the variety on their perspectives of uncertainty; that of the modellers focus being on the accumulated uncertainties associated with the outcomes of the model and the (robustness of) conclusions of the decision support exercise, while the policy makers' on how to value the outcomes regarding the goals and possibly conflicting objectives, priorities, and interests (Walker *et al*. 2003).

Notwithstanding the understanding for the need of an interdisciplinary research process, which will solve problems characterised by complexity and interconnectedness (Cash *et al*. 2002), and which emerge from the different actors involved in it, there is neither a commonly shared terminology nor agreement on a generic typology of uncertainties. Such a typology would solve issues such as problematic communication among policy analysts, among policy analysts, policy makers and stakeholders, and, finally, it would help in the identification and prioritisation of effective and efficient research and development activities for decision support. Therefore, the sources and types of uncertainty have to be identified. According to the uncertainty classification of Ascough *et al*. (2008) (Figure 3), uncertainty between lack of knowledge and uncertainty resulting from intrinsic variability in the systems or processes of interest are distinguished. Decision-making involved uncertainty is also considered as another type, as well as linguistic uncertainty.

The knowledge uncertainty is referring to the fact that there is a limitation to our knowledge, which can either increase or be reduced by additional research and empirical efforts. It is identical to the epistemic nature of uncertainty of Walker *et al*. (2003), which consists of the process understanding and the model uncertainty; the former representing the background scientific knowledge describing the system under consideration, and the latter, including the mechanistic parts of a model, from its designing, to its structure and function. The variability uncertainty has to do with the variability in natural and human systems, which refer to the stochastic nature of natural processes, and the influence of social structure and cultural standards (Bradshaw and Borchers 2000) on the environmental decision-making process. All these types of uncertainty lead to the decision-making uncertainty, due to mal-interpretation or miscommunication of model predictions, which can be part of the failure to embody social objectives in the policy analysis. Lastly, the linguistic uncertainty is describing the characteristics of our natural language, inhibiting constant clarity and precision.

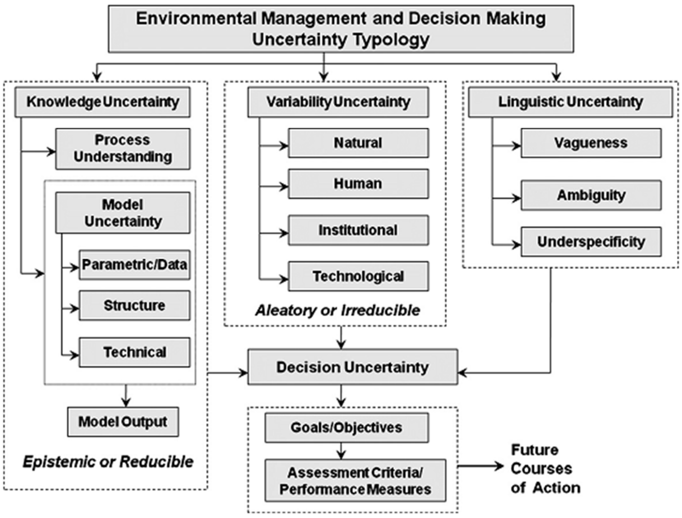

Figure 3: Description of uncertainty in environmental management and decision-making based on knowledge uncertainty, system variability, linguistic uncertainty, and decision-making uncertainty (Ascough *et al*. 2008).

In addition to these types of uncertainty, outlining more the scientists' perception, Bradshaw and Borchers (2000) introduced two more types, viewed by the public and policy makers perspective; first, there is uncertainty about the uncertainty and, secondly, the difficulty to translate science into decision-making as a result of lacking knowledge of scientific methods.

