Netherlands Environmental Assessment Agency

Copernicus Institute

MNP Report 550002010/2005
NWS-E-2005-122

A method for the analysis of assumptions in assessments

Exploring the value-ladenness of two indicators in the Fifth Dutch Environmental Outlook

P. Kloprogge[1], J.P. van der Sluijs[1], A.C. Petersen[2]

[1] Department of Science, Technology and Society, Copernicus Institute for Sustainable Development and Innovation, Utrecht University

[2] Information Services and Methodology (IMP), Netherlands Environmental Assessment Agency (MNP)

Contact:
J.P. van der Sluijs
Copernicus Institute for Sustainable Development and Innovation, Utrecht University
j.p.vandersluijs@chem.uu.nl

Netherlands Environmental Assessment Agency (MNP), P.O. Box 303, 3720 AH Bilthoven, the Netherlands;

Tel: +31-30-274 274 5; Fax: +31-30-274 4479; www.mnp.nl/en

Abstract

A method for the analysis of assumptions in assessments

Exploring the value-ladenness two indicators in the Fifth Dutch Environmental Outlook

When analysts do quantitative scientific assessments of complex policy problems, they have to make many assumptions in the chain of calculations that lead to the end results. These assumptions inevitably involve – to some degree – subjective judgements by the analysts. This report offers a method for analysing the ‘subjective component’, or ‘value-ladenness’, of such assumptions. ‘Value-ladenness’ is not reserved for politically controversial assumptions, since there are also epistemic, non-political components in the subjectivity of choices.

The proposed method enables analysts and stakeholders to conduct a well-structured discussion on potentially value-laden assumptions and their influence on the results of the assessment. An analysis of the nature and extent of the potential value-ladenness conveys the ‘weak links’ in the assessment and may lead to the decision to adjust the assessment. The method furthermore offers assistance in communicating crucial assumptions.

The method was tested *ex post* on the indicators ‘change in length of the growth season’ and ‘deaths and emergency hospital admittances due to exposure to ozone’ in the fifth Dutch Environmental Outlook, and led to a list of weak links in the calculation of these indicators.

Keywords:

value-ladenness, subjectivity, assumptions, assessments, pedigree matrix

Rapport in het kort

Een analysemethode voor aannames in assessments

Onderzoek naar de waardengeladenheid van twee indicatoren in de vijfde Milieuverkenning

Bij het opstellen van kwantitatieve wetenschappelijke assessments voor complexe beleidsproblemen zijn onderzoekers genoodzaakt veel aannames te maken in de keten van berekeningen die naar het eindresultaat leiden. Deze aannames zijn noodzakelijkerwijs – in meer of minder mate – gebaseerd op de subjectieve oordelen van de onderzoekers. Dit rapport biedt een methode voor de analyse van de 'subjectieve component', of 'waardengeladenheid, van dergelijke aannames. 'Waardengeladenheid' is niet voorbehouden aan politiek controversiële aannames, omdat er ook epistemische, niet-politieke componenten kleven aan de subjectiviteit van keuzes.

De voorgestelde methode stelt wetenschappers en stakeholders in staat een goed gestructureerde discussie te voeren over potentieel waardengeladen aannames en hun invloed op de resultaten van de assessment. Een analyse van de aard en grootte van de potentiële waardengeladenheid geeft een beeld van de 'zwakke schakels' in de assessment en kan aanleiding geven tot het herzien van de assessment. Verder biedt de methode hulp bij de communicatie rond cruciale aannames.

De methode werd *ex post* getest op de indicatoren 'verandering lengte groeiseizoen' en 'vroegtijdige sterfte en spoedopnamen in het ziekenhuis geassocieerd met ozon' uit de vijfde Milieuverkenning. Dit resulteerde in een lijst met zwakke schakels voor deze indicatoren.

Trefwoorden:

waardengeladenheid, subjectiviteit, aannames, assessments, pedigree matrix

Contents

Summary **7**

1 Introduction **9**

1.1 Assumptions in assessments 9
1.2 Objective of the study 12
1.3 Outline of the report 12

2 Analytical framework **13**

2.1 Assumptions and value-ladenness 13
2.2 A pedigree matrix for reviewing potential value-ladenness of assumptions 14

3 Setup of the test cases **21**

3.1 Indicators from the fifth Environmental Outlook as test cases 21
3.2 Method 22
3.2.1 Interviews 22
3.2.2 Document analysis 23
3.2.3 Expert workshop 23

4 Indicator 1: change in length of the growth season **25**

4.1 Description of the calculation chain 25
4.2 Identified assumptions 29
4.3 Key assumptions 31
4.4 Results of pedigree scores 34

5 Indicator 2: deaths and hospital admittances due to exposure to ozone **39**

5.1 Description of the calculation chain 39
5.2 Identified assumptions 42
5.3 Key assumptions 44
5.4 Results pedigree scores 44

