ARTICLE IN PRESS

Environmental Modelling & Software xxx (2009) 1–13

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

A method for the analysis of assumptions in model-based environmental assessments

Penny Kloprogge [a], Jeroen P. van der Sluijs [a,b,*], Arthur C. Petersen [c]

[a] Copernicus Institute for Sustainable Development and Innovation, Utrecht University, Utrecht, The Netherlands
[b] Centre of Economics and Ethics for the Environment and Development (C3ED), University of Versailles Saint-Quentin-en-Yvelines, Guyancourt, France
[c] Netherlands Environmental Assessment Agency (PBL), Bilthoven, The Netherlands

ARTICLE INFO

Article history:
Received 30 July 2008
Received in revised form
3 May 2009
Accepted 18 June 2009
Available online xxx

Keywords:
Uncertainty
Value-ladenness
Model assumptions
Assessments
NUSAP
Pedigree matrix

ABSTRACT

When analysts do model-based scientific assessments of complex environmental problems, they have to make many assumptions. This inevitably involves – to some degree – subjective judgements by the analysts. Although the potential value-ladenness of model-based assessments has been extensively problematized in literature, this has not so far led to a systematic strategy for analyzing this value-ladenness. In this article, a new method is presented to identify, review, and prioritize assumptions in order to assess the potential value-ladenness of important assumptions and to deal with these potentially value-laden assumptions in an explicit and transparent manner. The potential value-ladenness of the assumptions is analyzed using a so-called pedigree matrix. The matrix addresses epistemic (general and discipline-bound) and non-epistemic (socio-political and practical) values. The method can be applied by the analysts doing the assessment in collaboration with peers and stakeholders or by external reviewers. Here, the method is illustrated for the modelling chain that was used to calculate the indicator 'death and emergency hospital admittances due to the exposure to ozone' in the Fifth Dutch Environmental Outlook. The weakest links of the calculation chain were identified through a workshop. This method for the analysis of assumptions enables the analysts to make conscious, well-underpinned, transparent choices, and pinpoints the issues in the chain that are important to communicate to the audience of the assessment report.

1. Introduction

While making model-based assessments of complex environmental problems, analysts try to represent the causal chain of the system at hand, capture its dynamics and often to explore possible future developments in the system. The set of model calculations and other data operations that together produce the end results of the assessments, is here referred to as the 'calculation chain' behind a given quantified outcome of interest of an assessment.

Throughout the calculation chain the analyst will use the knowledge that is available to him/her at the time of the assessment. Not all of that knowledge has the status of well-established knowledge. Wherever uncertainties and knowledge gaps occur, the analyst will have to revert to assumptions. Assumptions also are frequently applied to simplify parts of the calculations. Assumptions are also frequently needed to connect submodels when they have different spatial, temporal or system scales that need to be bridged somehow. Assumptions can be made explicitly or implicitly. Often, an assumption explicitly made by the analyst, automatically implies additional, implicit assumptions.

In the following we discuss three reasons to deal more explicitly with assumptions: assumptions can lead to biased assessments (value-ladenness), assumptions can limit the quality of model-based environmental assessments, and dealing more explicitly with assumptions can improve current uncertainty assessment practice.

Since assumptions by definition cannot objectively be determined (since something is *assumed*), there always is an element of subjectiveness in assumptions. Two analysts assessing the same issue will not necessarily make the exact same assumptions in the calculation chain. Consequently, an assessment is not made up of objective, value-free scientific facts alone. For this reason, assessments can be considered to be value-laden to a certain degree.

Numerous studies from the history and sociology of science have problematized the classic distinction between facts and values. Scientific facts and knowledge claims, especially when

* Corresponding author. Copernicus Institute for Sustainable Development and Innovation, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. Tel.: +31 30 253 7631; fax: +31 30 253 7601.
E-mail address: j.p.vandersluijs@uu.nl (J.P. van der Sluijs).

1364-8152/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.envsoft.2009.06.009

produced at the science-policy interface, have been shown to be at least partially socially constructed and co-shaped by implicit or explicit negotiation processes. Observation has been shown to be theory-laden and cognitive authority of science is ultimately produced by boundary work and negotiation. These contexts of knowledge production and use produce value-ladenness in knowledge claims (Jasanoff, 1990; van der Sluijs et al., 1998; Jasanoff and Wynne, 1998; Huesemann, 2002).

The element of subjectiveness in the scientific domain is also problematized with regard to computer models, which have increasingly been used in assessments since the 1980s (see, for example, Hornberger and Spear, 1981; Keepin and Wynne, 1984; Schneider, 1997; Oreskes et al., 1994; van der Sluijs, 2002; Jakeman et al., 2006). These (and other) authors stress the importance of transparency about the value-laden assumptions in assessments. Keepin and Wynne (1984) argue for rigorous peer review and for testing the robustness and sensitivity of results. Jakeman et al. (2006) stress the importance of justifying the choice of model families and features, including assumptions. The need for sensitivity analysis is also stressed by Saltelli et al. (2000) and by Stirling (1999, 2001). Stirling found that the final results of risk assessment studies depend significantly on changes in starting assumptions. He also argues for the use of alternative framing assumptions, through which risk assessments result in a range of values, rather than discrete scalar numbers. This 'diversification' of assumptions is also advocated by Schneider (1997) who proposes that Integrated Assessment modellers provide users with a large range of value-containing options via menu-driven designs. Funtowicz and Ravetz (1993) stress the need for extended peer review, in which stakeholders and citizens are involved in the review process of science for policy in those cases where facts are uncertain, values in dispute, stakes high, and decisions urgent. In summary, it can be concluded that transparency, diversification of assumptions, extended peer review, and insight into the influence of assumptions on the outcomes of the assessment are seen as important elements in a strategy for dealing with value-laden assumptions.

Especially when dealing with complex issues that are surrounded by uncertainties and lack of knowledge, many assumptions have to be made. A lot of effort has been put into conceptual research (e.g., van der Sluijs, 1997; van Asselt, 2000; Walker et al., 2003; Janssen et al., 2005; van der Sluijs, 2005; Petersen, 2006; Maxim and van der Sluijs, 2007), and in the development of strategies for dealing with uncertainties. Examples of strategies are the PRIMA approach in which Cultural Theory is applied (van Asselt, 2000), the NUSAP method (Funtowicz and Ravetz, 1990; van der Sluijs et al., 2005), and the Guidance for Uncertainty Assessment and Communication (van der Sluijs et al., 2003, 2004; Petersen et al., 2003; Janssen et al., 2003, 2005; van der Sluijs et al., 2008). A concise overview of the different strategies is given by Refsgaard et al. (2007).

Some of these methods partly intervene with the assumptions in an assessment. When, for instance, applying Cultural Theory when dealing with uncertainties in an issue, the assumptions are set in accordance with an ideal type of value orientations. The assumptions are coloured by the perspective of that particular ideal type.

Not only in uncertainty management, but also in quality management assumptions constitute one of the elements that are tackled. In good practice guidelines (e.g., Lindfords et al., 1995; Refsgaard et al., 2005; Jakeman et al., 2006) and checklists (e.g., Risbey et al., 2005) attention is paid to the assumptions that are made and to the communication with regard to these assumptions.

None of these uncertainty and quality management methods, however, focuses specifically and systematically on the potential value-ladenness of the assumptions made in an assessment. In current uncertainty and quality management strategies the value-ladenness of assumptions is merely analyzed and handled in a general manner. Thus, although the value-ladenness of assumptions has been extensively problematized in the literature, this has not so far led to a systematic strategy for dealing with this problem. The central question in this paper is how assumptions in environmental assessments can be systematically identified, reviewed, and prioritized, in order to assess the potential value-ladenness of important assumptions and to deal with these potentially value-laden assumptions in an explicit and transparent manner.

In this study, we zoom in on the value-ladenness of assumptions, starting from the viewpoint of the analyst carrying out the assessment. In an earlier study (Kloprogge and van der Sluijs, 2002) it was shown that choices made by an analyst are affected by a range of factors. The choices of analysts are influenced by their knowledge, perspectives and situational factors. Arbitrariness can also play a role, in situations where the analyst has no reason to prefer one particular assumption to another. Based on the nature of factors influencing the choice for a certain assumption, we distinguish four different types of value-ladenness of assumptions: value-ladenness in a general epistemic sense (e.g., assumptions are coloured by the approach that the analyst prefers), in a disciplinary-bound epistemic sense (e.g., assumptions are coloured by the discipline in which the analyst was educated), in a socio-political sense (e.g., assumptions may be coloured by political preferences of the analyst), and in a practical sense (e.g., the analyst is forced to make simplifying assumptions due to time constraints). At first sight it may look strange that we include constraints having practical reasons in our typology of value-ladenness, but assumptions that are justified by a practical constraint can still lead to biased assessments as there is a potential to exploit references to such constraints to introduce assumptions that favour a politically desired outcome of an assessment. Also in case that there is no intentionality, practical constraints can introduce assumptions that lead to assessment results unduly favouring one position in a discourse over an other.