Having identified a classification of the types of uncertainty definitely enhances already the policy and decision-making process. However, there are several drawbacks in the addressing of uncertainty in an integrated assessment plan. To start with, there are not adequate methods and tools available, in order to address all uncertainties. This especially holds for uncertainty in model structure and uncertainty due to behavioural and societal variability, value diversity, technological surprise, ignorance and indeterminacy. However, there have been attempts to of methods in order to approach this range of uncertainties, like the dynamic identifiability analysis (DYNIA) of Wagener *et al.* (2003) for applications to analyse model structures or to estimate calibration parameters. Furthermore, virtually all current methods do not include evaluation of the impacts of specific uncertainties, because it is not considered a primary issue. Lastly, when there are no ways of indicating the magnitude and source of the underlying uncertainties – due to inadequate methods – actors other than scientists (i.e. decision-makers) find it difficult to deal with aggregated uncertainty measures (van Asselt and Rotmans 2002). Although it is inarguably a necessity to communicate uncertainty in the science/policy interface, when model-based decision support is considered (Walker *et al*. 2003), the uncertainty analysis has been doubted for its usefulness, as for example, by several scientists in water resources research, whose arguments are all refuted though by Pappenberg and Beven (2006). They conclude that uncertainty analysis is greatly influencing decisions in the direction of future predictions and, thus, in the decision and policy making in environmental matters. The reluctance of participating actors in these processes, to address and deal with uncertainty analyses, is only resulting in disputed results instead of the main goal which is a risk evaluation. In addition to this, Ascough *et al*. (2008) is stressing that is essential to address uncertainty in any comprehensive and defendable environmental decision-making situation, or the unreliability of the results with a consequential loss of public trust and confidence will be unavoidable.

5. Discussion and conclusions

Model validation has been debated since the 1960s, and has initiated several discussions among the scientific community, both from a philosophical and practical point of view. All arguments derive from the fact that human knowledge for natural systems is limited and little. Our perception of the systems' functions and characteristics is restricted, despite the series of observations and measurements which become increasingly more accurate and abundant compared to the past. In spite of the technology progress and achievements, there will always be elements and natural processes that we cannot have a grip on. Especially when future procedures are concerned, accurate predictions seem to be impossible, given the natural history of the world as we know it now. Thus, models built to describe natural systems, in order to explain their structure or to form predictions, are logically questioned regarding their validity of their performance. On the other hand, though, models have been a useful tool in actually gaining more knowledge of the natural processes for a long time, so their utility is relatively proven in practice.

The need to evaluate models is not only intrinsic to human nature and scientific precision, but it has a practical aspect and is essential to the purposes for which the models are built in the first place. The inability of model validation with the strict sense of the meaning, as Oreskes *et al*. (1994) define it, accepted as a notion by the scientific community, has been under great consideration in order to overcome it and, finally, assess the models. As presented in Part 2, there have been quite a few different theoretical views on model validation and whether it is possible or not, which they seem to converge after the statement that model validation of natural systems is impossible. Despite this convergence, there is not still a unified, universal terminology on model evaluation, resulting in difficulties in interactions within the scientific community, but also between the scientific community and the model users who are representing a large range of functions and interests. Oreskes *et al* (1994) view was received rather positively, however, scientists from various domains of environmental sciences (Part 3), had to face the questions that follow this ascertainment; what is the next step, how can models be evaluated and what are the consequences of non-validated models.

In Part 3 some approaches of the concept of model validity are outlined, originating from different perspectives in the environmental science field. The most popular model evaluation method is the comparison of observed data with model output data. The degree of the data match is proportional to the model assessment for the purpose initially built for. This history matching, as called differently, when successful, increases the model's credibility, because it reproduces real data. A model which can somewhat accurately reconstruct past events and parameters that have already been reported is more easily trusted to predict the future behaviour of the same system; notwithstanding uncertainties emerged at every stage of the model building and use (Part 4).