6 Discussion of the test case results and procedures **51**

6.1 Key assumptions 51
6.2 Pedigree scores 52
6.3 Evaluation of the workshop 53

7 A method for the analysis of assumptions **55**

8 Discussion and conclusions **63**

References **67**

Appendix A: List of interviewees and workshop participants **71**

Appendix B: Example of a pedigree scoring card **73**

Appendix C: Overview of the method for the analysis of assumptions in assessments **75**

Summary

When analysts do quantitative assessments, they have to make many assumptions in the chain of calculations leading to the end results. Assumptions cannot, by definition, be objectively determined as they contain a 'subjective component'. In the literature this value-ladenness of assessments is discussed extensively, but this has so far not led to a systematic strategy for dealing with value-laden assumptions. In this report, a method is described for analysing the potential value-ladenness of assumptions, answering the following research questions:

- How can assumptions in a calculation chain be mapped and prioritised?
- How can potential value-ladenness of assumptions be analysed?
- How can assumptions be handled, based on the inventory and prioritisation of assumptions, and analysis of potential value-ladenness?
- How can the potential value-ladenness of an assessment be communicated?

'Value-ladenness' is not reserved for politically controversial assumptions, since there are also non-political components in the subjectivity of choices. Four types of value-ladenness are distinguished depending on the context in which they originate, so value-ladenness can occur in the socio-political sense, disciplinary sense, epistemic sense and in the practical sense. Based on the NUSAP method by Funtowicz and Ravetz, a set of 'pedigree criteria' was developed to assess the potential value-ladenness and the influence of the assumptions on the end results of the assessment.

The method was tested *ex post* at an expert workshop, reflecting on the indicators 'change in length of the growth season' and 'deaths and emergency hospital admittances due to exposure to ozone' of the fifth Dutch Environmental Outlook (EO5). The first step was to make a list of assumptions for each indicator in the calculation chain. The list for the growth-season case contained 23 assumptions, the one for the ozone case contained 24. Next, the experts prioritised the assumptions, which resulted in a list of 7 key assumptions in the growth-season case and a list of 4 key assumptions in the ozone case. The potential value-ladenness of every key assumption was subsequently assessed by the experts using the pedigree criteria. The assumptions with a low pedigree score (that is, a high potential for value-ladenness) *and* a strong influence on the results of the assessment can be viewed as 'weak links' in the calculation chain. Combining the average scores on the pedigree criteria and the average scores on the estimated influence of the key assumptions resulted in a number of 'weak links' in the calculation chains of both indicators. The weak links identified for the growth-season indicator are:

- assumptions implied in the choice of a General Circulation Model.
- an assumption that the European Coordination and Global Competition scenarios were suitable for the EO5 analyses for the Netherlands, and that the choice for the range in global emission scenarios was suitable for the global analysis.

- an assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5.
- assumptions regarding the carbon cycle, including the feedback through land use.

The following weak links were identified for the ozone indicator:

- an assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the relative risk.
- an assumption that the global background concentration of ozone is constant.
- an assumption that the worst-case meteorological circumstances in the current period will also be worst-case meteorological circumstances in the future.
- an assumption that the developments in emission factors and volume growth are harmonised in the European context.

After a list of weak links is established, the nature and extent of the potential value-ladenness of these assumptions can be examined in more detail. On the basis of these findings, assumptions in the calculation chain can be adjusted or a sensitivity analysis can be added to the assessment. Finally, the insights gained from the previous steps can be used in communicating the main assumptions in an assessment: i.e. the key assumptions, the background to these assumptions and their implications in terms of robustness of results can be communicated to the users of the assessment. The method thus contains an analysis, a revision and a communication component.

The proposed method for analysing the assumptions helps then to raise awareness on assumptions, to systematically reflect upon them, to reconsider choices that were made, and to shape communication on crucial assumptions. The method thus enables a well-structured discussion among scientists and stakeholders on potentially value-laden assumptions in assessments. The cases showed that results of the method can be sensitive to the group composition and procedure details. Further research and testing can provide more insight into this sensitivity and into how stakeholders can be involved in the method.

1 Introduction

1.1 Assumptions in assessments

While making quantitative assessments of complex policy 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 assessments generally involve (model) calculations and other data operations that produce the end results of the assessments. We refer to this set of calculations and operations in an assessment as the ‘calculation chain’ behind a given outcome of interest of an assessment.

Throughout the causal chain the analyst will use the knowledge that is available to him 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 can be made explicitly or implicitly. Often, an assumption explicitly made by the analyst, automatically implies additional, implicit assumptions.

Since assumptions by definition cannot objectively be determined (for: something is assumed), there always is some kind of ‘subjective component’ that the analyst brings in while setting assumptions. Two analysts assessing the same issue will probably not make the exact same assumptions in the calculation chain. Consequently , an assessment is not made up of objective, value-free scientific facts alone. Because of this, assessments can be considered to be value-laden to a certain degree.

