CrossMark

A checklist for model credibility, salience, and legitimacy to improve information transfer in environmental policy assessments

G.A.K. van Voorn [a, *], R.W. Verburg [b, c], E.-M. Kunseler [d, e], J. Vader [b], P.H.M. Janssen [d]

[a] Biometris, Droevendaalsesteeg 1, P.O. Box 16, 6700 AA, Wageningen, The Netherlands
[b] LEI Wageningen UR, P.O. Box 29703, 2502 LS, The Hague, The Netherlands
[c] Copernicus Institute of Sustainable Development, Faculty of Geosciences, University of Utrecht, 3584 CS, Utrecht, The Netherlands
[d] PBL Netherlands Environmental Assessment Agency, Antonie van Leeuwenhoeklaan 9, 3721 MA, Bilthoven, The Netherlands
[e] Institute for Environmental Studies (IVM), VU University Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

A B S T R A C T

Modelers involved in environmental policy assessments are commonly confronted with the lack of uptake of model output by policy actors. Actors have different expectations of models, condensed into three quality criteria: credibility, salience, and legitimacy. The fulfilment of quality criteria is also dynamic as expectations vary, change, and possibly counteract each other. We present a checklist for modelers involved in model-based assessments that is aimed at the identification and monitoring of issues, limitations and trade-offs regarding model quality criteria. It draws upon the literature of integrated assessments as well as case study analysis of environmental policy assessments for the Dutch government, based on expert interviews and embedded experience. The checklist is intended to be consulted during assessments; its application may result in greater awareness among modelers involved in assessments regarding model quality criteria, and may positively affect the uptake of model-based knowledge from environmental policy assessments by policy actors.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Models[1] are simplified representations of real-life systems that allow for the combination of various and heterogeneous sources of knowledge, such as process knowledge, observational and experimental data, and expert judgment. Commonly models play an integral role in environmental policy assessments, in which they can fulfil various functions. For instance, models can have a heuristic, symbolic or relational role (Sterk et al., 2011); they can be tools to quantify the effects of alternative developments and policy scenarios (Schmolke et al., 2010), function as 'boundary objects' for participants and prospective users[2] of the environmental policy assessment to communicate and learn from each other (Borowski and Hare, 2007, and references therein; Jakeman et al., 2006), or serve to frame the assessment.

Over the years the role of models in assessments has also been criticized, showing structural differences in perspectives, interests and attitudes between different participants and users of the model-based assessments (Borowski and Hare, 2007, and

---

* Corresponding author.
  E-mail address: george.vanvoorn@wur.nl (G.A.K. van Voorn).

[1] By 'models' we mean a whole range of conceptual models, mathematical models, simulation models, (spatial) databases, and indicators. In the context of environmental policy assessments the term 'model' often refers to a model chain, i.e. a set of models in which the output of one model is input to the next.

[2] We make a distinction between 'participants' and 'users' of environmental policy assessments. Participants are modellers and researchers, but also stakeholders who serve as experts. Users are primarily policy actors to whom results are delivered, but they can also be other stakeholders. The distinction is not necessarily very strict, but serves to indicate there are different groups around environmental policy assessments with different stakes, views, opinions, etc. towards the assessment goals as well as the models.

references therein). This seems to correlate with the observation that there is a common lack of uptake and use of scientific information by policy actors and other non-scientists (Bauler, 2012; Lemos et al., 2012). In this respect, Kunseler et al. (2015) refer to the 'effectiveness' of an assessment, which is considered to be an emergent property based on the expectations that participants and users have regarding scientific assessment processes. The effectiveness is a proxy of the knowledge transfer between the various groups (participants and users) involved in the assessments. Although modelers and researchers may have the common perception that information is useful to users (e.g. Turnhout et al., 2013), the latter group may (and often does) think otherwise. In addition, it may not always be clear to the users how to use the results of the assessments. As a result the effectiveness of many environmental assessments is not as high as it could be due to (implicit) discrepancies among the various groups on how they judge the assessment processes and the usability of the provided results and insights (Kunseler et al., 2015).

For knowledge to be taken up by users it is essential to meet various expectations that are put forward by different participants and users (Liu et al., 2008; Meinke et al., 2006). As models are important carriers and production units of knowledge, these expectations also apply to models in particular. Expectations with regard to knowledge (and hence models) can be roughly aggregated into three quality criteria (Cash et al., 2002; Lusiana et al., 2011; McNie, 2007):

- **Credibility** concerns the scientific logic of the model and the soundness of the used knowledge. A model is deemed credible when concepts and processes in the model are considered acceptable as an approximation of the modelled system;
- **Salience** concerns the societal and political relevance of the use of the model in the assessment. A model is deemed salient when it plays a significant role in understanding and solving the policy issue at hand; its input is relevant to the issue, and its output can answer research questions that have been brought up in the context;
- **Legitimacy** concerns a fair representation of the views, values and concerns of involved stakeholders in the model used in the assessment. A model is deemed legitimate when these aspects are dealt with in the model and its inputs in an adequate way.

These three quality criteria serve to discuss the effectiveness of assessments in general (Kunseler et al., 2015; Meinke et al., 2006; Schut et al., 2013), but are equally applicable to models in particular (White et al., 2010).

Examples in the literature seem to suggest that the effectiveness of models in environmental policy assessments may be higher when expectations by stakeholders on model credibility, salience and legitimacy are properly addressed by modelers. In case the model does not meet the expectations of different participants and users there is an increased risk that the model and its output — and perhaps the conclusions or recommendations of the whole assessment in which the model is used — will be subject to criticism or may not be accepted by some or all of the users. One example of this concerns the efforts in the Global Biodiversity Assessment (Cash et al., 2002). In that particular case, the limited attention to salience issues caused the primary intended audience (parties to the Convention on Biological diversity) to have little interest in the kind of questions that were being asked by the scientific assessors. Information relevant to their decision making was not produced, and the assessment was largely ignored by the intended audience (Cash et al., 2002, and references therein).

In this paper we build upon the assumption that by taking the various expectations of participants and users into account the credibility, salience and legitimacy of models used in assessments will increase. It is further assumed that a proper reflection upon the perceptions on model quality among the various modelers, users and stakeholders participating in the assessment process is a necessary condition to increase the 'effectiveness' of information. This is not likely to be a trivial matter. There are at least four points that modelers have to be aware of.

The first point is that the three quality criteria do not necessarily have to be equally satisfied but should be balanced in the context in which models are produced and used. For instance, Lusiana et al. (2011) interviewed 122 potential users of a resource management model of various backgrounds and found that salience (i.e. the relevance of the model) was considered to be more important than credibility. At the same time legitimacy is considered to be an essential requirement for scientific knowledge to be transferred to non-scientific actors and to be translated into 'actionable knowledge', i.e. the science-policy interface (Meinke et al., 2006).

The second point is that the three quality criteria can also be counteracting. Trade-offs can result, especially under restrictions such as resource limitations (Cash et al., 2002). For instance, Ginger (2014) explored two different dimensions of legitimacy in model-based environmental planning cases (legitimacy based on procedure, and legitimacy based on scientific expertise), and found the two to be counteracting each other. It is therefore important to be explicit on the various aspects of the quality criteria and to address potential trade-offs between them.

The third point is the variability between modelers, users and participants in their perception of model credibility. Established scientific practice is aimed at the publication of models in peer-reviewed journals that first and foremost assess the scientific originality of the models (Schmolke et al., 2010). Furthermore, there are sets of standards for 'quality', such as the use of SI (International System of Units) units and following good modelling practices (STOWA/RIZA, 1999), which is coupled to the well-known model development cycle (Jakeman et al., 2006; see section 5 of this paper). Even if modelers have a shared view on credibility this does not exclude the possibility that non-modelers have different views.

The fourth point is that perceptions of the three quality criteria seem to be dynamic in nature (Sarkki et al., 2015) and may even be path-dependent. For example, Schut et al. (2013) discuss a case study in which at some point the credibility and legitimacy of specific research, which was earlier judged to be credible, salient, and legitimate, was openly questioned and contested by stakeholders as a result of interactions with partners that were not trusted. The dynamic nature of the quality criteria is also demonstrated by the case study we present in section 3 in this paper. It is conceivable that similar shifts can occur in the credibility, salience and legitimacy of models over the course of an assessment. This suggests that a monitoring of the quality criteria should occur at regular intervals.

This paper aims to develop a practical tool with suggestions for good practice for model developers to identify, avoid and help deal with issues regarding credibility, salience and legitimacy. It could also help in creating awareness among modelers involved in assessments, in particular concerning the four above-mentioned points. The selected format is a checklist with items for modelers to go through at a regular interval during model development and application. The evaluation of a model regarding the quality criteria cannot be seen separately from the policy issue context in which the model is developed, analyzed and applied. While not specifically aimed at models, trans-disciplinary approaches (implicitly) cover many aspects of model credibility, salience and legitimacy in suggesting ways for dealing with unstructured problems, i.e. problems in which there is great diversity and lack of knowledge

about both the societal discourses and the scientific aspects (Hoppe, 2002). Also, the role of stakeholders (as participants and users) in modelling is increasingly recognized, e.g. in participatory modelling and trans-disciplinary science (Voinov et al., 2016), as it is assumed that it increases the quality of models. The checklist therefore adopts a broader view than the well-known model development cycle and also incorporates elements of participatory modelling and trans-disciplinary science (see section 5).