The uncertainty issue in the model evaluation process is of great significance, due to the implications that it has at the decision and policy-making. Models are use in a great extent from decision and policy makers, in order to design strategies of dealing with a problem. Therefore, the result of this interaction between science and policy implementation is of public interest and, thus, can involve a broad range of participating actors often irrelevant to the model making procedure. The different perspectives and intentions are sources of miscommunication and conflict, if the gap created is not bridged. Traditionally, the roles of the different functions were separated and discrete, which renders the communication necessary, as well as explanations, instructions and feedback provision. Nowadays, there is a trend of domains overlapping, and of going towards an integrated approach of problem solving. This is consistent with the appearance of the integrated assessment models which have a multi-disciplinary character, and thus, include a variety of points of views, which makes the model more objective and, consequently, more credible by a broader public. Such examples are quite optimistic and available, as the Numeral Unit Spread Assessment Pedigree (NUSAP) method for multidimensional uncertainty assessment which can be applied to complex models (van der Sluijs *et al*. 2005). Lastly, avoidance of application of methods establishing and increasing the model function power and, thus, credibility, like the uncertainty analyses, is no longer regarded as reasonable, since the arguments against it have all been compromised (Pappenberger and Beven 2006).

6. Acknowledgments

I would like to thank dr. Patrick Bogaart for his cooperation, continuous guidance and support during the entire master thesis writing. Furthermore, I would like to thank dr. Stefan Dekker for his general supervision.

7. References

Aral, MustafaM. 2010. “Principles of Environmental Modeling.” *2*. http://dx.doi.org/10.1007/978-90-481-8608-2_2.

Araújo, Miguel B., and Antoine Guisan. 2006. “Five (or so) Challenges for Species Distribution Modelling.” *Journal of Biogeography* 33 (10): 1677–1688. doi:10.1111/j.1365-2699.2006.01584.x.

Araújo, Miguel B., Richard G. Pearson, Wilfried Thuiller, and Markus Erhard. 2005. “Validation of Species?climate Impact Models Under Climate Change.” *Global Change Biology* 11 (9): 1504–1513. doi:10.1111/j.1365-2486.2005.01000.x.

Araújo, Miguel B., Robert J. Whittaker, Richard J. Ladle, and Markus Erhard. 2005. “Reducing Uncertainty in Projections of Extinction Risk from Climate Change.” *Global Ecology and Biogeography* 14 (6): 529–538. doi:10.1111/j.1466-822X.2005.00182.x.

Ascough, J.C., H. R. Maier, J. K. Ravalico, and M. W. Strudley. 2008. “Future Research Challenges for Incorporation of Uncertainty in Environmental and Ecological Decision-making.” *Ecological Modelling* 219 (3-4): 383–399. doi:10.1016/j.ecolmodel.2008.07.015.

Van Asselt, M. B. A., and J. Rotmans. 2002. “Uncertainty in Integrated Assessment Modelling - From Positivism to Pluralism.” *Climatic Change* 54 (1-2). doi:10.1023/A:1015783803445.

Van Asselt, Marjolein, and Jan Rotmans. 1996. “Uncertainty in Perspective.” *Global Environmental Change* 6 (2): 121–157. doi:10.1016/0959-3780(96)00015-5.

Beissinger, S. R., and M. I. Westphal. 1998. “On the Use of Demographic Models of Population Viability in Endangered Species Management.” *Journal of Wildlife Management* 62 (3). doi:10.2307/3802534.

Beven, K. 2002. “Towards a Coherent Philosophy for Modelling the Environment.” *Proceedings of the Royal Society A-Mathematical Physical and Engineering Sciences* 458 (2026): 2465–2484. doi:10.1098/rspa.2002.0986.

Beven, K. J. 2009. *Environmental Modelling: An Uncertain Future?: An Introduction to Techniques for Uncertainty Estimation in Environmental Prediction*. Taylor & Francis.

Beven, Keith. 2002. “Towards an Alternative Blueprint for a Physically Based Digitally Simulated Hydrologic Response Modelling System.” *Hydrological Processes* 16 (2): 189–206. doi:10.1002/hyp.343.