Numerous studies from the history and sociology of science have problematised the classic distinction between facts and values. Scientific facts and knowledge claims, especially when produced in 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) even states with respect to environmental studies that ‘it is intrinsically impossible to carry out objective and value-free scientific research, and, that in fact, all environmental science is inherently biased by subjective opinions and values’.

Assumptions and models

Since the eighties of the twentieth century computer models are increasingly being used in complex assessments: they enable analysts to simulate reality and run several scenarios, thereby integrating knowledge from different disciplines. The models themselves, the use of models and the transparency of models have been criticised over the years (see e.g. Saltelli, 2002). Hornberger and Spear (1981) argued that, because of the many degrees of freedom in simulation models, ‘virtually any desired behaviour (can be produced), often with both plausible structure and parameter values’. In their publications several authors in the risk assessment and Integrated Assessment modelling field problematise the subjective component in models. According to Oreskes *et al.* (1994) we should wonder how much of a model is based ‘on observation and measurement of accessible phenomena, how much is based on informed judgment, and how much is convenience?’. In their critical review of IIASA energy scenarios Keepin and Wynne (1984) speak of ‘informal guesswork’ and a lack of peer review and quality control, ‘raising questions about political bias in scientific analysis’. Van der Sluijs (2002) states that ‘the building of an Integrated Assessment Model inevitably involves subjective choices and value-laden assumptions’. He argues that a lack of transparency with regard to these assumptions may lead to scandals and loss of trust in the scientific basis for policies, as was clearly demonstrated in the ‘de Kwaadsteniet affair’[1] and in the fore mentioned Keepin and Wynne IIASA affair. In a critique on (macro economic) climate risk assessment models, Schneider (1997) illustrates the ‘dangers that analytic methods with limited capabilities bring to the public debate given that not all potential users of IAM results will be aware of hidden values or assumptions that are inherent in all such tools.’

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 testing the robustness and sensitivity of results. 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 IA 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 important elements in a strategy for dealing with value-laden assumptions.

Assumptions and the management of uncertainty and quality

Especially when dealing with complex issues that are surrounded by uncertainties, many assumptions have to be made. The research community has put a lot of effort into conceptual

[1] In 1999, Hans de Kwaadsteniet, an employee of the Netherlands National Institute for Public Health and the Environment accused the institute of lying to and deceiving the public with studies based on computer models that were poorly validated and that were hardly based on measurements. This led to an extensive debate about the use of models in assessment studies in the Netherlands (van der Sluijs, 2002; van Asselt, 2000).

research on uncertainty (e.g., van der Sluijs, 1997; Walker *et al.*, 2003), and in the development of strategies for dealing with uncertainties. Examples of methods for dealing with uncertainty 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.*, 2002a, 2002b, 2005), the Guidance for Uncertainty Assessment and Communication (van der Sluijs *et al.*, 2003, 2004; Petersen *et al.*, 2003; Janssen *et al.*, 2003), and application of the precautionary principle (Harremoës *et al.*, 2002; UNESCO COMEST, 2005).

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 of models, methods and assessment studies assumptions are one of the elements that are tackled. In good practice guidelines (e.g., good practice guidelines for LCA (Lindfors *et al.*, 1995)) and checklists for practitioners (e.g., a checklist for quality assistance in environmental modelling (Risbey *et al.*, 2001), the 'Good Modelling Practice Handbook' (STOWA/RIZA, 1999), and the HarmoniQua State-of-the-Art Report on Quality Assurance in modelling related to river basin management (Refsgaard, 2002)) 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 subjective component of the assumptions made in an assessment. In current uncertainty and quality management strategies the subjective component of assumptions is merely analysed and handled in a general manner. Thus, dealing with the subjective component in assumptions has not yet been an issue in itself.

When assessments are discussed or criticised in the 'policy arena', assumptions also seem to receive only partial attention. Discussions tend to focus on assumptions that are viewed as policy relevant. For all assumptions however, including 'technical' and 'scientific' assumptions, it can be said that differing assumptions may lead to differing outcomes of the assessment, which may consequently lead to differing interpretations of the problem (e.g., the magnitude of the problem), which in their turn may lead to differing policy recommendations and measures. In this way all assumptions that influence the outcome of the assessment can turn out to be policy relevant and should be subject to critical review.

Although the subjective component of assumptions has been extensively problematised in literature, this has not so far led to a systematic strategy for dealing with this. To remedy this omission, a method for analysing and dealing with potentially value-laden assumptions was developed in this study.

1.2 Objective of the study

This study aimed at developing a method to systematically identify, prioritise and review assumptions 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 research questions are:

- How can assumptions in a calculation chain be mapped and prioritised?
- How can potential value-ladenness of assumptions be analysed?
- How can assumptions be handled, based on the inventory and prioritisation of assumptions and the analysis of potential value-ladenness?
- How can the potential value-ladenness of an assessment be communicated?