In the following section, we present the method that we designed for dealing with the potential value-ladenness of assumptions in environmental modelling. Next, one of the core elements used in this method, a 'pedigree matrix' to assess the potential value-ladenness in assumptions is elaborated on. The method was tested *ex-post* on two calculation chains in the 5th Dutch National Environmental Outlook, published in 2000. One of these cases, deaths and emergency hospital admittances due to exposure to ozone, is presented in this article to demonstrate the first three steps of the method. For other applications of the method, the reader is referred to Kloprogge et al. (2005) and Craye et al. (2009). The final section presents the conclusions and reflects on the applicability of the new method.

2. A method for analyzing the value-ladenness of assumptions

The method we introduce here for dealing with assumptions in model-based assessments consists of three sections: an analysis section, in which the assumptions in an assessment are identified and analyzed; a revision section, in which the assessment is altered or extended based on the analysis results; and a communication section, in which it is determined what should be communicated with respect to the assumptions in the assessment, based on the analysis. The method contains seven steps:

Analysis

1. Identify explicit and implicit assumptions in the calculation chain.
2. Identify and prioritize key-assumptions in the chain.

3. Assess the potential value-ladenness of key-assumptions.
4. Identify 'weak' links in the calculation chain.
5. Further analyze the potential value-ladenness of key-assumptions.

Revision

6. Revise/extend assessment:
   – sensitivity analysis of key-assumptions;
   – diversification of assumptions;
   – different choices in chain.

Communication

7. Communication:
   – key-assumptions;
   – alternatives and underpinning of choices regarding assumptions made;
   – influence of key-assumptions on results;
   – implications in terms of robustness of results.

All steps will be elaborated on below.

2.1. Identify explicit and implicit assumptions in the calculation chain

First of all, when analyzing assumptions in the calculation chain of an assessment, explicit and implicit assumptions have to be identified. Preferably, analysts construct a list with assumptions while doing the assessment. To identify implicit assumptions, it is important to repeatedly consider whether an assumption made implies other assumptions.

Each analyst, however, has limited knowledge and perspectives with regard to the assessment topic, and in consequence will have some 'blind spots'. It is therefore important that other analysts (peers) are involved in identifying assumptions. This can, for instance, be organized in the form of a workshop. Preferably stakeholders are involved as well to bring in their specific views and knowledge. In this way an inventory of assumptions is made that includes the controversial issues, from a scientific as well as from a social point of view. Identifying assumptions in a setting with peers and stakeholders can be viewed as part of an extended peer review process.

The aggregation level of the assumptions on the assumptions list can vary. An assumption can refer to a specific detail in the chain ('the assumption that factor x remains constant'), as well as refer to a cluster of assumptions on a part of the chain ('assumptions regarding submodel x'). Both types can be included in the analysis in a fruitful way, provided that the persons involved in the analysis of the assumptions share the same interpretation of the assumptions as formulated in the list.

2.2. Identify and prioritize key-assumptions in the chain

Since the time for an analysis of potential value-ladenness is always limited and not all assumptions will be of considerable influence on the assessment as a whole, the second step aims to identify and select the most important assumptions in the chain. Only these key-assumptions will be involved in the steps to come; the other assumptions will then no longer be considered. Ideally, this selection should be based on a comprehensive sensitivity analysis. However, a sensitivity analysis involving all assumptions will often not be attainable. Formulating and quantifying alternative assumptions in many cases requires a lot of effort. In some cases, a different assumption will even require a new model to be built.

To arrive at a selection of the most important assumptions the analysts who are carrying out the assessment, other experts and preferably stakeholders can be asked to indicate the estimated influence of the assumptions on the results of the assessment. An expert elicitation technique can be used in which the experts bring forward their opinions and argumentation on whether an assumption is of high or low influence on the indicator results. Based on the discussion the participants then can indicate their personal estimate regarding the magnitude of the influence, informed by the group discussion. Next, a group ranking can be established, based on the individual scores of the participants (see the example later on in the article).

It seems preferable to not restrict the number of key-assumptions too much. When involving more assumptions in the ranking exercise, better insight is obtained in the relative importance of assumptions throughout the calculation chain.

2.3. Assess the potential value-ladenness of key-assumptions

For the purpose of assessing the potential value-ladenness of assumptions we developed a so-called 'pedigree matrix' (see Table 1), based on the pedigree matrix applied in the NUSAP methodology for uncertainty management (Funtowicz and Ravetz, 1990). It contains several criteria with which the potential value-ladenness of assumptions can be reviewed. The pedigree matrix is discussed in more detail in Section 3.

The pedigree of each key-assumption is scored by the analysts, peers, and stakeholders on each of the criteria. Here, again a group discussion takes place first, in order for the participants to remedy each others' blind spots and to exchange arguments. It is the facilitator's job to make sure that the discussion does not slide off to a quick group consensus, but that there is an open discussion promoting critical review.

The order in which the key-assumptions are handled is determined by the group ranking established in step 2 of the method, starting with the key-assumption with the highest rank. In this way, if time proves to be too short to go over all the key-assumptions, the most important ones are not left out of the analysis.

2.4. Identify 'weak' links in the calculation chain

The pedigree matrix (see Section 3) is designed such that, as a rule of thumb, assumptions that score low on the pedigree criteria have a high potential for value-ladenness. Assumptions that, besides a low score on the criteria, also have a high estimated influence on the results of the assessment can be viewed as problematic weak links in the calculation chain. A tool to identify these assumptions is a diagnostic diagram (van der Sluijs et al., 2002); see Fig. 1 for the layout of such as diagram. The diagnostic diagram plots each assumption according to the estimated influence of the assumptions on the assessment results (*x*-axis) and the average pedigree scores (averaged over the six pedigree criteria and averaged of the experts) of the assumptions (*y*-axis). Assumptions that are situated in the upper right corner are in the 'danger zone' (i.e., high potential value-ladenness and high influence on the assessment), the ones in the lower left corner are in the 'safe zone' (i.e., low potential value-ladenness and low influence on the assessment). The assumptions that lie most in or towards the danger zone can be viewed as the most problematic assumptions in the calculation chain.

Please cite this article in press as: Kloprogge, P., et al., A method for the analysis of assumptions in model-based environmental assessments, Environ. Model. Softw. (2009), doi:10.1016/j.envsoft.2009.06.009

Table 1
The pedigree matrix for the assessment of the potential value-ladenness of assumptions.

Type of value-ladenness	Practical	General epistemic	General epistemic	Disciplinary-bound epistemic	Socio-political	Socio-political	Influence on results
Criteria → / Score ↓	Influence of situational limitations	(Im)plausibility	Choice space	(Dis)agreement among peers	(Dis)agreement among stakeholders	Sensitivity to view and interests of the analyst	
2	Choice assumption hardly influenced	The assumption is plausible	Hardly any alternative assumptions available	Many would have made the same assumption	Many would have made the same assumption	Choice assumption hardly sensitive	The assumption has only local influence
1	Choice assumption moderately influenced	The assumption is acceptable	Limited choice from alternative assumptions	Several would have made the same assumption	Several would have made the same assumption	Choice assumption moderately sensitive	The assumption greatly determines the results of the step
0	Totally different assumption had there not been limitations	The assumptions is fictive or speculative	Ample choice from alternative assumptions	Few would have made the same assumption	Few would have made the same assumption	Choice assumption sensitive	The assumption greatly determines the results of the indicator

2.5. *Further analyze the potential value-ladenness of key-assumptions*

Now that there is a general insight in the relative importance of the key-assumptions in the calculation chain, the nature of the potential value-ladenness of the individual key-assumptions can be explored. Based on inspection of the pedigree scores, it can be analyzed: (i) what types of value-ladenness possibly play a role and to what extent; (ii) to what extent there is disagreement on the pedigree scores among the participants; and (iii) whether changing assumptions is feasible and desirable. Potential motives for reconsidering assumptions are that the assumption is fictive or speculative (score 0 on plausibility) or that few peers and/or stakeholders would have made the same assumption (score 0 on agreement among peers/stakeholders). If hardly any alternative assumptions are available (score 0 on choice space) the options for changing assumptions is limited.