The paper is organized as follows. In section 2 we introduce our research approach underlying the development of the checklist. In section 3 we present a case study (that is discussed in more detail in Appendix A) and take a close look at the assessment processes of the Nature Outlook (NO) studies produced by the Netherlands Environmental Assessment Agency (PBL). Over the course of two decades the different editions of the NOs show clear changes in the complexity of the used models, followed by a shift in focus brought about by socio-political changes. These changes have had consequences in terms of credibility, salience, and legitimacy, and the case study illustrates how issues regarding the model quality criteria (including the four above-mentioned points) were experienced and addressed in practice. In section 4 we make an inventory of different factors affecting credibility, salience, and legitimacy that have been identified. In section 5 we provide the checklist, both in text to discuss the consecutive checklist items step by step based on the findings reported in section 4 and in summarized form as a table. The checklist has been fitted as much as possible to the well-known model development cycle for purposes of familiarity for modelers but is also inspired by guidelines available on trans-disciplinary science (Jahn, 2008; Bergmann et al., 2012; Lang et al., 2012; Jahn and Keil, 2015). In section 6 we discuss how the checklist may improve the transfer of knowledge from model output to policy actors and how it can be applied in practice. Section 7 provides some concluding remarks.

2. Methodology

The development of the checklist presented in this paper is based on an iterative cycle with a qualitative research design (Creswell, 2003), i.e. no formal statistical procedures have been used. As a start issues regarding model credibility, salience and legitimacy were identified from peer-reviewed and grey literature, as well as from the findings from our analysis of the NO case study (see next section). In addition, literature on conceptual models on how assessments are performed (in particular on trans-disciplinary science) was studied. This information was used to develop a first version of the checklist. The checklist was then validated and adapted according to findings from interviews with researchers with appropriate expertise, as well as additional literature findings.

Three data sources have been used. The first data source is scientific and grey literature on factors influencing the three model quality criteria and how assessments are (or should be) performed. Relevant literature was found with Scholar Google (https://scholar.google.com/) using search terms like 'model credibility', 'salience', 'legitimacy', 'modelling with stakeholders', 'model development cycle', 'interdisciplinary', 'trans-disciplinary', and 'participatory modelling'. Backtracking of references and (new) key words was used to follow promising leads. Occasionally additional literature was provided by people who were interviewed (see below). The different factors and the terms 'credibility', 'salience', and 'legitimacy' are not always explicitly mentioned in the literature; instead terms like 'stakeholder acceptance' are used, or they are mentioned only implicitly or casually in the discussion of case studies. In particular literature on modelling in water management, climate and agricultural assessments, and the role of models in trans-disciplinary science turned out to be of use. We also paid special attention to existing model development checklists or guidelines and reported case studies in which models were developed or model outputs were used for environmental policy development.

The second data source is provided by findings from interviews that were conducted in two rounds from August through November 2013 with 16 (anonymized) researchers involved in the Nature Outlook studies of the PBL. These researchers were all (former) employees at PBL or Wageningen University and Research and have professional experience with model-based environmental assessments and/or are trained in philosophical aspects of model development or communication between different stakeholders. An interview guide was prepared inspired by trans-disciplinary research questions (see Appendix B, translated from Dutch to English). Interviews were recorded on laptop (mostly in Dutch) and then transcribed; the transcripts were presented for member checking, as standard practice for ensuring validity in qualitative analysis (Creswell, 2003).

The third data source is the case study analysis that is presented in section 3 and Appendix A, in which all authors have been involved. The experience and proximity of the authors to the NO case study[3] served to gain insight into (partial) processes of real-life assessments (Flyvbjerg, 2006) by reconstructing the assessment processes of the NO. This allowed for the identification of issues emerging during the assessment processes and factors influencing model quality in this specific case. The case study also illustrates several of the previously mentioned points that modelers need to be aware of, in particular the dynamic nature of model credibility, salience and legitimacy. Our embedded research is informed by interpretive and naturalistic inquiry (Lincoln and Guba, 1985).

3. The Nature Outlook (NO)

The case study is introduced here for two reasons. First, it demonstrates several of the points that modelers have to be aware of regarding model quality criteria. In particular, this case study demonstrates the dynamic nature of how model quality criteria are fulfilled during a model-based environmental assessment. Second, it provides insight into factors that affect model quality criteria that serve as input to the development and evaluation of the checklist. The case study is discussed briefly here; for more detail we refer to Appendix A.

Since the end of the 1990's, PBL (the Netherlands Environmental Assessment Agency) has published the Nature Outlook at a regular interval, which is the primary nature policy assessment for the Dutch Ministry of Economic Affairs (responsible for nature-related policy issues). Four editions have appeared so far (1997, 2002, 2007, and 2012, henceforth abbreviated as NO1, NO2, NO3, and NO4, respectively) while currently (2015/2016) PBL is preparing a new Nature Outlook with a European scope (to be published in 2016). The different NOs are highly 'model-driven', i.e. models play a prominent part in these assessments. When analyzing the different NO editions two major trends become apparent. First, there is a clear trend of increasing model complexity in the earlier NO editions, followed by the adoption of a meta-modelling approach in later editions. Second, while the first editions were relatively mono-disciplinary, the later editions have shifted towards a multi-

---

[3] The authors of this paper were involved in the NO case study in the following ways. GVV is involved in the quality assessment of simulation models used in the NO. JV performed an analysis of the use and usefulness of different NOs, and interviewed 46 people in total from government, education, business, NGO's and research institutes (Vader et al., 2004; Vader and Bogaardt, 2014). EK evaluated the innovative features of the NO4 process, based on interviews with 22 people (project members and stakeholders) and document analysis (Kunseler et al., 2015). RV, JV and PJ were project team members of Nature Outlook 4.

or trans-disciplinary approach with a focus on policy studies and the involvement of various stakeholders.

With regard to the first trend, there is a clear evolution in terms of model complexity. This can be perceived as an attempt to improve the credibility of the models used in the NO. The models tended to become more complex to meet the desires to scientifically understand the ecological processes. However, at some point it was felt that the model complexity had over-shot, which eventually led to the questioning of model salience by researchers. This was primarily because simulation results took too much time: the increased running time presented severe limitations regarding model testing, validation and evaluation. To improve salience, model complexity was reduced by implementing a meta-modelling approach. This approach decreased simulation efforts and therefore model results were delivered faster. However, it also led to a new limitation. Together with the meta-modelling novel scientific and societal concepts were introduced in the nature policy debate. These included the concepts of ecosystem services and recreational values of nature areas. Implementing such concepts in the modelling would eventually strongly increase salience, yet time and budget limitations prevented appropriate testing and validation of the implementation of these concepts in the models. This undermined the credibility of the new models. In short, resource constraints invoked a trade-off between model quality criteria, corresponding to the findings reported by Cash et al. (2002).

With regard to the second trend, the early editions of the NOs embarked on global scenarios in which socio-economic drivers were the main focal point. One of the drawbacks of such an approach was the impracticability of implementing new ideas on nature policies within these scenarios. End users like policy makers had difficulties with the scenario approach, because the assessments did not provide much room for the development of policy options with regard to the subject of the study (namely biodiversity). Another issue related to salience was the long time horizon that was much ahead of political reality (e.g. Vader et al., 2004). By the time NO4 was published the nature policy was much more at the center of societal debate, and the role of scenarios in the NO was shifted towards a more 'normative' characteristic than an 'exploratory' one. A set of normative scenarios (named 'perspectives') was developed that reflected various societal views on nature. Different discourses, views, and values from different stakeholders were included in these 'perspectives'. This improved the legitimacy of the overall study – and the use of NO information by policy actors – but at the same time reduced the salience and legitimacy of the standard models used in the earlier NO editions. The shift away from the policy frame that was adopted in earlier editions of the NO (with biodiversity as the sole aim) also implied that other models would have to be developed and used to underpin the whole assessment. The legitimacy and salience of the newly developed models was considered higher than that of the standard models because of the stakeholder involvement in the discussion of model output.

Both trends (i.e. changes in model complexity, and the shift towards normative scenarios) demonstrate the point that the three model quality criteria are not equally satisfied and are not necessarily balanced (awareness point 1). Also, they illustrate the dynamical nature of the fulfilment of model quality criteria (awareness point 4). In fact, the whole process from NO1 through NO4 seems to illustrate a search for balance between the model quality criteria. In addition, the case study clearly demonstrates how model quality criteria can counteract each other and trade-offs may appear as a result from resource limitations (awareness point 2). Finally, it seems that many of the dynamics are driven by differences in the views of participating actors (researchers, policy makers, and later on also other stakeholders) regarding model credibility, salience, and legitimacy (awareness point 3).

By analyzing the dynamics in the NO we gained insight into the drivers for changes, the issues arising, and how they were dealt with (if at all). The findings of this analysis are reported at the end of Appendix A. It is one of the data sources used in this paper to derive a list of factors that influence model quality criteria. This list is presented in the next section.