The method was demonstrated and tested ex post on two indicators in the fifth Dutch Environmental Outlook (RIVM, 2000). The Dutch National Environmental Outlook (EO) is an assessment of key environmental indicators outlining different future scenarios for a time period of several decades, which is issued by the Netherlands Environmental Assessment Agency (MNP) (until January 2006 part of the National Institute for Public Health and the Environment (RIVM)).

1.3 Outline of the report

In chapter 2 (theoretical framework) it is further explored how and in what way assumptions can become value-laden. Chapter 2 also presents the core element of the method we developed in this study: a ‘pedigree matrix’ for the assessment of the potential value-ladenness of assumptions. How the two test cases were conducted is described in chapter 3. Chapters 4 and 5 present the results of the cases. The interpretation and comparison of the results and procedure of the two cases are discussed in chapter 6. The elements from the theoretical framework together with the experiences of the two test cases resulted in our method for analysis of potentially value-laden assumptions in assessments, which is presented in chapter 7. In chapter 8 the discussion and conclusions are presented.

2 Analytical framework

In section 2.1 we elaborate on how and in what way assumptions can become value-laden. For the review of the potential value-ladenness of assumptions we developed a so-called 'pedigree matrix', which is presented in section 2.2.

2.1 Assumptions and value-ladenness

In this study we zoom in on value-ladenness of assumptions, starting from the viewpoint of the analyst carrying out the assessment. Often, an analyst can choose from more than one option for a specific assumption. Making an assumption involves going through a choice process. In Kloprogge and van der Sluijs (2002) several phases that an analyst goes through in choice processes are distinguished:

1. Determining that an issue requires choices to be taken. (In case of implicit assumptions this step is skipped.)
2. Making an inventory of options to choose from
3. Choosing one or more options
4. Checking/evaluating the choice

Kloprogge and van der Sluijs (2002) have 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 4 types of value-ladenness of assumptions: value-ladenness in a socio-political sense (e.g., assumptions may be coloured by political preferences of the analyst), in a disciplinary sense (e.g., assumptions are coloured by the discipline in which the analyst was educated), in an epistemic sense (e.g., assumptions are coloured by the approach that the analyst prefers) and in a practical sense (e.g., the analyst is forced to make simplifying assumptions due to time constraints).

When reviewing assumptions, it is 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 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. For this purpose we designed a pedigree matrix containing criteria with which the room for value-ladenness can be explored.

2.2 A pedigree matrix for reviewing potential value-ladenness of assumptions

The idea of a pedigree matrix was introduced by Funtowicz and Ravetz (1990). It is one of the elements of the NUSAP method for uncertainty management (Funtowicz and Ravetz, 1990, van der Sluijs *et al.*, 2005). NUSAP is an acronym conveying five qualifiers of scientific information: **N**umeral, **U**nit, **S**pread, **A**ssessment, and **P**edigree. It is a heuristic for good practice in science for policy to address uncertainty in quantitative information. NUSAP has extended 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. The 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 www.nusap.net). Many of these criteria are hard to measure in an objective way. Assessment of pedigree involves qualitative expert judgement. To minimise 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 (see Table 2.1 for an example of a pedigree matrix). 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, based on the elements presented in section 2.1, we have developed a pedigree matrix to review the potential value-ladenness of assumptions. This matrix is presented in Table 2.2.

It contains the following criteria: influence of situational limitations, plausibility, choice space, agreement among peers, agreement among stakeholders, sensitivity to view and interests of the analyst, influence on results. The scale we used (score 0, 1 and 2) is similar to the scale used in previous NUSAP studies (see www.nusap.net for examples). Many pedigree matrices also contain a score 3 or 4. In those cases four, respectively five modes were used to span up the available judgements on the criteria. In our case not more than three were used; to our opinion no modes could be developed that added useful extra descriptions regarding the criteria.

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.

Score	**Theoretical Structure**	**Data input**	**Peer acceptance**	**Colleague consensus**
4	Established theory	Review	Total	All but cranks
3	Theory-based model	Historic/Field data	High	All but rebels
2	Computational model	Extrapolated	Medium	Competing schools
1	Statistical processing	Calculated	Low	Embrionic field
0	Definitions	Expert guess	None	No opinion

Table 2.1. An example of a pedigree matrix: pedigree matrix for research (Funtowicz and Ravetz, 1990)

The criteria for the review of potential value-ladenness are elaborated on below.

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.

We distinguished the scores: the choice for the assumption was hardly influenced by situational limitations (score 2), moderately influenced (score 1) and a totally different assumption would have been chosen had there not been any limitations (score 0).