2.6. *Revise/extend assessment*

Based on the analysis in step 5, it can be decided to change or broaden the assessment. As a minimum option, the assessment can be extended with a sensitivity analysis, which gives more information on the influence of weak links in the assessment. The robustness of the results of the indicator can then be analyzed in view of the critical key-assumptions.

Besides performing a sensitivity analysis, specific assumptions can be revised or diversified. In the case of revising an assumption, an alternative is chosen, that is, the assumption is replaced by a different assumption. In some cases however, it will be difficult to choose between alternative assumptions, since there might be differing views on the issue. If these assumptions have a high influence on the assessment as a whole, it can be decided to diversify the assumptions: the calculation chain is 'calculated' using several alternative assumptions in addition to the existing ones. In this way, additional assessments are formed, with differing outcomes, depending on what assumptions are chosen.

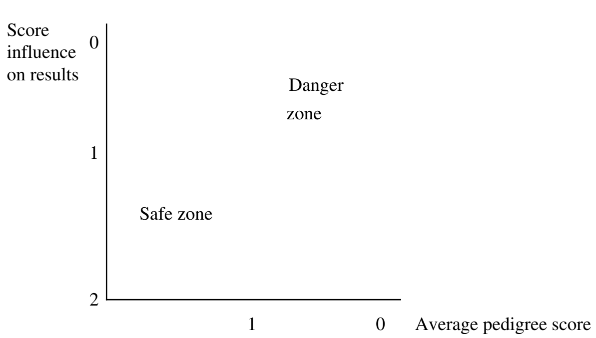

Fig. 1. Layout of a diagnostic diagram.

If several assumptions in the chain are involved, it may be possible to 'cluster' the assumptions in a consistent way, e.g., choose 'worst case' values for the assumptions or 'conservative' values for the assumptions. Based on the diversified assessment it may be possible to draw robust conclusions regarding the outcomes of interest of the assessment.

2.7. *Communication*

It is important to be explicit about potential value-ladenness in the chain and the effects of potentially value-laden assumptions on the outcomes of the assessment. Analogous to a patient information leaflet accompanying medicines, the presentation of the assessment results should be accompanied by information on:

- what are the key-assumptions in the calculation chain;
- what are the weak links in the chain;
- what were the alternatives and what is the underpinning of the choices that were made regarding assumptions;
- what is the robustness of the outcomes of interest in view of the key-assumptions.

The information and insights gained in steps 1–6 form the basis for this final step. The key-assumptions were identified and prioritized in step 2. The weak links in the chain were identified and analyzed in steps 4 and 5. Information on the alternatives and underpinning of the choices made regarding assumptions was gathered in the initial assessment and in step 6. For further guidance on uncertainty communication, the reader is referred to Wardekker et al. (2008).

3. The pedigree matrix

When tackling the value-ladenness of assumptions, it is virtually impossible to assess the value-ladenness of assumptions itself. This would require exact and detailed knowledge on what factors contributed to what extent to the analyst's choices. There is an

Please cite this article in press as: Kloprogge, P., et al., A method for the analysis of assumptions in model-based environmental assessments, Environ. Model. Softw. (2009), doi:10.1016/j.envsoft.2009.06.009

entangled web of factors influencing the choices made, part of which the analyst himself will be unaware of. However, the room for value-ladenness, the 'potential value-ladenness' can be addressed. This is done in step 3 of the method.

For this purpose we designed a 'pedigree matrix'. The idea of a pedigree matrix was introduced by Funtowicz and Ravetz (1990). It is part of the NUSAP system for uncertainty assessment (Funtowicz and Ravetz, 1990; van der Sluijs et al., 2005). NUSAP is an acronym conveying five qualifiers of scientific information: numeral, unit, spread, assessment, and pedigree. It extends the statistical approach to uncertainty with the methodological and epistemological dimensions by adding expert judgement of reliability (assessment) and systematic multi-criteria evaluation of the underpinning of numbers (pedigree). It combines quantitative and qualitative dimensions of uncertainty. On the one hand the sensitivity of results to spread in the numbers used in a calculation is examined, on the other hand the strength of these numbers is assessed, using pedigree. Pedigree addresses the strengths and weaknesses in the knowledge base behind a number by critically reviewing the production process of the number and the scientific status and underpinning of the number. A pedigree matrix is an aid for assessing the pedigree. It contains criteria that reflect the key components of the production process of policy relevant quantitative information (Funtowicz and Ravetz, 1990). These key components can vary with each special sort of information (e.g., specific key components for research information, emission monitoring data, environmental models; examples can be found on http://www.nusap.net). Many of these criteria are hard to measure in an objective way. Assessment of pedigree involves qualitative expert judgement. To minimize arbitrariness and subjectivity in measuring strength a pedigree matrix codes qualitative expert judgements for each criterion into a discrete numeral scale from 0 (weak) to 4 (strong) with linguistic descriptions (modes) of each level on the scale. Note that these linguistic descriptions are mainly meant to provide guidance in attributing scores to each of the criteria. It is not possible to capture all aspects that an expert may consider in scoring a pedigree in a single phrase. Therefore a pedigree matrix should be applied with some flexibility and creativity.

Here we have developed a pedigree matrix to review the potential value-ladenness of assumptions. This matrix is presented in Table 1. Our matrix deviates from Funtowicz and Ravetz original idea of pedigree matrices in two ways. First, their pedigree matrices used criteria addressing cognitive (e.g., empirical basis) and social (e.g., colleague consensus) phases in the production of knowledge. Based on interviews with modellers about their practices (Kloprogge and van der Sluijs, 2002) we extended this to include more phases that turned out to play a role in model building. Second, we used a three point scale (0–2) in stead of a five point scale. The resulting pedigree matrix contains the following criteria: influence of situational limitations, implausibility, choice space, (dis)agreement among peers, (dis)agreement among stakeholders, sensitivity to view and interests of the analyst, and influence on results.

Each of the criteria can be connected to one or two of the four types of value-ladenness we distinguished (value-ladenness in general epistemic sense, in a disciplinary-bound epistemic sense, in a socio-political sense, and in a practical sense). Three modes have been assigned to each of the criteria. Depending on the most suitable mode for a certain criterion, an assumption can score a 0, 1 or 2 on that criterion. The modes of all the criteria were arranged in such a way that the lower the score, the more room for value-ladenness an assumption contains.

The criteria, the scores and the column 'influence on results' are elaborated on below.

3.1. *Influence of situational limitations*

The choice for the assumption can be influenced by situational limitations, such as limited availability of data, money, time, software, tools, hardware and human resources. Without these restrictions, the analyst would have made a different assumption.

Although indirectly these limitations might be of a socio-political nature (e.g., the institute the analyst works for has other priorities and has a limited budget for the analyst's work), from the analyst's point of view these limitations are given. It can therefore be seen as primarily connected to value-ladenness in a practical sense.

3.2. *(Im)plausibility*

Although it is often not possible to assess whether the approximation created by the assumption is in accordance with reality, mostly an assessment can be made of the (im)plausibility of the assumption. If an analyst has to revert to fictive or speculative assumptions, because a plausible assumption is not attainable, the room for epistemic value-ladenness will often be larger. A fictive or speculative assumption also leaves room for potential disciplinary-bound epistemic and socio-political value-ladenness. This is, however, dealt with primarily in the criteria 'agreement among peers', and 'agreement among stakeholders' and 'sensitivity to view and interests of the analyst', respectively.

3.3. *Choice space*

In some cases an analyst has no choice but to make a certain assumption. In other cases several alternatives are available. The choice space indicates the degree to which alternatives were available to choose from when making the assumption In general, it can be said that a large choice space leaves more room for the epistemic preferences of the analyst. In other words: the potential for value-ladenness in a general epistemic sense will often be larger in case of a larger choice space. A large choice space will to some extent also leave more room for disciplinary-bound epistemic and socio-political value-ladenness. These are however primarily dealt with in the criteria 'agreement among peers', and 'agreement among stakeholders' and 'sensitivity to view and interests of the analyst', respectively.

3.4. *(Dis)agreement among peers*

An analyst makes the choice for a certain assumption based on his or her knowledge and perspectives regarding the issue. Other analysts having to make the same choice may choose a different assumption. The degree to which the choice of peers is likely to coincide with the analyst's choice is expressed in the criterion 'agreement among peers'.