4. Factors influencing model credibility, salience, and legitimacy

In this section we list various factors that influence how credibility, salience, and legitimacy of models and their use in assessments are interpreted and can dealt with. An overview of the listed factors can be found in Table 1A–C. The factors have been identified from the lessons learnt from the analysis of the NOs (see previous Section and Appendix A), the findings from literature review of scientific and grey literature, and expert interviews.

In Table 1 a distinction is made between factors in three ways. First, there is a division of factors based on the principal model quality criterion they affect. Table 1 is therefore divided into three subcategories (A through C) for this reason. Factors that affect model credibility are listed in Table 1A, those that affect model salience are listed in Table 1B, and those that affect model legitimacy are listed in Table 1C. In practice factors may affect more than one quality criterion, which may result in trade-offs (Cash et al., 2002). Second, there is a division between the generally positive and negative effects of factors. It should be noted that often the same factor can have a positive or negative effect, depending on the context. In Table 1A–C the positive factors are listed before the negative ones. Third, a division is made between factors based on the 'sphere of influence' of the modeler. One category contains factors that are reasonably within the scope of influence of the modeler and hence can and – we suggest – should be dealt with by effort of the modeler; in Table 1 we refer to this category as 'solvable' for the modeler. The other category contains factors that the modeler should be aware of but that are not within the direct range of responsibilities of the modeler; we refer to this as 'not solvable' for the modeler. The category of 'not solvable' also contains for instance institutional and organizational factors (concerning management barriers, funding possibilities, geographical barriers, etc.) that affect the uptake of information by non-scientific actors (Lemos et al., 2012), which we have not included in the list because of their indirect effects on model quality.

In the next section the listed factors and the lessons learned in the NO study case are used to develop the checklist for modelers.

5. Checklist for balancing model credibility, salience, and legitimacy

We introduce here a checklist for modelers to assess their model with regard to quality criteria and possibly enhance the model performance towards any of them (see Table 2 for the abbreviated form of the checklist). The checklist consists of a number of items. These are based on the table of factors that were identified in our case analysis and literature review in the previous section. Each item is put as guiding statement and informed by several guiding question. These statements and questions are explained in relation to the factors reported in section 4.

From the perspective of a modeler, the checklist can be compared to the well-known model development cycle (see Fig. 1). This cycle however does not fully cover the model development 'environment', and does not include all relevant aspects of credibility, salience, and legitimacy. Traditional model quality is focused

Table 1A

Factors that positively or negatively affect credibility. Key to footnotes: [1]The model may be practically infeasible to validate or verify (Oreskes et al., 1994), [2]The model complexity may be required by its application and may therefore be irreducible, [3]This may be solved by increasing data availability (if available) or reducing model complexity (if possible), [4]Modellers may not be capable of convincing certain stakeholders for various reasons.

Credibility (concerns the scientific logic of the model and the soundness of the used knowledge and information)			
Item	Factor	Main effect	Solvable?
C1	The model concept has been discussed and is according to well-established scientific knowledge (Schmolke et al., 2010).	+	yes
C2	There is a clear description and justification of assumptions and simplifications in the model, and the limitations are clear (Liu et al., 2008). Modelers will make assumptions and simplifications based on their disciplinary training and experience that are not necessarily shared by peers or stakeholders (Krueger et al., 2012). The communication of assumptions and simplifications and explaining why they are made allows stakeholders to provide feedback.	+	yes
C3	The inputs and scenarios of the model have been closely discussed — which requires an adequate verbalization — and are based on well-established scientific knowledge (Girod et al., 2009).	+	yes
C4	A quality assurance framework has been implemented: there is a proper description of the model, model input and output, and the management of data and model software is formalized.	+	yes
C5	The model has been successfully tested and verified, calibrated, and validated for purpose (i.e., meets specified performance requirements; Rykiel, 1994; Schmolke et al., 2010), in accordance with the modelling cycle (Refsgaard and Henriksen, 2004). For validation independent data have been used (Liu et al., 2008). Proper performance criteria have been selected to evaluate the model behavior (Bennett et al., 2013);	+	yes/no[1]
C6	The model input and scenarios and the model itself have been subjected to uncertainty analysis (Walker et al., 2003; Liu et al., 2008).	+	yes
C7	The model complexity has been judiciously increased to make the model more flexible in use, in order to increase model utility and confidence (Chwif et al., 2000);	+	yes/no[2]
C8	Poor performance of the model, as reflected by poor model validation results; The lack of reporting of validation efforts; Limited cross-validation character of data used for testing the model-performance, e.g. by re-using data that have previously been used in steps of the modelling process (e.g. model design, or calibration) to 'validate' the model;	−	yes/no[1]
C9	Over-shot model complexity that decreases confidence in the model. The loss of confidence can have resulted e.g. from a loss of understanding of how the model functions, a loss in the capacity to properly analyze the model (Chwif et al., 2000), or a lack of available data for model calibration (Wagener et al., 2001). Model complexity can for instance increase as a result of model adaptation to meet new requirements;	−	yes/no[3]
C10	The existence of pre-conceived and perhaps immovable conceptions among stakeholders may make it difficult for modelers to establish model credibility (Alexandrov et al., 2011);	−	yes/no[4]

around technical aspects of model credibility, such as model verification and validation, which is based on generally accepted standards for mathematics, science, and software engineering (see e.g. STOWA/RIZA, 1999). Existing checklists for model quality are accordingly based on (a variation of) the model development cycle. Model salience primarily targets the role and the use of the model in the assessment, while model legitimacy considers the views, values, interests, and concerns of the participants and users that are involved in the model development as well as the assessment. These aspects appear in steps 1, 3, and 4 of the model development cycle, but indirectly and implicitly. Hence this cycle does not properly touch upon the four points mentioned in the introduction in general (balance between model quality criteria, possible counteractions and trade-offs, variability in actor views on the quality criteria, and the dynamic nature), or many of the factors regarding model quality criteria specifically.

We therefore have to consider the scope of the model development cycle to be a broader one — in particular regarding steps 1, 3, and 4 — to deal also with model salience and legitimacy. For example, participants and users can and — we argue — should play an important role in the framing in step 1, and thus determine the role of the model in the assessment, as well as the uncertainty analysis. Also, the model validation can be based on scientific evaluation, but to participants and users it may be more important that the model is the right model for the role it is intended to fulfil in the assessment. The extent of the checklist is therefore taken broader, as it regards not only the model itself but also its 'environment', i.e. the whole assessment.

Typically assessments make use of more than one model, but the checklist presented below assumes there is one model under evaluation. At various points references are made to factors listed in Table 1. It is recommended to evaluate the different models separately, in particular when the roles of the different models in the assessment vary, but also to consider the interconnection and integration of models.

5.1. Describe the goal of the assessment. What is the issue, and what is the chosen frame?

Environmental outlooks typically explore impacts of future scenarios and policy designs. The first task is to clarify the policy issues at stake and the framing of these issues. The way issues are framed can affect the link between knowledge and policy action, the shape of the 'decision space' (room for policy makers to maneuver and their options for action according to the chosen frame), and which actors are considered to be important or not (White et al., 2010). The terminology used by participants and users should be brought in line in this stage (i.e. by establishing a common ontology, i.e. a shared and unambiguous terminology). A proper translation should occur of societal issues and policy evaluation questions into scientific questions or testable hypotheses in order for modelling to be possible (Bergmann et al., 2012; Kampen and Tamás, 2013). A clear and relevant frame, with common ontology and proper translation of policy questions to research questions will improve model salience (S1). The establishment of e.g. ontology together with stakeholders will improve model legitimacy (L1). Including end users as stakeholders in the early phases of the model-based assessment and model development will increase ownership and thus salience (S13) and may help avoid shifts in views on the model quality criteria later on resulting from the ad hoc inclusion of stakeholders. It should be avoided that specific stakeholders have too strong a voice to avoid degradation of legitimacy (L12).

A pitfall for any assessment is that the policy issue at hand may be 'wicked' (or unstructured), but is framed as a so-called 'structured policy problem' (cf. Hoppe, 2002). The latter term refers to a policy problem in which scientific knowledge is undisputed, agreed upon with little uncertainty while agreement among participants and users exists towards the frame of the social context of the policy problem. Assessments often involve unstructured policy problems, while even the problems themselves can be in dispute.

Table 1B
Factors that positively or negatively affect salience. Key to footnotes: [5]Descriptions may easily become outdated when the rate of stakeholder dynamics and discourse evolution is high, and it may prove difficult to keep them updated, [6]This expertise may not be (commonly) available (Allen et al., 2013), [7]A pre-requisite is that possible end users can be identified in the early phases.