Plausibility

Although it is often not possible to assess whether the approximation created by the assumption is in accordance with reality, mostly an (intuitive) assessment can be made of the plausibility of the assumption. Here, we distinguished three degrees of plausibility: the assumption is plausible (score 2), the assumption is acceptable (score 1) or the assumption is fictive or speculative (score 0).

Type of value-ladenness	**Practical**	**Epistemic**	**Epistemic**	**Disciplinary, epistemic**	**Socio-political**	**Socio-political**	
Criteria → **Score ↓**	**Influence situational limitations**	**Plausibility**	**Choice space**	**Agreement among peers**	**Agreement among stakeholders**	**Sensitivity to view and interests of the analyst**	**Influence on results**
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

Table 2.2. The pedigree matrix for the assessment of the potential value-ladenness of assumptions

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 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.

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: hardly any alternative assumptions available (score 2), limited choice from alternative assumptions (score 1), ample choice from alternative assumptions (score 0). 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 an 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 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.

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 sense and in an epistemic sense.[2] We distinguished three situations: many peers would have made the same assumption (score 2), several would have made the same assumption (score 1), few would have made the same assumption (score 0).

Potential socio-political value-ladenness influencing the analyst’s choice for a certain assumption is dealt with in the criteria ‘sensitivity to view and interests of the analyst’ and ‘agreement among stakeholders’.

[2] There is a link to controversy, as not all peers would agree to the same assumption if there was controversy regarding the issue of the assumption. However, if the majority of peers would choose the same assumption, still the score would be 2 (‘many peers would have made the same assumption’). The occurrence of controversies in the scientific field thus is not always visible in the score. Reasoned the other way around, a score of 0 (‘few peers would have made the same assumption’) does not imply that there are controversies surrounding the assumption: it is possible that all peers agree on the issue, yet that the analyst for some reason has chosen a different assumption. The same applies to the criterion ‘agreement among stakeholders’.

Agreement among stakeholders

Stakeholders, though mostly not actively involved in carrying out assessments, may also choose 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 'agreement 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.

We distinguished three situations: many stakeholders would have made the same assumption (score 2), several would have made the same assumption (score 1), few would have made the same assumption (score 0).

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.

Three levels of sensitivity are distinguished: the choice for the assumption is hardly sensitive to the views and interests of the analyst (score 2), the choice for the assumption is moderately sensitive (score 1), the choice for the assumption is sensitive (score 0).

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 analyse the influence on outcomes of interest of the assessment (the component 'spread' in the NUSAP method). 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 built. The pedigree matrix therefore includes a column 'influence on results' for a rough indication of the influence on the assessment results.

We distinguished three levels of influence: the assumption has only influence locally in the calculation chain (score 2), the assumption greatly determines the results of the step (score 1) and the assumption greatly determines the results of the indicator (score 0).

The pedigree matrix 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.*, 2002b). 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 of the assumptions (averaged over the 6 pedigree criteria) (y axis). Assumptions that are situated in the upper right corner are in the 'danger zone' (i.e., high potential value-ladenness and strong influence on the assessment results), the ones in the lower left corner are in the 'safe zone' (i.e., low potential value-ladenness and weak influence on the assessment results). See Figure 2.1 for the layout of such a diagram.

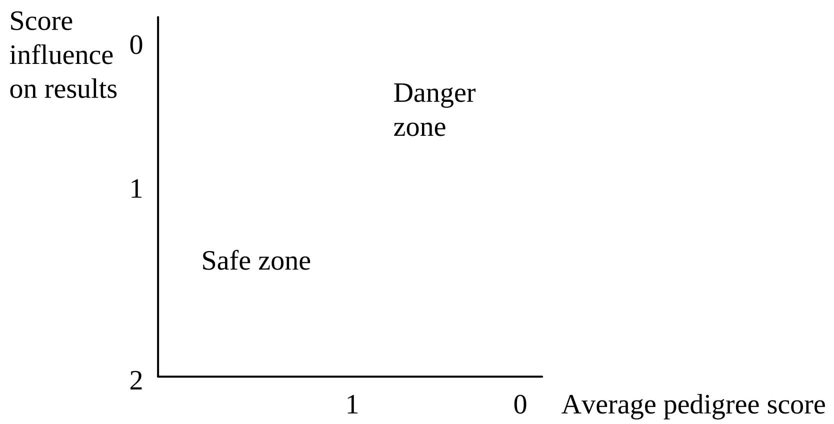


Figure 2.1. A diagnostic diagram for the identification of weak links in the calculation chain (adapted from van der Sluijs *et al.*, 2002b).

3 Setup of the test cases

In section 3.1 the Dutch National Environmental Outlooks and the two selected indicators from the fifth Environmental Outlook are introduced. The method applied in the test cases is described in section 3.2.