———. 2006. “A Manifesto for the Equifinality Thesis.” *The Model Parameter Estimation Experiment MOPEX MOPEX Workshop* 320 (1–2): 18–36. doi:10.1016/j.jhydrol.2005.07.007.

Botkin, D. B. 1993. *Forest Dynamics: An Ecological Model*. Oxford University Press, USA. http://books.google.com/books?hl=nl&lr=&id=8aZM4fWNOJQC&oi=fnd&pg=PR15&dq=Forest+Dynamics+botkin&ots=tFYRTCtzTd&sig=nNYig4G_V1dWQiQUNeqWNgszAVA.

Boyce, Mark S., Pierre R. Vernier, Scott E. Nielsen, and Fiona K. A. Schmiegelow. 2002. “Evaluating Resource Selection Functions.” *Ecological Modelling* 157 (2–3): 281–300. doi:10.1016/S0304-3800(02)00200-4.

Bradshaw, G. A., and J. G. Borchers. 2000. “Uncertainty as Information: Narrowing the Science-policy Gap.” *Conservation Ecology* 4 (1): 7–7.

Bugmann, H. 2001. “A Review of Forest Gap Models.” *Climatic Change* 51 (3-4). doi:10.1023/A:1012525626267.

Cash, D., W. Clark, F. Alcock, N. Dickson, N. Eckley, and J. Jäger. 2002. “Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making.” http://papers.ssrn.com/sol3/papers.cfm?abstract_id=372280.

Cash, D. W., W. C. Clark, F. Alcock, N. M. Dickson, N. Eckley, D. H. Guston, J. Jäger, and R. B. Mitchell. 2003. “Knowledge Systems for Sustainable Development.” *Proceedings of the National Academy of Sciences* 100 (14): 8086–8091.

Caswell, H. 1976. “The Validation Problem.” *Systems Analysis and Simulation in Ecology* 4: 313–325.

Caswell, Hal. 1988. “Theory and Models in Ecology: A Different Perspective.” *Ecological Modelling* 43 (1–2): 33–44. doi:10.1016/0304-3800(88)90071-3.

Chang, J. C., and S. R. Hanna. 2004. “Air Quality Model Performance Evaluation.” *Meteorology and Atmospheric Physics* 87 (1-3). doi:10.1007/s00703-003-0070-7.

Corwin, DL, PJ Vaughan, and K. Loague. 1997. “Monitoring Nonpoint Source Pollutants in the Vadose Zone with GIS.” *Environmental Science & Technology* 31 (8): 2157–2175. doi:10.1021/es960796v.

Cummings, LE, and NJ Cayer. 1993. “Environmental-Policy Indicators - a Systems-Model.” *Environmental Management* 17 (5): 655–667. doi:10.1007/BF02393727.

Demeritt, David. 2001. “The Construction of Global Warming and the Politics of Science.” *Annals of the Association of American Geographers* 91 (2): 307–337. doi:10.1111/0004-5608.00245.

Erasmus, Barend F. N., Albert S. Van Jaarsveld, Steven L. Chown, Mrigesh Kshatriya, and Konrad J. Wessels. 2002. “Vulnerability of South African Animal Taxa to Climate Change.” *Global Change Biology* 8 (7): 679–693. doi:10.1046/j.1365-2486.2002.00502.x.

Finnveden, Göran, Michael Z. Hauschild, Tomas Ekvall, Jeroen Guinée, Reinout Heijungs, Stefanie Hellweg, Annette Koehler, David Pennington, and Sangwon Suh. 2009. “Recent Developments in Life Cycle Assessment.” *Journal of Environmental Management* 91 (1): 1–21. doi:10.1016/j.jenvman.2009.06.018.