	Salience (concerns the societal and political relevance of the use of the model in the assessment)		
Item	Factor	Main effect	Solvable?
S1	There is a clear framing of the policy issue (i.e. the policy issue is structured), and the frame is considered relevant by the involved stakeholders, or it is agreed that the policy issue is unstructured and needs to be tackled. There is a common ontology (a shared and unambiguous terminology between users). The key policy questions are clear, and these have been translated to clear scientific questions (Liu et al., 2008; Bergmann et al., 2012; Kampen and Tamás, 2013).	+	yes
S2	The model plays a role in the success of the assessment according to criteria that have been set in the early stages of the assessment (Bergmann et al., 2012; Podestá et al., 2013). For example, a criterion for success is that certain management objectives are met, and the model has contributed to obtain this result.	+	yes/no[5]
S3	Model inputs and scenarios have been developed such that they match the framing of the issue. For instance, the scenarios and inputs of the model used in an outlook of future policy effects are capable of quantifying the effects of non-intervention and intervention futures (Girod et al., 2009). An appropriate design approach has been adopted to design the required types of scenarios (Börjeson et al., 2006).	+	yes
S4	There is a clear description of the intended role of the model. Outlooks and assessments can differ greatly in their goals and scope; a description of the role of the model is based on the framing of the policy issue and the defined scientific questions. The model is up-to-date, i.e. it has been developed or adjusted such that it matches the intended role. The expectations towards the model are clear.	+	yes/no[5]
S5	There is a clear description of the limitations of the model input and scenarios, and the model itself (including the output). This includes aspects about spatial extents and time horizons (Liu et al., 2008), validity, and the theoretical and empirical underpinning of the model.	+	yes/no[5]
S6	Natural resource management decisions commonly face trade-offs between costs of remedial actions and losses resulting from inaction that depend in a nonlinear way on model output, often with critical thresholds. Conveying information about uncertainties in model output to the decision makers can change decisions (Bastin et al., 2013), and thus increases model relevance.	+	yes
S7	Model inputs and scenarios are comprehensible for stakeholders (Girod et al., 2009), as are the model outputs. Model outputs match the requirements set in collaboration with stakeholders (Schmolke et al., 2010), and provide answers to the defined scientific questions.	+	yes
S8	The model meets specific expectations by stakeholders. These are for instance with regard to speed or accuracy, e.g. for making on-the-spot calculations (for instance in serious gaming), or about the model's ability to properly deal with the effects of non-intervention and intervention futures in combination with involved scenarios and inputs (Girod et al., 2009).	+	yes/no[5]
S9	The 'model-on-the-shelf' phenomenon: Under severe time constraints and in the absence of relevant models it can be very appealing to make use of models that are readily available, despite that these may have been developed with a different purpose in mind. In practice there are always differences in the type of knowledge required for different assessments. The straightforward re-use of models should be discouraged; instead it should be properly judged whether available models meet stakeholder demands before they are being used.	−	yes
S10	Model role descriptions that have not been agreed upon, which are undocumented, or subject to change over the course of the assessment. Models may then be found to (no longer) bear sufficient relevance to the policy issue. Misunderstandings about the role and importance of models between modelers and other stakeholders can decrease effectiveness (Borowski and Hare, 2007).	−	yes
S11	Delivering model output without consideration of sources of uncertainty (Refsgaard et al., 2007; Liu et al., 2008). The result may be that the output cannot be used to answer the research question.	−	yes
S12	An increased model complexity necessary to describe environmental systems (which can improve model credibility) requires a higher level of technical expertise to promote the application and usability (salience) of the models for decision-making	−	yes/no[6]
S13	Refraining to include end users in early phases of the project. Such users may not feel obligated to contribute to the modelling or make use of the output (McIntosh et al., 2011). This may in turn reduce model salience as the input and feedback of these end users is missing in the development;	−	yes/no[7]

This means there is little to no consensus or certainty about the norms, values, views etc. of various participants and users regarding the issues at stake (affecting model legitimacy), and/or there is significant uncertainty with regard to the relevant scientific knowledge (affecting model salience and credibility). This implies that views on model quality criteria may also change during the assessment, and model developers may be faced with trade-offs and changes regarding model development and application. Although a modeler cannot be expected to take a lead in dealing with the structure of a policy problem, he/she should certainly be aware of how this pitfall can be conditioned by the (lack of) available models. Many systems are inherently complex with many sources of uncertainty, and allow for many different stakeholder views and concerns. The complexity and uncertainty should be properly considered for model legitimacy (L7) to avoid an improper balance between model quality criteria. The definition of a limited set of clear scientific questions together with participants and users helps in gaining a stronger focus in the research — and modelling — activities (Liu et al., 2008) and possibly avoid large shifts in views on model quality criteria. By making a clear linkage between these questions it is also more likely that the integration at the end is better (Bergmann et al., 2012). The modeler can steer efforts towards obtaining clear and clearly linked research questions by pushing for a choice of frame (and documenting the justification of this choice), the establishment of a common ontology, and allowing for sufficient feedback between the involved actors. This reduces the probability that stakeholders feel a lack of ownership (L14) and that the project team may even dissolve (L15), and may reduce the probability that views on quality criteria change.

5.2. Describe who the stakeholders are, what they want with the model and the assessment in general, and what their role and (potential) influence is. Provide a justification for the inclusion and exclusion of stakeholders

As discussed earlier stakeholder involvement can be crucial for policy issues dealt with in an environmental assessment. For instance, stakeholders may help in identifying conflicts and improving the management of socio-ecological systems (S6), and their involvement can improve model legitimacy (L2). In addition, stakeholders may hold a different view on model credibility than model developers. It is getting more common that stakeholders are involved in different roles (i.e. as participant or user) in modelling exercises. If stakeholders are to be involved, they should certainly be involved in the description of ontologies and provide information about their expectations towards the model, e.g. about what they expect with regard to the frame, the type of output, and the acceptable level of uncertainty in the output (L1, L2, L4, L6, L7). These expectations can vary depending on the intended role of the model as well as on the stakeholder views, values, concerns, and

Table 1C

Factors that positively or negatively affect legitimacy. Key to footnotes: [2]The model complexity may be required by its application and may therefore be irreducible, [3]This may be solved by increasing data availability (if available) or reducing model complexity (if possible), [5]Descriptions may easily become outdated when the rate of stakeholder dynamics and discourse evolution is high, and it may prove difficult to keep them updated, [8]A modeler may have no control over stakeholder involvement for various reasons, e.g. because of institutional organization, [9]This may be partly addressed by setting up an effective interaction framework, but severe resource limitations may prevent this possibility.

Legitimacy (concerns a fair representation of the views and concerns of involved stakeholders in the model)			
Item	Factor	Main effect	Solvable?
L1	A common ontology has been established between scientists and stakeholders (Liu et al., 2008; Bergmann et al., 2012) – which may take some time. This aids effective and efficient communication between scientists and stakeholders.	+	yes
L2	Stakeholder-based modelling aids the identification of conflicts and the improvement of management of socioecological systems (Angelstam et al., 2013). There is a broad and balanced participation of relevant actors. There is a documented inventory of the views, concerns, and demands by various relevant stakeholders, and these have been translated into demands with regard to the model.	+	yes
L3	There is maintenance of active and continuous support to and dialogue with stakeholders (Liu et al., 2008; Schmolke et al., 2010). There is frequent and transparent communication with stakeholders about the model and its input, scenarios, and output (Volkery et al., 2008; Girod et al., 2009). Interactions between scientists and stakeholders are frequent, brief, focused, and repetitive in nature (Podestá et al., 2013). Participatory involvement or trans-disciplinary approaches are recommended when policy issues are highly unstructured (cf. Hoppe, 2002; Girod et al., 2009; Mahmoud et al., 2009; Bergmann et al., 2012).	+	yes
L4	The model inputs and scenarios cover all views and concerns relevant to stakeholders in an appropriate way, e.g. different scenarios are considered to cover possible ranges of drivers relevant to the policy issue (Mahmoud et al., 2009), and stakeholders agree with the use of the input and scenarios.	+	yes
L5	The model concept has been developed in strong collaboration with stakeholders (Angelstam et al., 2013). It covers all views and concerns relevant to stakeholders in an appropriate way and all stakeholders agree with the use of the model.	+	yes
L6	The model output types are designed according to requirements discussed with stakeholders (Schmolke et al., 2010). Model outputs are comprehensible to all stakeholders (White et al., 2010), and the contribution of model output to the policy issue is broadly recognized.	+	yes
L7	There is a proper consideration of sources of uncertainty and discussion with stakeholders about sources of uncertainty and uncertainty estimates (Janssen et al., 2003; Refsgaard et al., 2007; Liu et al., 2008; Mahmoud et al., 2009; Schmolke et al., 2010).	+	yes
L8	Delivering the model and/or its inputs and outputs as reductionist science, i.e. without any consideration of stakeholder views or concerns (Meinke et al., 2006; Volkery et al., 2008). This may also produce a kind of false legitimacy as models may appear objective but in fact can contain hidden values and assumptions (Ginger, 2014). The 'model-on-the-shelf' phenomenon (see salience) also applies here.	–	yes
L9	Increased model complexity may lead to a loss of procedural legitimacy because more complex models are also less accessible (Ginger, 2014; see also factors that affect model salience).	–	yes/no[2,3]
L10	Poor monitoring of stakeholder involvement, or the use of improper formats for stakeholder involvement.	–	yes/no[5,8]
L11	A bias in the involvement of stakeholders, e.g. occurring through snow-balling (the inclusion of people through networks of those already involved) or the inclusion of only the 'usual suspects'.	–	yes/no[8]
L12	The sole adoption of a democratic principle to make decisions in the assessments (for instance, on establishing ontology or model aim). This is a form of procedural legitimacy that may harm legitimacy based on (scientific or other) expertise, especially in the case where stakeholders are involved with strong advocacy goals (Ginger, 2014).	–	yes/no[8]
L13	The dominant involvement of stakeholders who are known to oppose each other and of who will likely not come closer (cf. e.g. Schut et al. (2013), or a failure to report the existence of opposing views. The development of scenarios and models can easily be stalled by the existence of diverging interests, conflicting views, and hidden agendas (Volkery et al., 2008).	–	yes/no[8]
L14	A 'lack of ownership' – also referred to as 'not invented here'. Decisions are implemented with less conflict and more success when stakeholders are directly involved (Voinov and Bousquet, 2010).	–	yes/no[8]
L15	A lack of frequent and repetitive interactions with stakeholders. This tends to result in the dissolution of project teams into smaller units (Podestá et al., 2013), in turn preventing the input and response from stakeholders to new developments;	–	Yes/no[9]