3.1 Indicators from the fifth Environmental Outlook as test cases

Approximately once every four years the Dutch government issues a National Environmental Policy Plan (NEP). The NEP indicates the policies the government plans to implement in the following four years. As input for each NEP, the Netherlands Environmental Assessment Agency (MNP) (until January 2006 part of the National Institute for Public Health and the Environment (RIVM)) prepares an assessment of key environmental indicators outlining different future scenarios for a time period of several decades: the National Environmental Outlook (EO).

Since the EO is of high importance for the Dutch environmental policy, and is situated on the interface of science and policy, this assessment is an interesting case for demonstrating and testing a method for the analysis of potentially value-laden assumptions. Besides this, the EO's have been subject to an extensive uncertainty study (van Asselt, 2000; van Asselt *et al.*, 2001). In the study a method was developed in which uncertainties can be approached from different perspectives. One of the recommendations of the report was to design a 'guidance' for uncertainty management. This recommendation has lately resulted in a 'Guidance for Uncertainty Assessment and Communication' (van der Sluijs *et al.*, 2003, 2004; Petersen *et al.*, 2003; Janssen *et al.*, 2003). The Guidance among others contains a tool catalogue with tools for the assessment of uncertainties (van der Sluijs *et al.*, 2004). The method for analysis of assumptions presented in this report is included in the catalogue.

Within the MNP 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 MNP, 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 MNP/RIVM) 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.

We applied our method for the analysis of assumptions ex post to two indicators in the fifth EO (RIVM, 2000), 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 study we examined two indicators: ‘change in length of the growth season’ (RIVM, 2000; p. 71 for the European scale and page 91 for the global scale) and ‘deaths and emergency hospital admittances due to exposure to ozone’ (RIVM, 2000; page 208). These were chosen based on the interest of the advisory board group involved in this study, policy relevance, the differing problem characteristics (complex versus more or less structured), the use of computer models in the calculation chain (IMAGE[3] and EUROS[4], respectively) and the availability of information on the assessments carried out for the EO5.

3.2 Method

In order to identify assumptions in the calculation chain of each indicator first an overview of each of the modelling chains behind the two indicators had to be reconstructed. For this purpose we used the so called ‘EO-explorer’, a Microsoft Access-database developed by the MNP in which the MNP members documented for each indicator the information flow leading to the end result of the indicator. The EO-explorer lists in diagrams who created what information, with which models and data and to whom the information was passed on. With this information, supplemented with information from interviews with the RIVM persons who were involved in the EO5 and the EO5 documentation, the two calculation chains were reconstructed.

Next, information was gathered on the choices that were made in the calculation chains. Assumptions were deducted from these choices and from the EO5 documentation.

The list of assumptions was checked and completed in a workshop with experts. In this workshop, the assumptions were subsequently prioritised and – using the pedigree matrix – reviewed with regard to potential value-ladenness.

A description of the interviews, analysis of EO5 documentation and the workshop is given below.

3.2.1 Interviews

The key persons involved in the chains at the time of the EO5-study were interviewed for this study. The goal of the interviews was to reconstruct the calculation chains that were used in

---

[3] IMAGE stands for Integrated Model to Assess the Greenhouse Effect. A description can be found in Alcamo *et al.*, 1998.

[4] EUROS stands for EURopean Operational Smog model. A description can be found in Van Loon, 1996.

the assessments of the two indicators, and to identify and analyse the choices and assumptions that were made in the assessments. For this purpose the analysts who carried out the calculations in the chains were interviewed, together with the EO5 project members who were responsible for the chapters in which the chosen indicators were presented. A list of the persons who were interviewed for both cases can be found in Appendix A.

3.2.2 Document analysis

Only part of the information on which the assessment of the indicators is based, is documented in the EO5. More information on the calculation procedure and the intermediary results can be found in RIVM background reports. Some of these reports were written as background reports of the EO5, others were issued as ‘regular’ RIVM-reports. These reports were studied. In addition, for information on IMAGE and EUROS books, reports, CD-roms and information on the RIVM intranet were studied.

The documentation analysis provided insight into the steps in the calculation chains of the indicators and on choices and assumptions that were made throughout the chain.

3.2.3 Expert workshop

Objectives of the workshop

Based on the information from the interviews and the EO5 documentation, a list was compiled of assumptions that were made throughout the calculation chains. In the expert workshop the list of assumptions was checked and completed. After that, the assumptions were prioritised and reviewed using the pedigree matrix, presented in section 2.2. Finally, the method used in the workshop was evaluated.

The participants

The participants all had expertise on (part of) the calculation chain of one of the indicators. Some of the participants had contributed to the EO5 assessments. They did not only have expertise on parts of the chain, but also had detailed knowledge of the choices made during the EO5 assessment. 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 several experts from other institutes.

The group of the growth season case consisted of 8 participants, including a facilitator and a note taker. The ozone group consisted of 5 participants, including a facilitator and a note taker. In the ozone group a second note taker was present, who did not take part in the exercises. The ozone group had been planned with 8 participants. Three participants had to cancel shortly before the workshop took place. See Appendix A for a list of the participants.