These choices may be partly determined by the disciplinary training of the peers, and by their epistemic preferences. This criterion can thus be seen to be related to value-ladenness in a disciplinary-bound epistemic sense and in a general epistemic sense.

3.5. *(Dis)agreement among stakeholders*

Stakeholders, though mostly not actively involved in carrying out assessments, may also opt for a different assumption in case they were asked to make one. The degree to which the choice of stakeholders is likely to coincide with the analyst's choice is expressed in the criterion 'disagreement among stakeholders'. This will often have to do with the socio-political perspective of the stakeholders on the issue at hand and this criterion can therefore be seen as referring to value-ladenness in a socio-political sense.

3.6. Sensitivity to view and interests of the analyst

Some assumptions may be influenced, consciously or unconsciously, by the view and interests of the analyst making the assumption. The analyst's epistemic preferences, and his cultural, disciplinary and personal background may influence the assumption that is eventually chosen. The influence of the analyst's disciplinary background on the choices and the influence of his epistemic preferences are taken into account in the criteria 'agreement among peers', 'plausibility' and 'choice space'. In this criterion the focus is on the room for value-ladenness in a socio-political sense.

3.7. Influence on results

In order to be able to pinpoint important value-laden assumptions in a calculation chain it is not only important to assess the potential value-ladenness of the assumptions, but also to analyze the influence on outcomes of interest of the assessment (the 'spread' qualifier in the NUSAP system). Ideally, a sensitivity analysis is carried out to assess the influence of each of the assumptions on the results. In most cases, however, this will not be attainable: formulating and quantifying alternative assumptions in many cases requires a lot of effort. In some cases, a different assumption will even require a new model to be developed. The pedigree matrix therefore includes a criterion 'influence on results' which in the absence of sensitivity analysis is based on expert judgement.

The pedigree matrix is designed such that assumptions that score low on the pedigree criteria have a high potential for value-ladenness.

4. The case of the ozone-indicator in the fifth Dutch National Environmental Outlook

The Dutch National Environmental Outlook (EO) is an assessment of key environmental indicators outlining different future scenarios for a time period of several decades. It is prepared by the Netherlands Environmental Assessment Agency (PBL; the Dutch acronym was MNP until May 2008), which made part of the National Institute of Public Health and the Environment (RIVM) until 2006. The EO serves as input for the National Environmental Policy Plan that is issued by the Dutch government every four years.

Within the PBL for each EO a project team is set up with project team members coordinating parts of the assessment. After the issues to be included in the assessment are selected, several analysts across different departments of the PBL, together with institutions in its network, carry out (model) calculations to arrive at the results of the indicators on the selected issues. The contents of the assessment are partly gathered from previous (EO and other) assessments, and part of the assessment is carried out specifically for the EO. Model calculations play an important role in the assessments. In a 'model chain' of soft-linked computer models the effects for the environment for different scenarios are calculated. Beside model calculations, other calculations and operations take place. Many assumptions have to be made in the calculation chains, especially since the output of a computer model in the chain often does not fit the requirements of input for the next model or operation in the chain.

The complexity of these assessments and their role in the Dutch science-policy interface makes the EO's an interesting test case. Another reason for choosing EO's as test case is that they have been subject to an extensive uncertainty study (van Asselt, 2000; van Asselt et al., 2001). One of the recommendations of the report was to design a 'guidance' for uncertainty management. This recommendation has resulted in a 'Guidance for Uncertainty Assessment and Communication' (see Section 1).

We focused our analysis on the fifth EO, which, at the time of this study, was the most recent EO published. In the EO5, hundreds of indicators are presented that indicate the (future) pressure on or state of the Dutch, European or global environment. The indicators provide insight in potential developments regarding climate, nature and biodiversity, health and safety and the living environment in the time period 2000–2030.

In this case study, we aim to identify key-assumption in the model calculation chain for the indicator 'deaths and emergency hospital admittances due to exposure to ozone' and assess their potential value-ladenness. This is done by applying the first four steps of the method developed in Sections 2 and 3. A further aim of this case study is to test and evaluate the method.

4.1. Expert workshop

For our case study on the ozone-indicator a 4 h expert workshop was organized, in which the first four steps of the method were applied. The ozone group consisted of five participants. Participants were selected based on expertise on (part of) the calculation chain behind the indicators. Some of the participants had contributed to the EO5 assessments. In order to obtain a diverse group regarding expertise and regarding the involvement in the EO5 calculations, RIVM participants of other departments were invited as well as experts from other institutes.

Before the workshop took place the calculation chain was 'reconstructed': all calculations and operations that led to the eventual results of the indicator were identified. This was done by analysing the text in the EO5, background documents and by conducting interviews with the RIVM analysts involved. A description of the steps taken in the calculation chain can be found in Box 1.

4.2. Workshop procedure

Before the workshop all participants received the description of the calculation chain. During the workshop itself the facilitator gave a short elaboration on these descriptions and presented the list of assumptions that had been compiled beforehand. The participants were asked if they had any comments on these assumptions and were asked to complete the list with important assumptions they thought were missing.

Each participant received a set of 'scoring cards'. Each card contained one of the assumptions for which the pedigree scores could be filled in on the card (see Kloprogge et al., 2005, for an example of a scoring card).

They also received several blank cards on which they could fill in additional assumptions that were identified during the workshop. The participants were then asked to select seven cards containing the assumptions that, according to them, seemed most important in the calculation chain. They were asked to sort these seven cards (assumptions) from most important to least important. They filled in their ranking on a form containing a list of all assumptions. Everyone was asked to mark the most important assumption with a '1', the second most important with a '2', etc., till '7'. In order to obtain a group ranking, these scores were reversed (i.e., 1 becomes 7, 2 becomes 6, etc.). The scores of all the participants were then added per assumption. Next, the assumptions were ranked in order of diminishing total score, thus expressing the group ranking.

Next, the scoring cards were filled in, i.e., the assumptions were scored on the pedigree criteria. This was done card by card, starting with the card that received the highest priority. Each criterion of a certain assumption was discussed briefly in the group. After the

Box 1. Description of the calculation chain

Under the influence of sunlight several tropospheric reactions take place involving NO_x and volatile organic compounds (VOC). In this process ozone is formed. Exposure of humans to ozone can cause several health effects, ranging from a light decrease in lung capacity to respiratory problems that require emergency hospital admittances, and death (RIVM, 2000).

The ozone formed is part of a pollution mix, which makes it difficult to determine the exact relationship between ozone exposure and its effects. Due to ethical reasons, clinical testing on humans of the effects of exposure can only be performed at relatively low doses.

In the EO5, the indicator 'Untimely deaths and emergency hospital admittances for respiratory, heart and pulmonary affections in the Netherlands associated with ozone' was included in the paragraph 'Loss of health related to environmental quality' (of the chapter 'Environment in the Netherlands') (see Fig. 2).

In Fig. 2 the number of expected deaths and hospital admittances in several age categories in the year 2030 are presented for two scenarios European Coordination (EC) and Global Competition (GC) taken from the Netherlands Bureau for Economic Policy Analysis (CPB, 1997). Also, the number of deaths and hospital admittances in 1995 are shown.

In the assessment of the expected number of deaths and the expected number of emergency hospital admittances due to the exposure to tropospheric ozone in the years 2010, 2020 and 2030, several calculation/modelling steps can be distinguished:

1. Determining societal/demographical developments.
2. Determining VOC and NO_x emissions in the Netherlands and abroad.
3. Determining O_3 concentrations.
4. Determining potential exposure to O_3.
5. Determining the number of deaths/hospital admittances.

Each step is described in more detail below.

1. Determining societal/demographical developments

Future societal and demographical developments in the Netherlands in the EO5 are based on two macro-economic scenarios that were developed by the Netherlands Bureau for Economic Policy Analysis (CPB) for the years 1995–2020 (RIVM, 2000). In the original CPB study three scenarios were developed: Divided Europe (DE), European Coordination (EC) and Global Competition (GC) (CPB, 1997). Due to the favourable economic developments in the years following the study, divided Europe was no longer viewed as a likely development path. Therefore the EO5 focuses on the EC and GC scenarios (RIVM, 2000).

The CPB scenarios cover the time period until 2020. The RIVM extended the main features of these scenarios (population, GDP and consumption) to the year 2030 (RIVM, 2000).