desires. A clear overview of stakeholder expectations helps in developing a model that balances the model quality criteria, and may expose trade-offs that have an origin in model legitimacy. A failure to involve stakeholders in model development is a major threat to the legitimacy of models (L14). Stakeholders can also have other roles, for instance they can play a role in the framing of the problem and the design of the conceptual model, provide acceptance criteria for model testing and validation, or provide data that is used in model calibration. This should be clear as well.

The involvement of stakeholders will generally improve model legitimacy (L5) and also salience. This requires that a stakeholder analysis is performed to gain insight into the views of stakeholders on the subject, and on their potential role and contribution in the assessment process. Such an analysis can be combined with

Table 2

The checklist for modelers regarding model quality criteria, consisting of the items listed in the main text. Each item contains several action (bullet) points. The different action points in one item do not necessarily have to be executed in the exact order they are listed; rather they are to be executed iteratively or simultaneously. In particular item 6 should be considered from the start of the assessment. Likewise, the items themselves should not be viewed as static, but rather checks at regular intervals during the modelling process should occur to evaluate each of the items.

Item	Modeler's actions
1 Describe the goal of the assessment	■ Determine if the assessment issue is wicked/unstructured or not; ■ Establish a common ontology in collaboration with stakeholders (participants and users) and document this; ■ Determine what the frame of the assessment is in collaboration with stakeholders and document this; ■ Determine in collaboration with stakeholders what the policy questions are, how they are translated to research questions and document this.
2 Describe who the stakeholders are, what they want with the model, and what their role and influence is	■ Push to perform a stakeholder and discourse analysis a priori to the project to explore the contribution of relevant stakeholders in the development of the assessment in general and to the model (and/or scenarios/input/output) in particular; ■ Identify possible stakeholder bias and immovable stakeholder perceptions; ■ Push to have documentation generated that includes stakeholder role description as well as justification thereof.
3 Describe the role of the model in the assessment	■ Include a clear description of the model goal in the model documentation, including a reference to documentation where this goal has been decided upon together with stakeholders; ■ Make an inventory of stakeholder views and expectations regarding the role of the model; ■ Perform a check in collaboration with stakeholders of the current model version in relation to the applicability to the objectives of the assessment; ■ If applicable, establish in collaboration with stakeholders a plan for possible adaptations of the current model or for a new model design; ■ Check at a regular basis if the objective of the assessment has not changed (intentionally or not) such that the model has to be revised.
4 Describe the model scenarios including key assumptions, and how they have been tested against logical structure and socio-political reality	■ Develop scenarios in collaboration with stakeholders. Document all assumptions, and their justification; ■ Check all scenarios against scientific knowledge, and the logical and socio-political reality; ■ Provide an uncertainty analysis regarding scenario input and assumptions; ■ Check if the scenarios can be handled with the available model and if not, discuss alternatives with stakeholders; ■ Determine clear descriptions of who delivers what.
5 Describe the model and its relevant output	■ Design the model in collaboration with stakeholders, and document all model assumptions, and their justification; ■ Check the model against scientific knowledge, and the logical and socio-political reality; ■ Perform model testing, verification, validation, and uncertainty analysis, and communicate model limitations and uncertainties to stakeholders; ■ Determine together with stakeholders the decision space (e.g. existing trade-offs); ■ Present model output in such a way that it is comprehensible to stakeholders, e.g. by designing 'policy levers' in collaboration with stakeholders; ■ Work on a common ground to deliver joint ownership of assessment results (of which the model output is a part of).
6 Analyze the project process and planning and identify project limitations	■ Identify possible limitations and trade-offs within the assessment to model quality criteria; ■ Determine which limitations and trade-offs are solvable or not, and document and communicate the findings; ■ If trade-offs exist, justify choices regarding the possible improvement of one quality criterion at the expense of the other quality criteria; ■ Adopt an appropriate format for involving stakeholders, and maintain an active and continuous dialogue with stakeholders; ■ Divide available resources between model development and use, and stakeholder involvement

discourse analysis unravelling the 'stories' behind framing the problem. A lack of monitoring stakeholder involvement, or the use of an improper way to involve stakeholders, presents a threat to model legitimacy (L10), and may result in a lack of ownership among end users (L14) and even dissolution of the project team (L15). A modeler can 'push' a stakeholder analysis and should do so as early as possible in the process to reduce the risk of changes in model legitimacy. Any 'sampling bias' in the stakeholder population (e.g. through 'snow-balling' or the selection of the 'usual suspects'; L11) should be avoided in order to have a fair representation of stakeholder views and concerns, but types of representativeness can be contested (c.f. Warleigh, 2003; Ginger, 2014; L12, L13). Nonetheless, justification of stakeholder inclusion and exclusion should be made in relation to the modelling process and the assessment in general. There may be reasons to exclude

stakeholders (in certain steps), for instance resulting from a lack of contribution in terms of knowledge, or incapability or unwillingness for joint learning. The existence of immovable conceptions among certain stakeholders can make it difficult to establish model credibility (C10).

5.3. Describe the role or function of the model in the assessment. What are the demands for model design and model use?

Depending on the type of assessment and the question that has to be answered the role or function of a model can differ. Three roles are distinguished in a multi-stakeholder context (cf. Sterk et al., 2011), namely a heuristic role for learning, a symbolic role to put issues on the political agenda, and a relational role to create communities. In particular for models in a heuristic role a further
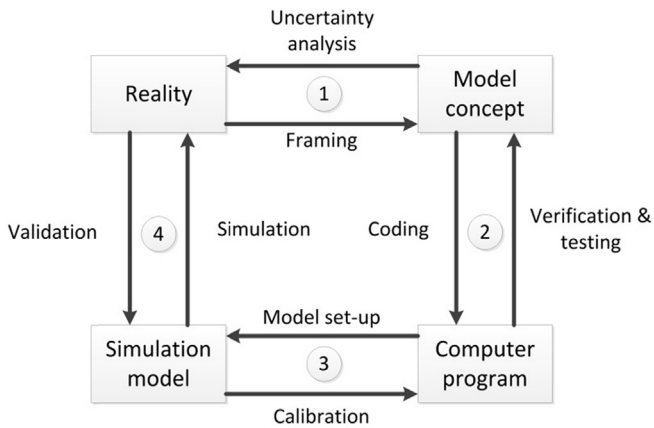
Fig. 1. Depicted is a conceptualization of the modelling cycle (after Refsgaard and Henriksen, 2004). It should be noted that the term 'cycle' should be considered in a dynamic way here, as there are typically feedbacks. The cycle serves to secure the credibility of modelling exercises. In many steps participants and users may be included, which may improve the model salience and legitimacy. 1/ The model concept is designed based on reality by framing, while uncertainty analysis (critical reflection) improves the model concept. Both can be done together with participants and users. 2/ The model concept is codified as a computer program, which is improved by testing and verification. In this step the role for participants and users is limited to providing user acceptability criteria and software utility demands. 3/ The computer program is adapted as a simulation model. In this adaptation, calibration may be used. Participants and users can provide data, expert opinion, and criteria of acceptance. 4/ The simulation model is run to mimic reality. Model validation for its purpose in the assessment by participants and users can be based on scientific evaluation, but also on grounds of salience and legitimacy (is it the right model?).

specification can be made, e.g. a model can be used for the quantitative exploration of future projections or alternative scenarios, in the evaluation of previously implemented policies, or as tool for discussions in thinking about the studied system.

Clear role descriptions can enhance model credibility, salience, and legitimacy, and thus may improve the balance between these quality criteria. The role that a model plays affects the requirements with regard to various modelling aspects, such as what is included in the model and what is not, model input handling, model output, calculation speed, accuracy, and acceptable uncertainty bounds. Model credibility is increased by a clear description of the model, its assumptions and simplifications, and its limitations in use (C2), a description of model input and output (C3), and the use of a quality assurance framework (C5). Model salience is increased by checking the model description to the (agreed upon) description of the intended role in the assessment (S4), and monitoring if no changes occur in this role during the assessment (S10). The instant re-use of existing models should be avoided (S9), because it is important to observe model limitations (S5) to answer the question if the model can indeed fulfil its intended role or not. Model legitimacy can be increased by documenting the expectations of participants and users towards model input, scenarios, concept, and output, as well as their justification (L2, L4). Ideally all stakeholders should agree to some degree with the use of the model in the assessment to avoid a loss of legitimacy (L5). The maintenance of an active dialogue with stakeholders on the role of the model positively affects model legitimacy (L3). An active dialogue reduces the risk of a loss of model salience because (changes in) perceptions on model quality criteria were missed.