Workshop programme

Four hours were available for the workshop. Both groups worked separately. In advance all participants had received a short description of the entire calculation chain. During the workshop itself the facilitators 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 Appendix B for an example of a scoring card (in Dutch)). 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 7 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 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.

There was not enough time during the workshop to fill in all scoring cards. The participants were requested to fill in the remaining cards later on and to send them to the workshop organiser.

Finally, the method used during the workshop was evaluated in a short plenary session. Everyone was asked to give a short first reaction to the workshop. The participants were furthermore requested to fill in an elaborate evaluation form later on (see Appendix C (in Dutch)).

4 Indicator 1: change in length of the growth season

Section 4.1 presents a description of the calculation chain for the indicator 'change in length of the growth season'. The major assumptions in this chain are presented in section 4.2. The assumptions that the workshop participants identified as most important in the chain are listed in section 4.3. Finally, in section 4.4 the results of the pedigree scoring exercise in the workshop are presented.

4.1 Description of the calculation chain

The growth season is the yearly time period in which precipitation and temperature enable plant growth (RIVM, 2000). In the EO5, the approach of the Food and Agriculture Organisation (FAO) was used: growth season was defined as the yearly time period in which the temperature and soil moisture exceed certain values (interview).

Since climate change may affect temperature and precipitation, it may cause changes in the length of the growth season.

In the EO5, the indicator 'change in length of the growth season' was included in the chapter 'Environment on the global scale' and 'Environment on European scale'. The figures printed in the EO5 are reproduced below (Figure 4.1 and 4.2). The figures indicate the increase or decrease of the number of days of the growth season on average per decennium for the period 1990 to 2050, for the world and Europe, respectively.

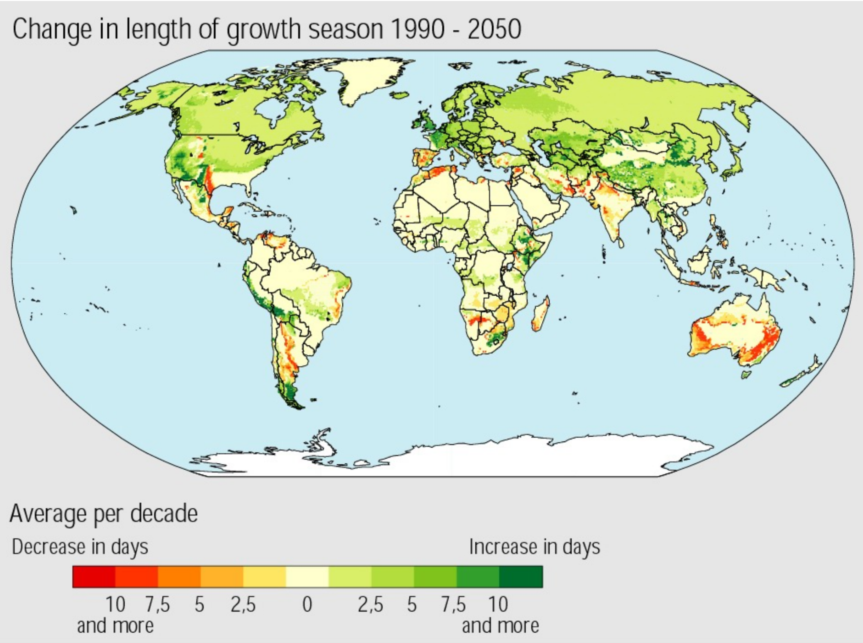


Figure 4.1. EO5-figure indicator 'change in length of the growth season' in the chapter 'Environment on the global scale' (RIVM, 2000; p. 91; translation PK).