Grimm, V., K. Frank, F. Jeltsch, R. Brandl, J. Uchmanski, and C. Wissel. 1996. “Pattern-oriented Modelling in Population Ecology.” *Science of the Total Environment* 183 (1-2): 151–166. doi:10.1016/0048-9697(95)04966-5.

Guisan, Antoine, and Niklaus E. Zimmermann. 2000. “Predictive Habitat Distribution Models in Ecology.” *Ecological Modelling* 135 (2–3): 147–186. doi:10.1016/S0304-3800(00)00354-9.

Hall, Charles AS, and John W. Day Jr. 1977. “Ecosystem Modeling in Theory and Practice: An Introduction with Case Histories.” http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=5835861.

Heikkinen, Risto K., Miska Luoto, Miguel B. Araujo, Raimo Virkkala, Wilfried Thuiller, and Martin T. Sykes. 2006. “Methods and Uncertainties in Bioclimatic Envelope Modelling Under Climate Change.” *Progress in Physical Geography* 30 (6). doi:10.1177/0309133306071957.

Hilborn, R., C. J. Walters, and D. Ludwig. 1995. “Sustainable Exploitation of Renewable Resources.” *Annual Review of Ecology and Systematics* 26: 45–67.

Holling, Crawford S. 1978. “Adaptive Environmental Assessment and Management.” *Adaptive Environmental Assessment and Management.* http://www.cabdirect.org/abstracts/19800666996.html.

HOLZBECHER. 1997. *Remarks on the Paper by E.J. Rykiel, Entitled: “Testing Ecological Models: The Meaning of Validation” (Ecol. Model. 90, 229-244, 1996)*. Vol. 102. Ecological Modelling. Amsterdam, PAYS-BAS: Elsevier.

Jakeman, A. J., R. A. Letcher, and J. P. Norton. 2006. “Ten Iterative Steps in Development and Evaluation of Environmental Models.” *Environmental Modelling & Software* 21 (5): 602–614. doi:10.1016/j.envsoft.2006.01.004.

Jeffers, John Norman Richard. 1972. *Mathematical Models in Ecology: The 12th Symposium of the British Ecological Society, Grange-Over-Sands, Lancashire, 23-26 March 1971*. Blackwell Scientific Publications.

Khaiter, P. A., and M. G. Erechtchoukova. 2007. *From Complex to Simple in Environmental Simulation Modelling*.

Kirchner, J. W., R. P. Hooper, C. Kendall, C. Neal, and G. Leavesley. 1996. “Testing and Validating Environmental Models.” *Science of the Total Environment* 183 (1-2). doi:10.1016/0048-9697(95)04971-1.

Knutti, Reto, Reinhard Furrer, Claudia Tebaldi, Jan Cermak, and Gerald A. Meehl. 2010. “Challenges in Combining Projections from Multiple Climate Models.” *Journal of Climate* 23 (10). doi:10.1175/2009JCLI3361.1.

Kobayashi, K., and M. U. Salam. 2000. “Comparing Simulated and Measured Values Using Mean Squared Deviation and Its Components.” *Agronomy Journal* 92 (2). doi:10.1007/s100870050043.

KONIKOW, LF, and JD BREDEHOEFT. 1992. “Groundwater Models Cannot Be Validated.” *Advances in Water Resources* 15 (1): 75–83. doi:10.1016/0309-1708(92)90033-X.

Levins, R. 1966. “The Strategy of Model Building in Population Biology.” *American Scientist*: 421–431.

Van Lieshout, M., RS Kovats, MTJ Livermore, and P. Martens. 2004. “Climate Change and Malaria: Analysis of the SRES Climate and Socio-economic Scenarios.” *Global Environmental Change-Human and Policy Dimensions* 14 (1): 87–99. doi:10.1016/j.gloenvcha.2003.10.009.

Martínez-Meyer, Enrique, A. Townsend Peterson, and William W. Hargrove. 2004. “Ecological Niches as Stable Distributional Constraints on Mammal Species, with Implications for Pleistocene Extinctions and Climate Change Projections for Biodiversity.” *Global Ecology and Biogeography* 13 (4): 305–314. doi:10.1111/j.1466-822X.2004.00107.x.