A modeler should ensure that all documentation on the role of the model is available and agreed upon (or objections are noted), and that he/she understands what is expected of the model. The documentation may reveal issues surrounding expectations of what the model should do. If such issues become apparent they should be communicated to all participants and users. It should be

avoided that the model is delivered as reductionist science without consulting the stakeholders (L8).

5.4. Describe the model scenarios including key assumptions, and how they have been tested against logical structure and socio-political reality. Also describe to what extent stakeholders are involved in the creation and discussion of scenarios and input, and how

Scenarios are crucial to many environmental assessments and are often used to drive models. Scenarios are a way of making knowledge and experience broadly available (e.g. by drawing up a map and by listing the key assumptions) and can function as communication tool (Volkery et al., 2008). Input can consist of drivers in scenarios (e.g. atmospheric carbon dioxide levels, food prices, etc.), but also of additional input that is needed to run the model (e.g. soil and land-use maps, precipitation, etc.). Scenarios can include policy objectives or different transitions towards a policy objective.

The quality of the model performance is limited by the quality of what goes into the model (the well-known 'garbage in, garbage out' principle). Model credibility is improved by checking input and scenarios against well-established scientific knowledge (C3) and by subjecting these to uncertainty analysis (C6). Issues may arise though regarding the balance in the model quality criteria. Scenarios may be scientifically sound, yet inappropriate because they do not match views held by participants and users (compromising legitimacy) or are not relevant given the context (compromising salience). Girod et al. (2009) for instance analyzed the evolution of credibility, salience, and legitimacy during the development of the different IPCC emission scenarios, and they observed an increase in credibility but a decline in salience due to the existence of inappropriate classifications, and of a relatively high number of baseline scenarios. The salience of input and scenarios can be improved by matching these to the context (S3). These should also be comprehensible for stakeholders (S7) and in line with stakeholder expectations of the scenarios and what the model can do in general (S8). The legitimacy of scenarios is increased by covering the various views of stakeholders (L4), as was for instance done in the latest edition of the NO. A proper approach to stakeholder involvement should be followed to obtain scenarios and input that are credible, salient, and legitimate (Börjeson et al., 2006; Volkery et al., 2008; Girod et al., 2009; Mahmoud et al., 2009). A modeler can and should discuss with participants and users on how scenarios are developed and what input is delivered by whom. Arrangements should be made such that all scenario assumptions and data sources are tested both against logical structure and socio-political reality, and all should be documented.

5.5. Describe the model and its (relevant) output. Has the model including its output (including uncertainty estimates) been discussed with participants and users? what were the main concerns, if any?

Most modelers are well aware of technical issues regarding model credibility from a scientific point of view. The model should be discussed and be developed according to well-established scientific knowledge (C1). There should be a clear description of the model concept (C2), and the model should be subjected to testing, verification, validation (C5) and uncertainty analysis (C6). A lack of model performance or of reporting validation effort negatively affects model credibility (C8).

Fulfilling the technical demands towards model credibility in principle should present a no-regret good practice. However, spending resources on technical model credibility may invoke

trade-offs between model quality criteria. It is not necessarily implied that stakeholders hold the same view on model credibility as modelers, and credibility may counteract or be out of balance in comparison to salience and legitimacy. Even though the model may be taken for granted by certain participants and users (analogous to that many of us accept that a car drives without knowing how it exactly works) model complexity should be monitored. Model complexity that has been judiciously increased may positively affect model credibility (C7), but over-shot complexity generally leads to a loss of confidence (C9) and a loss of procedural legitimacy (L9). Moreover, there is the risk of loss of salience and legitimacy because there may be an increase in the level of technical expertise that is required to run the more complex model (S12), which may consume considerable resources that are then not spent on ful-filling demands regarding model salience and legitimacy, and may exclude participants and users. Model salience is increased by discussing not only the technical but also the application limita-tions of the model (S5), including monetary and legal ones. Model legitimacy is improved by developing the model concept in collaboration with stakeholders (L5) and meeting expectations of stakeholders (S8).

Model output is often the actual focal point in the knowledge transfer between modelers, participants and users. It is therefore crucial that model output is presented in a format that is comprehensible to all participants and users (L6). This increases acceptance and a feeling of ownership among them (L14). A valid option is, for instance, the inclusion of adjustable model parameters – dubbed 'policy levers' – that reflect variables that are important to policy actors (White et al., 2010). Results of scientific research that are not made available in a format that is comprehensible to policy actors – i.e. an absence of 'policy levers' – cause a major limitation of knowledge uptake by these actors (Liu et al., 2008). Model limitations should be properly understood to avoid misuse and misinterpretation of model output (Alexandrov et al., 2011). A proper scientific treatment and communication of uncertainty bounds is often vital to improve the credibility, salience, and legitimacy (Refsgaard et al., 2007; Liu et al., 2008), especially when trade-offs (e.g. between costs and risks) are involved and decisions have to be made in the face of uncertainty (S6). A lack of commu-nicating output uncertainty is a major threat to model salience (S11) and legitimacy (L14). This requires that all relevant model assumptions are documented and discussed with participants and users (Kloprogge et al., 2011). A modeler is responsible for this communication and should provide up-to-date and clear docu-mentation of the model concept, the main assumptions and sim-plifications, and the justification of the choices made about model concept and output.

5.6. Analyze the project process and planning and identify project limitations. Are sufficient resources available for model development, adaptation, and application? are sufficient resources available for stakeholder involvement? is there sufficient room for iterative steps and feedback steps? are there any other limitations that might negatively affect credibility, salience, and legitimacy?

There may be various factors over the course of a project in which a model-based outlook or assessment is done that may negatively affect the three model quality criteria. Not all of them will be 'solvable' for the modeler (defined earlier in Section 4). This does not mean that such factors are not of interest to him/her and/ or he/she should not be involved in any communication about these factors to others. Several of these factors have already been mentioned. Limitations in project resources may in particular be a factor that invokes strong trade-offs in credibility, salience, and legitimacy (Cash et al., 2002).

Model development and use cannot be aimed solely at model credibility, and appropriate stakeholders have to be included to discuss their views on model credibility, the intended role of the model in the assessment (affecting salience), and to safeguard the inclusion of stakeholder expectations and views (affecting legiti-macy). Therefore resources have to be divided between model development, adaptation and use, and proper stakeholder involvement and a communication strategy. This includes the se-lection of stakeholders (i.e. as participants and users), providing information to participants and users, and handling feedback. Refraining to include end users in early phases of the project may negatively affect salience (S13) and lead to a loss of feelings of ownership (L14), because expectations of these end users regarding model quality may differ from the views by those involved in earlier steps of the model development process. This in turn may present the modeler with a situation in which the model quality criteria are not well-balanced and possibly even counter-acting, while having insufficient resources to remedy this situation. Several rounds of interactions may be needed to decide on the frame of the issue, and to set the demands for the required scenarios and input, model design, model output, and output formats, and to discuss and learn from the outcomes. Sarkki et al. (2015) stress the importance of providing sufficient room (and thus also resources) for iterative cycles, in particular given the dynamic nature of the perception of credibility, salience and legitimacy by the various actors. An active and continuous dialogue with stakeholders improves legitimacy (L3), while the lack of support for it negatively affects salience and legitimacy (S13, L10, L14, L15). The dialogue should be aimed at keeping a proper balance in stakeholder expectations and views to avoid bias in stakeholder involvement (L11, L13). It should also prevent a loss of legitimacy based on expertise (L12), a continued existence of pre-conceived and immovable conceptions among stakeholders (if any; C10, L13), and the possible dissolution of the project team (L15).