The population data in the CPB scenarios were based on the study 'Bevolking en arbeidsaanbod' (Population and labour supply) carried out by the Statistics Netherlands (CBS) and the CPB (CBS/CPB, 1997). The EO5 made use of the more detailed information of the CBS/CPB study (among others the age distribution of the population).

2. Determining VOC and NO_x emissions in the Netherlands and abroad

The societal and demographical developments of the fore mentioned long term scenarios together with the emission related policy measures that already had been agreed on ('fixed policy') formed the basis for the calculations of the VOC and NO_x emissions in the years 2010, 2020 and 2030. It was assumed that all emission related policy measures agreed on by the year 2000 will be implemented and that no new policy measures are taken (van Wee et al., 2001).

For the Netherlands the VOC and NO_x emissions for both the EC and GC variant in the years 2010, 2020 and 2030 were calculated for about 20 types of activities. The EC and GC scenarios were not specified for other countries. For the VOC and NO_x abroad one scenario was used for about five clusters of activities (interview). This scenario was mainly based on trend analysis and extrapolation of that trend assuming fixed policy.

3. Determining O_3 concentrations

The emission data of step 2 were used in runs of the 'EURopean Operational Smog model' (EUROS). EUROS is a model that describes chemical transformations, transport processes and deposition processes of several air pollution compounds. It computes reactions involved in the formation of ozone.

EUROS was used to calculate the diurnal 8 h maxima concentrations on a grid level of 50×50 km for the years 2010, 2020 and 2030. Since the formation of ozone is heavily determined by the meteorological conditions, and may therefore vary considerably from year to year, a 'worst case approach' was used. Several runs were done with the 1990 emissions, using the meteo data of different years. The meteo data of the year with the highest ozone formation was used in the EO5-runs (Eerens and van Dam, 2001).

Since the EUROS model was reasonably new at the time of the EO5, extra runs were done with the LOTOS model (interview; Blom and Roemer, 1997). The results were compared.

4. Determining potential exposure to O_3

Using a GIS application, the geographically explicit ozone concentrations resulting from step 3 and the geographically explicit information on the number of people in the different age categories from step 1 were combined. This yielded information per age category on how many people would be potentially exposed to different levels of ozone concentrations according to the EUROS calculations and population prognoses (interview).

5. Determining the number of deaths/hospital admittances due to exposure

The calculation of the number of possible deaths and emergency hospital admittances caused by exposure to the computed ozone concentrations was based on two epidemiological studies, carried out in the Netherlands (RIVM, 2000). In both studies (Vonk and Schouten, 1998 regarding hospital admittances and Hoek et al., 1997 regarding the number of deaths) an analysis was made of the relationship between ozone concentrations and deaths/hospital admittances by analysing data on measured ozone concentrations and official records stating deaths and health problems that required emergency hospital admittance. In these studies epidemiological analyzes yielded relative risk (RR) data on deaths and hospital admittances respectively for different age categories. The RR indicates the chances of developing a disease in an exposed group compared to those of a non-exposed group. The RR is calculated by dividing the incidence of the disease in the exposed group by the incidence of the disease in the non-exposed group.

The number of expected deaths for the years 2010, 2020 and 2030 was obtained by combining the RRs for the different age categories with the information on the calculated ozone concentrations that these age categories are potentially exposed to (interview).

discussion each participant individually filled in the score on that criterion. This procedure ensured that all experts based their individual evaluations on the same information shared by the group.

4.3. Final list of assumptions

An initial list of 18 assumptions was drafted during this study based on information from the interviews, information in the EO5 background reports and information in documentation on the EUROS model. Some of the assumptions in the list were explicitly stated in the aforementioned sources. Other, more implicit assumptions, were logically deducted from information on the assessment. After the changes and additions by the participants of the workshop the list of assumptions in the calculation chain of the indicator 'deaths and hospital admittances due to exposure to ozone' contained 24 assumptions.

Step 1: Determining societal/demographical developments.

(1) Assumption that the CPB scenarios were suitable for the societal-demographical developments in the EO5 (based on information in the EO5 (RIVM, 2000)).
(2) Assumption that for the analysis of long term environmental problems the macro-economic scenarios did not require adjustments based on the realisations in 1996–1998 (based on information in van Wee et al., 2001).
(3) Assumption that Divided Europe was no longer a plausible scenario (based on information in van Wee et al., 2001).
(4) Assumption that trends between 2010 and 2020 could be extrapolated to 2030 (based on information in van Wee et al., 2001).

Step 2: Determining VOC and NO_x emissions in the Netherlands and abroad

(5) Assumption that insufficiently specified policy directions did not need to be taken into account (also no 'scenario colouring') (based on information in van Wee et al., 2001).
(6) Assumption that fixed policy will be executed completely (though it is not lived up to for 100%) (van Wee et al., 2001).
(7) Assumption that emissions abroad will not differ between the EC and GC scenario (van Wee et al., 2001).
(8) Assumption that the emissions abroad in 2010 will be equal to the national emission ceilings from that year on (established in the framework of the UNECE Convention on Long-range Transboundary Air Pollution) (van Wee et al., 2001).

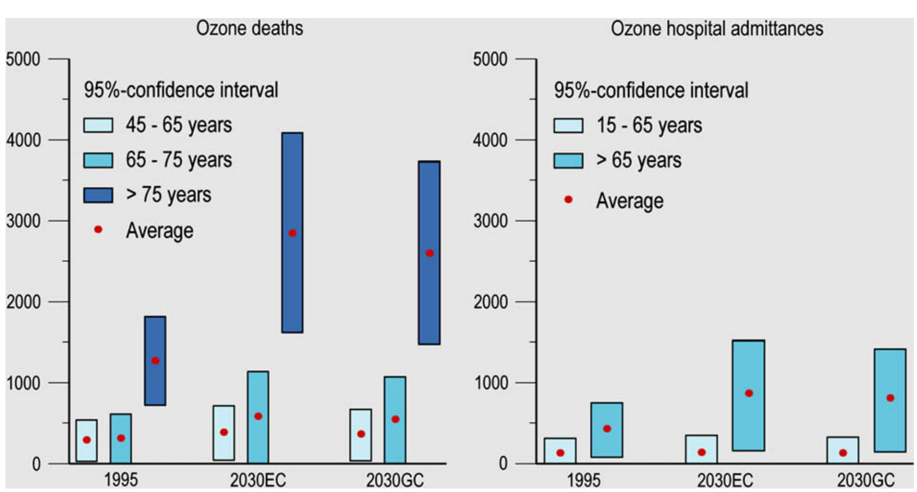

Fig. 2. Indicator 'deaths and hospital admittances due to exposure to ozone' (from RIVM, 2000).

(9) Assumption that the sectoral emissions abroad in 2010–2030 will experience the same development as the sectoral emissions in the Netherlands in that time period (van Wee et al., 2001).
(10) Assumption that the developments in emission factors and volume growth are harmonized in European context (identified in the workshop).

Step 3: Determining O_3 concentrations

(11) Assumption that the calculations needed to be carried out based on a worst case scenario for meteorological circumstances (based on information in Eerens and van Dam, 2001).
(12) Assumption that the worst case meteorological circumstances in the current time period will also be worst case meteorological circumstances in the future (based on information in Eerens and van Dam, 2001).
(13) Assumption that the global background concentration of ozone is constant (identified in the workshop).

Step 4: Determining potential exposure to O_3

(14) Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells (based on information in Eerens and van Dam, 2001).
(15) Assumption that the ozone concentration is representative for the exposure to ozone (identified in the workshop).

Step 5: Determining the number of deaths/hospital admittances due to exposure

(16) Assumption that deaths and hospital admittances related to ozone were relevant for the EO5 (based on information in the EO5; RIVM, 2000).
(17) Assumption that in case of ozone there is a linear dose-effect relationship (interview).
(18) Assumption that the degree of exposure of the future population will be similar to that of the population that lived during the time period of the epidemiological studies (based on interview information).
(19) Assumption that no better treatment methods will be developed (based on interview information).
(20) Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone (based on interview information).
(21) Assumption that death and illness are related to 8 h average ozone concentrations (top ozone concentrations are therefore not considered) (Vonk and Schouten, 1998; Hoek et al., 1997).
(22) Assumption that there is a direct causal relationship between ozone and death (identified in the workshop).
(23) Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR (identified in the workshop).
(24) Assumption that the Dutch epidemiological data are adequate for the whole of the Netherlands (identified in the workshop).