Nathan, R., UN Safriel, and I. Noy-Meir. 2001. “Field Validation and Sensitivity Analysis of a Mechanistic Model for Tree Seed Dispersal by Wind.” *Ecology* 82 (2): 374–388. doi:10.2307/2679866.

Oreskes, N., and K. Belitz. 2001. “Philosophical Issues in Model Assessment.” *Model Validation: Perspectives in Hydrological Science* 23. http://history.ucsd.edu/_files/faculty/oreskes-naomi/PhilIssuesModelAssessOreskes.pdf.

Oreskes, N., K. Shraderfrechette, and K. Belitz. 1994. “Verification, Validation, and Confirmation of Numerical-Models in the Earth-Sciences.” *Science* 263 (5147): 641–646. doi:10.1126/science.263.5147.641.

Pappenberger, F., and K. J. Beven. 2006. “Ignorance Is Bliss: Or Seven Reasons Not to Use Uncertainty Analysis.” *Water Resources Research* 42 (5): n/a–n/a. doi:10.1029/2005WR004820.

Parker, Dawn C., Steven M. Manson, Marco A. Janssen, Matthew J. Hoffmann, and Peter Deadman. 2003. “Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review.” *Annals of the Association of American Geographers* 93 (2): 314–337. doi:10.1111/1467-8306.9302004.

Perrin, C., C. Michel, and V. Andréassian. 2001. “Does a Large Number of Parameters Enhance Model Performance? Comparative Assessment of Common Catchment Model Structures on 429 Catchments.” *Journal of Hydrology* 242 (3–4): 275–301. doi:10.1016/S0022-1694(00)00393-0.

Pontius, RG, D. Huffaker, and K. Denman. 2004. “Useful Techniques of Validation for Spatially Explicit Land-change Models.” *Ecological Modelling* 179 (4): 445–461. doi:10.1016/j.ecolmodel.2004.05.010.

Popper, Karl R. 1959. “THE LOGIC or SCIENTIFIC DISCOVERY.” http://terra-geog.lemig2.umontreal.ca/donnees/geo1512/POPPER_1961_Logic_chapter1.pdf.

Rastetter, Edward B. 1996. “Validating Models of Ecosystem Response to Global Change.” *Bioscience* 46 (3): 190.

Reddy, K. R., R. H. Kadlec, E. Flaig, and P. M. Gale. 1999. “Phosphorus Retention in Streams and Wetlands: A Review.” *Critical Reviews in Environmental Science and Technology* 29 (1). doi:10.1080/10643389991259182.

Refsgaard, JC, and HJ Henriksen. 2004. “Modelling Guidelines - Terminology and Guiding Principles.” *Advances in Water Resources* 27 (1): 71–82. doi:10.1016/j.advwatres.2003.08.006.

Refsgaard, Jens Christian. 1997. “Parameterisation, Calibration and Validation of Distributed Hydrological Models.” *Journal of Hydrology* 198 (1–4): 69–97. doi:10.1016/S0022-1694(96)03329-X.

Reynolds, JH, and ED Ford. 1999. “Multi-criteria Assessment of Ecological Process Models.” *Ecology* 80 (2): 538–553. doi:10.1890/0012-9658(1999)080[0538:MCAOEP]2.0.CO;2.

Rykiel, EJ. 1996. “Testing Ecological Models: The Meaning of Validation.” *Ecological Modelling* 90 (3): 229–244. doi:10.1016/0304-3800(95)00152-2.

Saltelli, A., S. Tarantola, and F. Campolongo. 2000. “Sensitivity Analysis as an Ingredient of Modeling.” *Statistical Science* 15 (4): 377–395.