6. Discussion

The objective of this paper is to present a practical checklist for modelers who are involved in environmental assessments to identify and when possible address issues that arise during the assessment regarding the three model quality criteria credibility, salience and legitimacy. Assessments generally make use of models in various capacities. As stakeholders are becoming increasingly involved in environmental policy (Voinov et al., 2016) these models need to meet various expectations set by stakeholders. Four major points that should be considered by modelers are 1/ the three model quality criteria do not necessarily have to be equally satisfied but should be balanced in the context in which models are pro-duced and used; 2/ the three model quality criteria may be coun-teracting, and trade-offs can be invoked between them, especially under for instance resource limitations; 3/ modelers, users and participants may hold different views regarding model credibility; and 4/ perceptions of the three quality criteria may be dynamic and path-dependent. These four points have also been found to appear in the NO case study presented in this paper. In addition, various factors have been identified from the literature and case study that may positively or negatively affect any or all of the model quality criteria. In cases where the model performance towards some or all quality criteria is lacking, this can have strong negative effects, not only for the model itself but also for an assessment as a whole. A possible result may be that a 'gap' appears in the transfer of knowledge from scientists to policy actors and other stakeholders (Meinke et al., 2006; Borowski and Hare, 2007; McIntosh et al., 2011; Bauler, 2012; Lemos et al., 2012), reducing the effectiveness of the assessment. Alignment between the quality perceptions of

participants and users may be important in preventing such a 'gap', and may depend on factors like the understanding and interpretation of the results by all involved in the assessment, but also on a shared ownership of the model in relation to decision making (Kelly et al., 2013; Voinov and Bousquet, 2010). There seems to be ample justification for providing guidance to modelers on how to deal with issues regarding model quality criteria in the development of models for use in environmental policy assessments. Jahn and Keil (2015) provide guidelines and profiles for quality standards valued by specific groups of actors in trans-disciplinary research, namely program managers or donors, researchers, and policy makers. Their guideline for researchers is valuable to map the environment of models in assessments and certainly is of use for modelers who are involved in assessments. Nevertheless, none of the questions on their guideline is specifically aimed at models or model use. The checklist presented here is intended as an aid to modelers in identifying issues regarding model quality criteria that specifically deal with model development and use.

The practical application of the checklist is a point for discussion. The first aspect of application is *when* to apply the checklist. We argue that modelers should go through the checklist as soon as they are involved in an assessment. Early efforts can avoid problems later on. We also explicitly point to the dynamic nature of the model quality criteria (Kunseler et al., 2015). As for instance is demonstrated in the NO case study, changes in stakeholder expectations and if/how the model meets these expectations may occur during a project (Sarkki et al., 2015). The checklist should therefore also be re-applied at regular intervals during the project so that actions can be undertaken by the modeler if possible.

The second aspect is *how* to apply the checklist. We advise that modelers not to do this alone, but to involve others, in particular stakeholders from the project as well as other team members, in dealing with the efforts related to the checklist issues. Stakeholder involvement in environmental modelling is increasingly discussed in the literature (Voinov et al., 2016), and it stands to reason this involvement includes thinking about issues that affect model credibility, salience and legitimacy. It requires a proper way of stakeholder involvement, and this should be conceptualized as continued two-sided communication during the whole process (Allen et al., 2013).

The primary goal of the checklist, as we see it, is to create awareness among modelers and provide a code of good practice. For this purpose the items on the checklist target the four main points that we identified in this paper. The goal of the model and the proper balance between model quality criteria cannot be assessed without understanding the broader context of the model application. Hence the first item on the list is to describe the goal of the assessment. The third item then asks for the role of the model in the assessment. If the model is found to be unsuitable for the intended application it may need to be adapted. If the role of a model is unclear or it is not agreed upon this may result in the loss of model credibility, salience and/or legitimacy. The three quality criteria do not necessarily have to be equally satisfied but should be balanced in view of the intended model application. The fourth item focuses on scenarios and input. Assumptions behind scenarios may be contested by stakeholders, and uncertainties may exist that are not acceptable to some or all involved. Again, a balance between quality criteria has to be achieved. The fifth item discusses the model itself and the model output. Views on the credibility of the model may differ and should be clarified. There may be discussion about the usefulness of the model output for decision-making, e.g. because of uncertainties, and model salience and legitimacy may certainly improve when model output is presented in a way that is understandable to stakeholders.

The establishment of communication with stakeholders is a necessary condition, and should be done in a proper way, which is why it is a separate item on the checklist. It will help in establishing a common ontology, the views of others on the model in terms of quality criteria (including what entails model credibility), in thinking about ways to present the model and its output in a way that is understandable for stakeholders, and so forth. The exclusion of relevant stakeholders in some or all of the model development phases increases the risk of changes or mismatches in expectations towards model application and performance, and thus increases the risk of improperly balanced or counter-acting model quality criteria. It also increases the risk of wasting resources for model development because many adaptations have to be made to the model later on.

The three quality criteria seem to be dynamic in nature (Sarkki et al., 2015), and proper monitoring of stakeholder views will increase the probability that changes are observed that occur during the process. We argue that the modeler ensures that there is an active and frequent monitoring of stakeholder satisfaction, understanding and involvement with regard to the model and the modelling project (as part of the assessment). Allen et al. (2013) for instance installed a communications working group with the core task of analyzing the scientist and stakeholder perceptions, understandings, and level of satisfaction during the project. They did this through interviews and surveys that used self-reporting questionnaires. Project communication was facilitated by regular working group meetings, monthly integration meetings, bi-annual meetings of all involved scientific personnel, and an e-mail service to update involved people on project progress. Through such a process of stakeholder interaction room is available to act upon changes, e.g. if stakeholders are not content what can and should be reasonably done to remedy this? Opacity can be achieved by documenting the monitoring results and the actions that were undertaken.

The checklist does not provide if-then-else options on actions to be undertaken by the modeler. If factors are solvable it means that action can and perhaps should be carried out by the modeler. However, resources may be limited or other limitations may exist that prevent action by the modeler; in particular financial and time constraints can invoke strong trade-offs between model quality criteria (Cash et al., 2002). This is for instance clearly demonstrated in the NO4: time limitations prevented proper validation of newly developed models, which were more salient and legitimate than the established models but at the possible expense of being less credible. There may furthermore be factors that can affect the model quality criteria and play a role in the information transfer 'gap' that cannot be expected to be addressed by modelers alone. For example, Lemos et al. (2012) mention institutional and organizational factors (mentioned earlier) that affect the legitimacy of models (and thus the effectiveness) but that cannot be realistically dealt with by researchers.

Because of the above the last item on the checklist is aimed specifically at the process of model development and application by asking for possible issues that may arise from planning. The best practice for modelers is to try and identify factors in the model environment and make them explicit to the other participants and users, without assuming the responsibility to solve them, in particular when limitations in resources present modelers with trade-offs between model quality criteria. An explicit exposure of limitations and trade-offs may help in making choices what aspects of model quality should be targeted under the existing constraints. Also for this it is necessary to know the views of involved stakeholders on model quality criteria, priorities therein, and whether or not these views are subject to change. We advise the modeler to communicate detected limitations and trade-offs explicitly to the

other participants and users, and to make them jointly responsible for any decisions regarding them. It is up to the modeler to maintain a proper and up-to-date documentation and management system to aid communication and reproducibility, and to keep documentation of (changes in) model input, output, assumptions and simplifications, their justification, testing, verification, calibration, validation, and uncertainty analysis, as well as (changes in) decisions made by participants and users on all relevant issues (such as ontology, model goal, use of output, stakeholder views, etc.) and the justifications of these (why were these decisions made, and who made them?).

7. Concluding remarks

The checklist presented in this paper is intended as a practical tool to make modelers involved in environmental assessments aware of the broad scope of options, limitations and trade-offs that exist with regard to meeting model quality criteria in science-policy settings. It involves some good practice that largely consists of following no-regret actions to be taken by modelers. In particular, it is aimed at creating awareness among modelers regarding four relevant points, namely 1/ the three quality criteria do not necessarily have to be equally satisfied but should be balanced in the context in which models are developed and applied; 2/ the three quality criteria can work counteracting and present the modeler with trade-offs, especially under restrictions such as resource limitations; 3/ there is variability between modelers, users and participants in their perception of not only salience and legitimacy, but also model credibility; and 4/ perceptions of the three quality criteria seem to be dynamic in nature and may even be path-dependent. The checklist is intended to help modelers identify factors that affect model quality that should be made explicit without necessarily having the possibility to solve them. It should not be considered as static, and future experience may warrant an update of the checklist. We expect that by following this checklist models can be improved in terms of credibility, salience, and legitimacy, and hence help in meeting an important necessary requirement to improve the uptake of model-based output by policy makers. This may in turn positively add to the effectiveness of environmental policy assessments.

Acknowledgements

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.envsoft.2016.06.003.

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. Environ. Modell. Softw. 26, 328–336.

Allen, E., Kruger, C., Leung, F.-Y., Stephens, J.C., 2013. Diverse perceptions of stakeholder engagement within an environmental modeling research team. J. Environ. Stud. Sci. 3, 343–356.

Angelstam, P., Andersson, K., Annerstedt, M., Axelsson, R., Elbakidze, M., Garrido, P., Grahn, P., Ingemar Jönsson, K., Pedersen, S., Schlyter, P., Skärbäck, E., Smith, M., Stjernquist, I., 2013. Solving problems in social-ecological systems: definition, practice and barriers of transdisciplinary research. Ambio 42, 254–265.

Bastin, L., Cornford, D., Jones, R., Heuvelink, G.B.M., Pebesma, E., Stasch, C., Nativi, S., Mazzetti, P., Williams, M., 2013. Managing uncertainty in integrated environmental modelling: the UncertWeb framework. Environ. Modell. Softw. 39, 116–134.

Bauler, T., 2012. An analytical framework to discuss the usability of (environmental) indicators for policy. Ecol. Indic. 17, 38–45.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Modell. Softw. 40, 1–20.

Bergmann, M., Jahn, T., Knobloch, T., Krohn, W., Pohl, C., Schramm, E., 2012. Methods for Transdisciplinary Research — a Primer for Practice. Campus Verlag GmbH, Frankfurt-am-Main.