4.4. Key-assumptions

After the list of assumptions in the workshop had been checked and completed, the key-assumptions – the most important assumptions in the calculation chain of the indicator deaths and hospital admittances due to ozone – were identified by the workshop participants.

The results of the ranking exercise are presented in Table 2. The ranking resulted in 14 key-assumptions, which we have here labelled I–XIV. The ten assumptions that are not mentioned in the table received no points at all: they were not mentioned as one of the seven most important assumptions by any of the experts and are thus considered to be less important by the group.

4.5. Results for pedigree scores – step 3 in method

After the key-assumptions had been identified, the participants of the workshop assessed the potential value-ladenness of the assumptions. Starting with the key-assumption with the highest rank, scoring cards were filled in for the assumptions. During the workshop there was enough time available to complete the scoring cards of seven key-assumptions. The participants each individually completed the remaining seven scoring cards later on and sent them to the workshop organizer. Hence, the scoring on the criteria of these assumptions took place without group discussion.

The scores that were filled in on the cards were processed after the workshop had finished. In Table 2 the average scores of the group members on the pedigree criteria are listed together with the standard deviations. Strictly spoken the scales used are ordinal, but the scales are designed such that the intervals between the possible scores are more or less equal, so it still makes sense to use mean and standard deviation. Also, the mean scores averaged on all six pedigree criteria is given. All criteria were weighed equal.

In Fig. 3 the pedigree score results of the workshop are presented in diagrams. Using a diagnostic diagram the weakest links in the chain of assumptions can be identified. In Fig. 4 the average pedigree score (averaged on the six pedigree criteria and averaged on all experts) per assumption is plotted against the average score on ‘influence on results’. The assumptions most situated in the upper right corner of the graph can be viewed as the weakest links in the chain of assumptions. In this case, these are the assumptions:

- Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR (I).
- Assumption that the global background concentration of ozone is constant (IX).
- Assumption that the worst case meteorological circumstances in the current time period will also be worst case meteorological circumstances in the future (IV).
- Assumption that the developments in emission factors and volume growth are harmonized in European context (V).

4.6. Evaluation of the workshop

At the end of the workshop, the participants were asked to fill in an evaluation form. In general, the identification of assumptions was not found to be difficult. All respondents felt that the main assumptions in the chain were indeed identified. The ranking exercise was found quite easy to do. With respect to the filling in of the pedigree scoring cards, most participants indicated that the criteria were clear to them and that they felt they filled in meaningful scores (as opposed to arbitrary scores). However, several times it was indicated that the criterion ‘disagreement among stakeholders’ is hard to assess. Some participants suggested involving stakeholders in pedigree workshops. With respect to the usefulness and applicability of the method, most participants thought this was a useful exercise.

Table 2
Average scores (average over the five participants) on the pedigree criteria and standard deviations.

Assumptions	Situational limitations		Plausibility		Choice space		Agreement peers		Agreement stakeholders		Sensitivity views analyst		All pedigree criteria		Influence on results	
	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.	Avg.	St. dev.
Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR (I)	0,2	0,4	0,0	0,0	0,0	0,0	1,4	0,5	1,2	0,4	1,4	0,9	0,7	0,8	0,2	0,4
Assumption that emissions abroad will not differ for the EC and GC scenario (II)	0,4	0,9	0,2	0,4	0,2	0,4	0,2	0,4	1,2	0,4	0,8	0,4	0,5	0,6	1,1	0,9
Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells (III)	0,0	0,0	0,0	0,0	0,6	0,9	1,8	0,4	1,2	0,4	1,8	0,4	0,9	0,9	0,4	0,5
Assumption that the worst case meteorological circumstances in the current time period will also be worst case meteorological circumstances in the future (IV)	0,8	0,4	0,0	0,0	0,0	0,0	0,8	0,4	0,6	0,9	1,8	0,4	0,7	0,8	0,6	0,5
Assumption that the developments in emission factors and volume growth are harmonised in European context (V)	0,2	0,4	1,0	0,0	0,0	0,0	1,0	0,0	1,6	0,5	0,4	0,5	0,7	0,7	0,6	0,5
Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone (VI)	2,0	0,0	0,0	0,0	1,8	0,4	2,0	0,0	1,2	0,4	1,4	0,5	1,4	0,8	0,2	0,4
Assumption that in case of ozone there is a linear dose-effect relationship (VII)	1,8	0,4	1,4	0,5	0,2	0,4	2,0	0,0	1,0	0,0	0,2	0,4	1,1	0,8	0,6	0,5
Assumption that the ozone concentration is representative for the exposure to ozone (VIII)	0,4	0,5	1,0	0,7	1,2	0,8	1,0	0,0	1,0	1,0	1,0	1,0	0,9	0,7	0,8	0,8
Assumption that the global background concentration of ozone is constant (IX)	1,0	0,0	0,0	0,0	0,0	0,0	0,2	0,4	0,4	0,5	1,8	0,4	0,6	0,7	0,4	0,5
Assumption that the emissions abroad in 2010 will be equal to the national emission ceilings from that year on (established in the framework of the UNECE Convention on long-range transboundary air pollution) (X)	1,0	0,7	1,2	0,8	1,2	0,8	1,2	0,8	1,2	0,4	1,4	0,9	1,2	0,7	0,8	0,4
Assumption that the Dutch epidemiological data are adequate for the whole of the Netherlands (XI)	0,8	1,1	1,8	0,4	0,2	0,4	1,4	0,9	1,4	0,5	0,8	1,1	1,1	0,9	0,2	0,4
Assumption that the sectoral emissions abroad in 2010–2030 will experience the same development as the sectoral emissions in the Netherlands in that time period (XII)	1,0	0,7	1,4	0,5	0,2	0,4	1,6	0,5	1,2	0,4	1,0	1,0	1,1	0,7	1,4	0,5
Assumption that there is a direct causal relationship between ozone and death (XIII)	1,4	0,9	1,2	0,4	1,4	0,5	1,2	0,8	1,2	0,4	1,2	0,8	1,3	0,6	1,2	0,8
Assumption that the CPB scenarios were suitable for the societal-demographical developments in the EO5 (XIV)	1,0	1,0	2,0	0,0	1,2	0,8	1,8	0,4	1,4	0,5	1,2	0,8	1,4	0,7	1,0	0,7

Please cite this article in press as: Kloprogge, P., et al., A method for the analysis of assumptions in model-based environmental assessments, Environ. Model. Softw. (2009), doi:10.1016/j.envsoft.2009.06.009

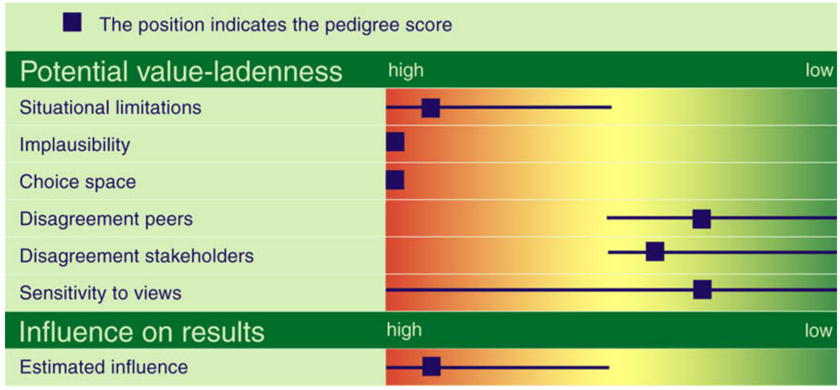

Fig. 3. Example of a pedigree chart for the assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the Relative Risk parameter.

5. Discussion and conclusions

In this article we presented and demonstrated a new method to identify, prioritize, analyze and deal with key-assumptions in assessments. This responds to a need felt among assessment practitioners and also expressed in literature dealing with the issue of the value-ladenness of assumptions. The method involves an analysis part, a revision part and a communication part. We developed a typology of value-ladenness that distinguishes between practical, general epistemic, disciplinary-bound epistemic and socio-political value-ladenness in the key-assumptions of an assessment. These types have been operationalized using a pedigree matrix (Table 1). The criteria defined in the matrix are implausibility, disagreement among peers, disagreement among stakeholders, choice space, influence of situational limitations, and sensitivity to view and interests of the analyst. In this way, the potential for value-ladenness in the production process of an assumption can be explored. This information, combined with information on the estimated influence of the assumption on the assessment results, helps to identify 'weak links' in the chain of calculations made in an assessment.