Saltelli, Andrea. 2002. “Sensitivity Analysis for Importance Assessment.” *Risk Analysis* 22 (3): 579–590. doi:10.1111/0272-4332.00040.

Saltelli, Andrea, Marco Ratto, Stefano Tarantola, Francesca Campolongo, European Commission, and Joint Res Ctr Ispra. 2006. “Sensitivity Analysis Practices: Strategies for Model-based Inference.” *Reliability Engineering & System Safety* 91 (10-11): 1109–1125. doi:10.1016/j.ress.2005.11.014.

Scheffer, Marten, and Stephen R. Carpenter. 2003. “Catastrophic Regime Shifts in Ecosystems: Linking Theory to Observation.” *Trends in Ecology & Evolution* 18 (12): 648–656. doi:10.1016/j.tree.2003.09.002.

Schneider, Stephen H. 1997. “Integrated Assessment Modeling of Global Climate Change: Transparent Rational Tool for Policy Making or Opaque Screen Hiding Value-laden Assumptions?” *Environmental Modeling & Assessment* 2 (4) (December 1): 229–249. doi:10.1023/A:1019090117643.

Shaw, Raymond A. 2003. “PARTICLE-TURBULENCE INTERACTIONS IN ATMOSPHERIC CLOUDS.” *Annual Review of Fluid Mechanics* 35 (1): 183–227. doi:10.1146/annurev.fluid.35.101101.161125.

Shugart, Herman H. 1984. *A Theory of Forest Dynamics. The Ecological Implications of Forest Succession Models.* Springer-Verlag. http://www.cabdirect.org/abstracts/19850602322.html.

Van der Sluijs, JP, M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey. 2005. “Combining Quantitative and Qualitative Measures of Uncertainty in Model-based Environmental Assessment: The NUSAP System.” *Risk Analysis* 25 (2): 481–492.

STANLEY, TR. 1995. “Ecosystem Management and the Arrogance of Humanism.” *Conservation Biology* 9 (2): 255–262. doi:10.1046/j.1523-1739.1995.9020255.x.

Vanclay, J. K. 1994. “Modelling Forest Growth and Yield: Applications to Mixed Tropical Forests.” *School of Environmental Science and Management Papers*: 537.

Vanclay, J. K., and J. P. Skovsgaard. 1997. “Evaluating Forest Growth Models.” *Design, Performance and Evaluation of Models for Forest Stand Dynamics* 98 (1): 1–12. doi:10.1016/S0304-3800(96)01932-1.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta. 2003. “Towards Reduced Uncertainty in Conceptual Rainfall-runoff Modelling: Dynamic Identifiability Analysis.” *Hydrological Processes* 17 (2): 455–476. doi:10.1002/hyp.1135.

Wainwright, John, and Mark Mulligan. 2004. *Environmental Modelling: Finding Simplicity in Complexity*. Wiley.

Walker, W. E., P. Harremoës, J. Rotmans, der Sluijs van, M. B. A. van Asselt, P. Janssen, and von Krauss Krayer. 2003. “Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support.” *Integrated Assessment* 4 (1): 5–17. doi:10.1076/iaij.4.1.5.16466.

Warren, LA, and EA Haack. 2001. “Biogeochemical Controls on Metal Behaviour in Freshwater Environments.” *Earth-Science Reviews* 54 (4): 261–320. doi:10.1016/S0012-8252(01)00032-0.

Weaver, Andrew J., Michael Eby, Edward C. Wiebe, Cecilia M. Bitz, Phil B. Duffy, Tracy L. Ewen, Augustus F. Fanning, et al. 2001. “The UVic Earth System Climate Model: Model Description, Climatology, and Applications to Past, Present and Future Climates.” *Atmosphere-Ocean* 39 (4): 361–428. doi:10.1080/07055900.2001.9649686.

Young, P. C. 2002. “Advances in Real-time Flood Forecasting.” *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences* 360 (1796). doi:10.1098/rsta.2002.1008.