Börjeson, L., Höjer, M., Dreborg, K.-H., Ekvall, T., Finnveden, G., 2006. Scenario types and techniques: towards a user's guide. Futures 38, 723–739.

Borowski, I., Hare, M., 2007. Exploring the gap between water managers and researchers: difficulties of model-based tools to support practical water management. Water Resour. Manage. 21, 1049–1074.

Cash, D., Clark, W.C., Alcock, F., Dickson, N., Eckley, N., Jäger, J., 2002. Salience, Credibility, Legitimacy and Boundaries: Linking Research, Assessment and Decision Making. Research Programs, John F. Kennedy School of Government, Harvard University.

Chwif, L., Barretto, M.R., Paul, R.J., 2000. On simulation model complexity. In: Joines, J.A., Barton, R.R., Fishwick, P.A. (Eds.), Proceedings of the 2000 Winter Simulation Conference, pp. 449–455.

Creswell, J.W., 2003. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Sage Publications, Thousand Oaks/London/New Delhi.

Flyvbjerg, B., 2006. Social science that matters. Foresight Eur. 38–42.

Ginger, C., 2014. Integrating knowledge, interests and values through modelling in participatory processes: dimensions of legitimacy. J. Environ. Plan. Man. 57, 643–659.

Girod, B., Wiek, A., Mieg, H., Hulme, M., 2009. The evolution of the IPCC's emissions scenarios. Environ. Sci. Policy 12, 103–118.

Hoppe, R., 2002. Cultures of public policy problems. J. Comp. Policy Anal. Res. Pract. 4, 305–326.

Jahn, T., 2008. Transdisciplinarity in the practice of research. In: Matthias Bergmann/Engelbert Schramm (Hg.): Transdisziplinäre Forschung. Integrative Forschungsprozesse verstehen und bewerten. Campus Verlag, Frankfurt/New York, pp. 21–37.

Jahn, T., Keil, F., 2015. An actor-specific guideline for quality assurance in transdisciplinary research. Futures 65, 195–208.

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

Janssen, P.H.M., Petersen, A.C., Van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R., 2003. RIVM/MNP guidance for uncertainty assessment and communication. Available via. http://www.nusap.net/sections.php?op=viewarticle&artid=17.

Kampen, J.K., Tamás, P., 2013. Should I take this seriously? A simple checklist for calling bullshit on policy supporting research. Qual. Quant. 48, 1213–1223.

Kelly (Letcher), R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S.H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., Van Delden, H., Voinov, A.A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environ. Modell. Softw. 47, 159–181.

Kloprogge, P., Van Der Sluijs, J.P., Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. Environ. Modell. Softw. 26, 289–301.

Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K., 2012. The role of expert opinion in environmental modelling. Environ. Modell. Softw. 36, 4–18.

Kunseler, E.-M., Tuinstra, W., Vasileiadou, E., Petersen, A.C., 2015. The reflective futures practitioner: balancing salience, credibility and legitimacy in generating foresight knowledge with stakeholders. Futures 66, 1–12.

Lang, D.J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., Swilling, M., Thomas, C.J., 2012. Transdisciplinary research in sustainability science: practice, principles, and challenges. Sustain. Sci. 7, 25–43.

Lemos, M.C., Kirchhoff, C.J., Ramprasad, V., 2012. Narrowing the climate information usability gap. Nat. Clim. Change 2, 789–794.

Lincoln, Y.S., Guba, E.G., 1985. Naturalistic Inquiry. Sage Publications, Newbury Park, CA.

Liu, Y., Gupta, H., Springer, E., Wagener, T., 2008. Linking science with environmental decision making: experiences from an integrated modeling approach to supporting sustainable water resources management. Environ. Modell. Softw. 23, 846–858.

Lusiana, B., Van Noordwijk, M., Suyamto, D., Mulia, R., Joshi, L., Cadisch, G., 2011. Users' perspectives on validity of a simulation model for natural resource management. Intern. J. Agric. Sustain. 9, 364–378.

Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R.,

Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., Van Delden, H., Waldick, R., White, D., Winter, L., 2009. A formal framework for scenario development in support of environmental decision-making. Environ. Modell. Softw. 24, 798—808.

McIntosh, B.S., Ascough II, J.C., Twery, M., Chew, J., Elmahdi, A., Haase, D., Harou, J.J., Hepting, D., Cuddy, S., Jakeman, A.J., Chen, S., Kassahun, A., Lautenbach, S., Matthews, K., Merritt, W., Quinn, N.W.T., Rodriguez-Roda, I., Sieber, S., Stavenga, M., Sulis, A., Ticehurst, J., Volk, M., Wrobel, M., Van Delden, H., El-Sawah, S., Rizzoli, A., Voinov, A., 2011. Environmental decision support systems (EDSS) development — challenges and best practices. Environ. Modell. Softw. 26, 1389—1402.

McNie, E., 2007. Reconciling the supply of scientific information with user demands: an analysis of problem and review of the literature. Environ. Sci. Policy 10, 17—38.

Meinke, H., Nelson, R., Kokic, P., Stone, R., Selvaraju, R., Baethgen, W., 2006. Actionable climate knowledge: from analysis to synthesis. Clim. Res. 33, 101—110.

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

Podestá, G.P., Natenzon, C.E., Hidalgo, C., Ruiz Toranzo, F., 2013. Interdisciplinary production of knowledge with participation of stakeholders: a case study of a collaborative project on climate variability, human decisions and agricultural ecosystems in the Argentine Pampas. Environ. Sci. Policy 26, 40—48.

Refsgaard, J.C., Henriksen, H.J., 2004. Modelling guidelines — terminology and guiding principles. Adv. Water Resour. 27, 71—82.

Refsgaard, J.C., Van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process — a framework and guidance. Environ. Modell. Softw. 22, 1543—1556.

Rykiel, E.J., 1994. Testing ecological models: the meaning of validation. Ecol. Model. 90, 229—244.

Sarkki, S., Tinch, R., Niemelä, J., Heink, U., Waylen, K., Timaeus, J., Young, J., Watt, A., Nesshöver, C., Van den Hove, S., 2015. Adding 'iterativity' to the credibility, relevance, legitimacy: a novel scheme to highlight dynamic aspects of science—policy interfaces. Environ. Sci. Policy 54, 505—512. http://dx.doi.org/10.1016/j.envsci.2015.02.016.

Schmolke, A., Thorbek, P., DeAngelis, D.L., Grimm, V., 2010. Ecological models supporting environmental decision making: a strategy for the future. Trends Ecol. Evol. 25, 479—486.

Scholar Google, https://scholar.google.com/.

Schut, M., Van Paassen, A., Leeuwis, C., 2013. Beyond the research—policy interface.

Boundary arrangements at research-stakeholder interfaces in the policy debate on biofuel sustainability in Mozambique. Environ. Sci. Policy 27, 91—102.

Sterk, B., Van Ittersum, M.K., Leeuwis, C., 2011. How, when, and for what reasons does land use modelling contribute to societal problem solving? Environ. Modell. Softw. 26, 310—316.

STOWA/RIZA, 1999. Smooth Modelling in Water Management, Good Modelling Practice Handbook. STOWA report 99—05, Dutch Dept. of Public Works, Institute for Water Management and Waste Water Treatment report 99.036, ISBN 90-5773-056-1. Available from: http://harmoniqua.wur.nl/public/Reports/Existing%20Guidelines/GMP111.pdf.

Turnhout, E., Waterton, C., Neves, K., Buizer, M., 2013. Rethinking biodiversity: from goods and services to "living with". Conserv. Lett. 6, 154—161.

Vader, J., Smits, M.J.W., Vreke, J., Dagevos, J.C., 2004. Usefulness and Necessity of Nature Outlooks (Nut en noodzaak van Natuurverkenningen). In Dutch:. Wageningen, Natuurplanbureau — vestiging Wageningen, Planbureaurapporten 16.

Vader, J., Bogaardt, M.J., 2014. Nature Outlook 2 Years Further. Use of the Nature Outlook 2010-2040 (Natuurverkenning 2 jaar later. Over gebruik en doorwerking van Natuurverkenning 2010-2040). In Dutch: (WOt internal document, Wettelijke Onderzoekstaken Natuur en Milieu, Wageningen UR, Wageningen).

Voinov, A., Bousquet, F., 2010. Modelling with stakeholders. Environ. Modell. Softw. 25, 1268—1281.

Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P., 2016. Modelling with stakeholders — next generation. Environ. Modell. Softw. 77, 196—220.

Volkery, A., Ribeiro, T., Henrichs, T., Hoogeveen, Y., 2008. Your vision or my model? Lessons from participatory land use scenario development on a European scale. Syst. Pract. Action. Res. 21, 459—477.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5 (1), 13—26.

Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Krayer Von Kraus, M.P., 2003. Defining uncertainty — a conceptual basis for uncertainty management in model-based decision support. Integr. Assess. 4, 5—17.

Warleigh, A., 2003. Democracy in the European Union. Sage Publications.

White, D.D., Wutich, A., Larson, K.L., Gober, P., Lant, T., Senneville, C., 2010. Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theatre. Sci. Public Policy 37, 219—232.