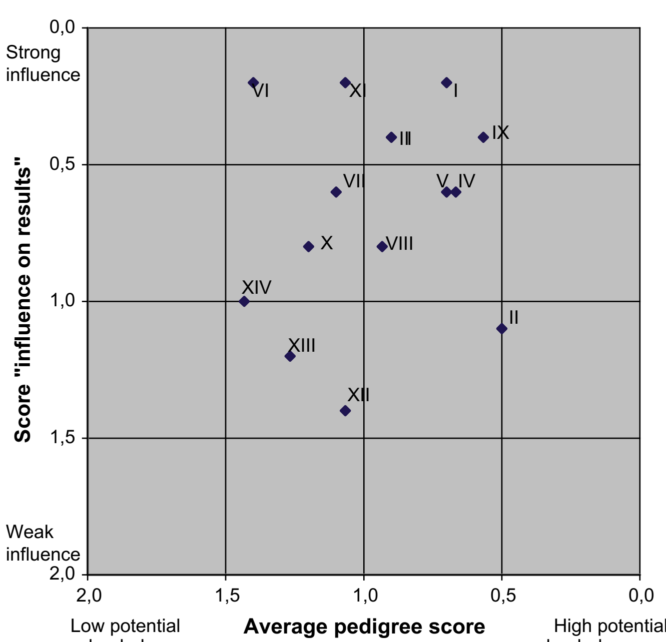

Fig. 4. Diagnostic diagram for indicator 'deaths and hospital admittances due to exposure to ozone'. The numbers I–XIV refer to the 14 key-assumptions listed in Table 2.

We tested the method on one indicator from the Dutch Fifth Environmental Outlook (EO5), 'deaths and emergency hospital admittances due to tropospheric ozone'. We identified implicit and explicit assumptions in the calculation chain by systematic mapping and deconstruction of the calculation chain, based on document analysis, interviews, and critical review. The resulting list of key-assumptions was reviewed and completed in a workshop. Analysis of the calculation chain of the selected indicator yielded a list of 24 assumptions. Fourteen key-assumptions were selected by the workshop participants as the most important ones, and prioritized. Combining the results of pedigree analysis and estimated influence, it became apparent that in the EO5 study all uncertainties other than the statistical uncertainty in relative risk had not been quantified. Specifically, the results identified as the weakest links of the calculation chain: (i) the assumption that the global background concentration of ozone is constant over the 30-year time horizon; (ii) the assumption that the worst case meteorological circumstances remain constant over time; and (iii) the assumption that the developments in emission factors and volume growth are harmonized in the European context. In their evaluation of the application of the method, participants found that the main assumptions in the chain were indeed identified, that the pedigree criteria were clear and their scores meaningful.

The method can be applied during the development of the assessment or after the assessment has already been carried out. In the latter case insight will be gained in potentially value-laden assumptions in the chain, but if the assessment has already been finalized, revisions are no longer possible. If the assessment has already been documented it neither will be possible to include the insights of the analysis in the assessment documentation. It may help however in communication surrounding the assessment and in extended peer review of the assessment. Applying the method during the assessment process is preferred: changes can be made to improve the assessment, and the results of the assumption-analysis can be included in the documentation.

In the daily practice of science-policy interface institutions such as the PBL, there are major challenges, institutional aspects and resource limitations that hamper large scale use of the method. The experts involved have busy schedules and the tasks of assessing and reporting consume the vast majority of the time allocated to a given regulatory assessment. Uncertainty and assumptions assessments are not a core task, so they seldom get priority. Thus, – as one reviewer of this paper phrased it – the question is not so much one of "how to?", but "when to?". It is not realistic to request that the full method becomes a standard daily practice. Instead it can be applied selectively to those cases where the policy relevance of the issue of assumptions is highest. Kloprogge et al. (2007) and Wardekker et al. (2008) suggest seven situation-specific factors to help identify such cases: (1) when being wrong in one direction could carry more serious consequences than being wrong in the other; (2) when uncertain outcomes can have a large influence on policy advice; (3) when indicators are close to a policy goal or threshold; (4) when there is the possibility of large effects or catastrophic events; (5) in cases of societal controversy; (6) when value-laden choices are in conflict with interests or views of stakeholders; and (7) when public distrust in outcomes that show low risk can be expected.

An other issue in the application of the method is the question of expertise. Slottje et al. (2008) provide general practical guidance on how to arrive at a balanced sample of experts. For this particular method, two issues require further reflection. First, the calculation chains involved often span a broad spectrum of expertises and disciplines. When asked to assess the assumptions dealing with a part of the calculation chain that does not fall directly within their field, experts tend to be reluctant to answer. In some cases this may be justified (i.e., they are truly not competent enough to provide a meaningful evaluation), while in other cases it may not be (i.e., they perceive themselves as incompetent when in fact, they are relatively competent). It should be noted, however, that participants do not decide what assumptions will be made in a study. In stead their role is one of reflective quality control. Second, if only one expert from each segment of the calculation chain is involved, it makes it difficult to generate a reflexive dialogue on the basis for the expert opinions expressed and disagreements encountered. However, due to the knowledge, perspectives, habits and preferences every analyst has, some blind spots may occur. It is therefore advisable to include peers in the analysis, from the same institute and preferably from other institutes as well. Stakeholders also have their knowledge and perspectives on the issues at hand. Therefore it is valuable to involve them in the process, if possible, as was indeed advised by participants in their evaluation of the workshop. In this way, this method for the analysis of assumptions can be used as a tool for extended peer review processes (see Craye et al., 2009, for an example). Involving stakeholders further increases the resource intensiveness of the method, which limits large scale applicability in the science-policy interface.

Still, the assumption-analysis enables the analysts to make a conscious, well-underpinned, transparent choice, increases reflexivity, and pinpoints the issues in the chain that are important to communicate to the audience of the assessment report.

Acknowledgements

The research on which this article reports was performed in the period 2002–2005 at the Netherlands Environmental Assessment Agency (and was funded by Utrecht University). We would like to thank all interviewees and workshop participants for their contributions to this study. We are also grateful for the valuable comments of the advisory committee, which consisted of Laurens Brandes, Anton van der Giessen, Leon Janssen, Peter Janssen and Rob Maas. Finally, we would like to thank Wim Turkenburg and Jerry Ravetz for their comments on an earlier draft of this article.

References

Blom, J.G., Roemer, M.G.M., 1997. Description of the 3D LOTOS Model. Part I: Dynamics. Report MAS-N9701. CWI, Amsterdam, The Netherlands.
CBS/CPB, 1997. Bevolking en arbeidsaanbod: drie scenario's tot 2020 [Population and Labour Supply: Three Scenarios to 2020]. Sdu uitgevers, The Hague, The Netherlands.
CPB, 1997. Economie en fysieke omgeving [Economy and Physical Environment]. Sdu uitgevers, The Hague, The Netherlands.
Craye, M., Laes, E., van der Sluijs, J., 2009. Re-negotiating the role of external cost calculations in the Belgian nuclear and sustainable energy debate. In: Pereira Guimaraes, A., Funtowicz, S. (Eds.), Science for Policy. Oxford University Press, Oxford.
Eerens, H.C., van Dam, J.D. (Eds.), 2001. Grootschalige luchtverontreiniging en depositie in de Nationale Milieuverkenning 5 [Large-Scale Air Pollution and Deposition in the National Environmental Outlook 5]. Report 408129016, RIVM. Bilthoven, The Netherlands.
Funtowicz, S.O., Ravetz, J.R., 1990. Uncertainty and Quality in Science for Policy. Kluwer Academic Publishers, The Netherlands.
Funtowicz, S.O., Ravetz, J.R., 1993. Science for the post-normal age. Futures 25, 735–755.
Hoek, G., Verhoeff, A., Fischer, P., 1997. Daily Mortality and Air Pollution in the Netherlands, 1986–1994. Report 1997-481. Wageningen University, Wageningen, The Netherlands.
Hornberger, G.M., Spear, R.C., 1981. An approach to the preliminary analysis of environmental systems. Journal of Environmental management 12, 7–18.
Huesemann, M.H., 2002. The inherent biases in environmental research and their effects on public policy. Futures 34, 621–633.
Jakeman, A.J., Letcher, R.A., Norton, J.P., 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling and Software 21 (5), 602–614.
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: Quickscan Hints & Actions List. RIVM/MNP and Utrecht University, Bilthoven and Utrecht, The Netherlands.
Janssen, P.H.M., Petersen, A.C., van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R., 2005. A guidance for assessing and communicating uncertainties. Water Science & Technology 52 (6), 125–131.
Jasanoff, S., 1990. The Fifth Branch: Science Advisers as Policymakers. Harvard University Press, Cambridge, MA.
Jasanoff, S., Wynne, B., 1998. Science and decisionmaking. In: Rayner, S., Malone, E.L. (Eds.), Human Choice and Climate Change. The Societal Framework, Vol. 1. Battelle Press, Columbus, OH, pp. 1–87.
Keepin, B., Wynne, B., 1984. Technical analysis of IIASA energy scenarios. Nature 312, 691–695.
Kloprogge, P., van der Sluijs, J.P., 2002. Choice processes in modelling. In: Rizzoli, A.E., Jakeman, A.J. (Eds.), Integrated Assessment and Decision Support, Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society. iEMSs, Manno, Switzerland, pp. 96–101.
Kloprogge, P., van der Sluijs, J.P., Petersen, A.C., 2005. A Method for the Analysis of Assumptions in Assessments: Exploring the Value-Ladenness of Two Indicators in the Fifth Dutch Environmental Outlook. Report 550002010. Netherlands Environmental Assessment Agency and Utrecht University, Bilthoven and Utrecht.
Kloprogge, P., Van der Sluijs, J.P., Wardekker, J.A., 2007. Uncertainty Communication: Issues and Good Practice. Department of Science, Technology and Society. Utrecht University, Utrecht. Available from. http://www.nusap.net/guidance.
Lindfords, L.G., Christiansen, K., Hoffmann, L., Virtanen, Y., Juntilla, V., Hanssen, O.J., Rønning, A., Ekvall, T., Finnveden, G., 1995. Nordic Guidelines on Life-Cycle Assessment. Nordic Council of Ministers, Copenhagen.
Maxim, L., van der Sluijs, J.P., 2007. Uncertainty: cause or effect of stakeholders' debates? Analysis of a case study: the risk for honey bees of the insecticide Gaucho®. Science of the Total Environment 376, 1–17.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the Earth sciences. Science 263, 641–646.
Petersen, A.C., 2006. Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice. Het Spinhuis Publishers, Apeldoorn, The Netherlands and Antwerp, Belgium.
Petersen, A.C., Janssen, P.H.M., van der Sluijs, J.P., Risbey, J.S., Ravetz, J.R., 2003. RIVM/MNP Guidance for Uncertainty Assessment and Communication: Mini-Checklist & Quickscan Questionnaire. RIVM/MNP and Utrecht University, Bilthoven and Utrecht, The Netherlands.
Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H., Kassahun, A., 2005. Quality assurance in model based water management: review of existing practice and outline of new approaches. Environmental Modelling and Software 20 (10), 1201–1215.
Refsgaard, J.C., van der Sluijs, J.P., Hojberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process: a framework and guidance. Environmental Modelling and Software 22 (11), 1543–1556.
Risbey, J.S., van der Sluijs, J.P., Kloprogge, P., Ravetz, J.R., Funtowicz, S., Corral Quintana, S., 2005. Application of a checklist for quality assistance in environmental modelling to an energy model. Environmental Modeling and Assessment 10 (1), 63–79.
RIVM, 2000. Nationale Milieuverkenning 5, 2000–2030 [National Environmental Outlook 5, 2000–2030]. RIVM, Bilthoven, The Netherlands.
Saltelli, A., Chan, K., Scott, E.M. (Eds.), 2000. Sensitivity Analysis. John Wiley & Sons Publishers, Chichester, England.
Schneider, S.H., 1997. Integrated assessment modeling of global climate change: transparent rational tool for policy making or opaque screen hiding value-laden assumptions. Environmental Modeling and Assessment 2, 229–248.

Slottje, P., van der Sluijs, J.P., and Knol, A.B., 2008. Expert Elicitation – Methodological suggestions for its use in environmental health impact assessments. RIVM Letter Report 630004001/2008, National Institute for Public Health and the Environment, Bilthoven, 56 pp. Available from: http://www.nusap.net/downloads/reports/Expert_Elicitation.pdf.
Stirling, A., 1999. On Science and Precaution in the Management of Technological Risk, Vol. I: A Synthesis Report of Case Studies. Report EUR 19056 EN, European Commission Institute for Prospective Technological Studies, Seville, Spain.
Stirling, A., 2001. Inclusive deliberation and scientific expertise: precaution, diversity and transparency in the governance of risk. PLA Notes 40, 67–71.
van Asselt, M.B.A., 2000. Perspectives on Uncertainty and Risk: The PRIMA Approach to Decision Support. Kluwer, Dordrecht.
van Asselt, M.B.A., Langendonck, R., van Asten, F., van der Giessen, A., Janssen, P.H.M., Heuberger, P.S.C., Geuskens, I., 2001. Uncertainty and RIVM's Environmental Outlooks: Documenting a Learning Process. ICIS/RIVM, Maastricht/Bilthoven, The Netherlands.
van der Sluijs, J.P., 1997. Anchoring amid Uncertainty: On the Management of Uncertainties in Risk Assessment of Anthropogenic Climate Change. PhD dissertation, Utrecht University, Utrecht, The Netherlands.
van der Sluijs, J.P., 2002. A way out of the credibility crisis around model-use in integrated environmental assessment. Futures 34, 133–146.
van der Sluijs, J.P., 2005. Uncertainty as a monster in the science policy interface: four coping strategies. Water Science and Technology 52 (6), 87–92.
van der Sluijs, J.P., van Eijndhoven, J.C.M., Shackley, S., Wynne, B., 1998. Anchoring devices in science for policy: the case of consensus around climate sensitivity. Social Studies of Science 28 (2), 291–324.
van der Sluijs, J.P., Potting, J., Risbey, J., van Vuuren, D., de Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S.O., Kloprogge, P., Nuijten, D., Petersen, A.C., Ravetz, J.R., 2002. Uncertainty Assessment of the IMAGE/TIMER B1 CO_2 emissions scenario, using the NUSAP method. Report 410200104. Dutch National Research Program on Climate Change, Bilthoven, The Netherlands.
van der Sluijs, J.P., Risbey, J.S., Kloprogge, P., Ravetz, J.R., Funtowicz, S.O., Corral Quintana, S., Guimarães Pereira, A., De Marchi, B., Petersen, A.C., Janssen, P.H.M., Hoppe, R., Huijs, S.W.F., 2003. RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance. Utrecht University and RIVM/MNP, Utrecht and Bilthoven, The Netherlands.
van der Sluijs, J.P., Janssen, P.H.M., Petersen, A.C., Kloprogge, P., Risbey, J.S., Tuinstra, W., van Asselt, M.B.A., Ravetz, J.R., 2004. RIVM/MNP Guidance for Uncertainty Assessment and Communication: Tool Catalogue for Uncertainty Assessment. Utrecht University and RIVM/MNP, Utrecht and Bilthoven, The Netherlands.
van der Sluijs, J.P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J., Risbey, J., 2005. Combining quantitative and qualitative measures of uncertainty in model based environmental assessment: the NUSAP system. Risk Analysis 25 (2), 481–492.
van der Sluijs, J.P., Petersen, A.C., Janssen, P.H.M., Risbey, J.S., Ravetz, J.R., 2008. Exploring the quality of evidence for complex and contested policy decisions. Environmental Research Letters 3, 024008 (9 pp).
van Wee, G.P., Kuijpers-Linde, M.A.J., van Gerwen, O.J., 2001. Emissies en kosten tot 2030 bij het vastgesteld milieubeleid. Achtergronddocument bij de Nationale Milieuverkenning 5. Report number 408129013. RIVM, Bilthoven, The Netherlands.
Vonk, J.M., Schouten, J.P., 1998. Daily Emergency Hospital Admissions and Air Pollution in the Netherlands, 1982–1986 and 1986–1995. Report VROM 96.140072. University of Groningen, Groningen, The Netherlands.
Walker, W.E., Harremoës, P., Rotmans, J., van der Sluijs, J.P., van Asselt, M.B.A., Janssen, P.H.M., Krayer von Krauss, M.P., 2003. Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment 4 (1), 517.
Wardekker, J.A., van der Sluijs, J.P., Janssen, P.H.M., Kloprogge, P., Petersen, A.C., 2008. Uncertainty communication in environmental assessments: views from the Dutch science-policy interface. Environmental Science & Policy, 11, 627-641. doi:10.1016/j.envsci.2008.05.005.