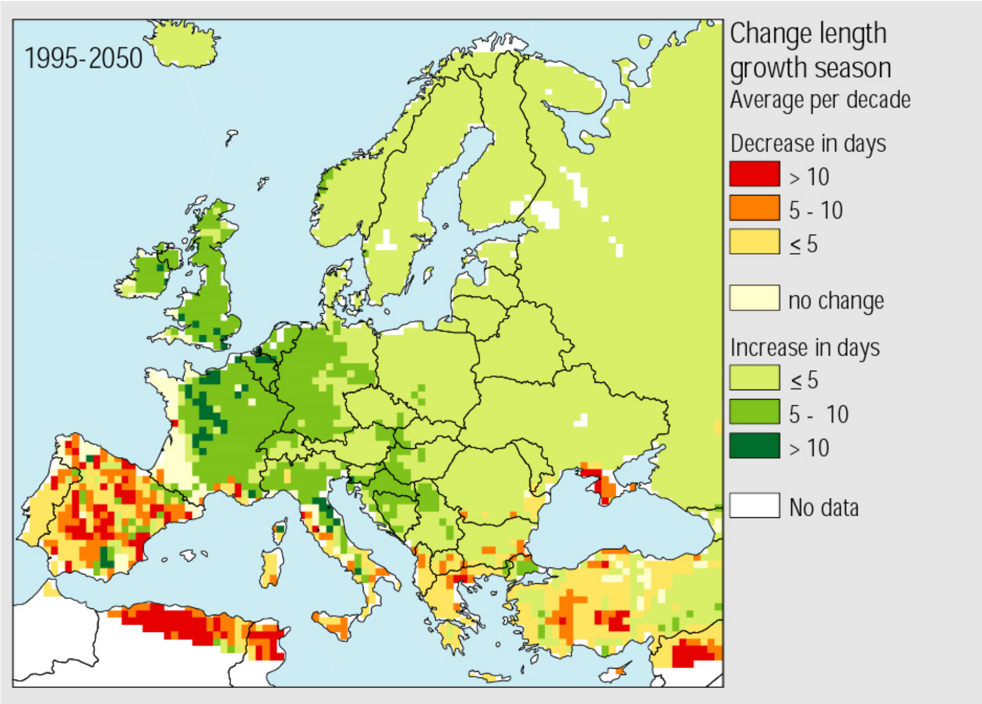


Figure 4.2. EO5-figure indicator 'change in length of the growth season' in the chapter 'Environment on European scale' (RIVM, 2000; p. 71; translation PK).

In the assessment of the expected changes in the length of the growth season in the time period 1995 - 2050 several calculation/modelling steps can be distinguished:

1. Determining societal/demographical developments
2. Determining global emissions of greenhouse gasses, sulphate aerosols and ozone precursors
3. Determining the changes in GHG concentrations, radiative forcing and the resulting change in global mean annual temperature
4. Determining changes in temperature and precipitation on grid-scale
5. Determining diurnal changes in temperature and precipitation on grid-scale
6. Determining soil moisture on grid-scale
7. Determining the change in the length of the growth season on grid-scale

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 the Global Competition (GC) and the European Coordination (EC) scenario developed by the Netherlands Bureau for Economic Policy Analysis (CPB) (RIVM, 2000; CPB, 1997). For indicators concerning the European and global scale, scenarios on societal and demographical developments on a global scale are needed. These global developments in the EO5 are based on the A1 and B2 scenario of the Special Report on Emission Scenarios (SRES) of the Intergovernmental Panel on Climate Change (IPCC) (Nakicenovic and Swart, 2000). These two scenarios were chosen, since, out of the 4 SRES-scenarios, they resemble the GC-scenario and the EC-scenario the most (RIVM, 2000; interview).

For the A1 scenario only the population data were processed exogenously. The economic growth was calculated using the Worldscan model (CPB, 1999). These calculations had already been carried out by the RIVM for the purpose of the SRES report (interview).

In case of the B2 scenario no IMAGE calculations had been carried out for population growth and economic growth. These data were taken from the so called ‘marker scenario’ of the SRES (interview).

2. *Determining global emissions of greenhouse gasses, sulphate aerosols and ozone precursors*

For the B2 scenario, the emissions due to energy use and industrial activity were directly taken from the marker scenario B2, as presented in the SRES report (interview).

For the A1 scenario, the RIVM had done emission calculations herself for the purpose of the SRES report. Using the RIVM-model ‘Targets IMage Energy Regional’ (TIMER) version 0.4 the societal and demographic data from step 1 were used to calculate future energy use (interview). The emissions from energy use and industrial activity were then calculated by the TIMER Emission Module (TEM) (interview). Emissions related to land use and from biotic sources were calculated in the Terrestrial Environment System (TES), which is part of the RIVM-model IMAGE 2.1.2 (Alcamo *et al.*, 1998; interview).

3. *Determining the changes in GHG concentrations, radiative forcing and the resulting change in global mean annual temperature*

The changes in global mean annual temperature were calculated by the Atmosphere-Ocean System (AOS) of the IMAGE 2.1.2 model. Based on the emissions, the Atmospheric Composition model (part of AOS) first computed the atmospheric concentrations of the GHGs and sulphate aerosols (Alcamo *et al.*, 1998).

The CO_2 flux between the atmosphere and plant communities in the ocean was calculated by the Ocean Biosphere and Chemistry model (part of AOS). The CO_2 flux between the terrestrial environmental system and the atmosphere was calculated in the TES-model.

Based on the resulting concentrations of the GHGs and sulphate aerosols, the Zonal Atmospheric Climate model (part of AOS) computed the radiative forcing. The next step was calculating the heat balance of the atmosphere and ocean. With these results the changes in global mean annual temperature were determined (Alcamo *et al.*, 1998).

4. Determining changes in temperature and precipitation on grid-scale

Using results from General Circulation Models (GCMs) and a current climate database, the change in global mean annual temperature was translated into changing patterns of precipitation and surface temperature on grid-scale[5] (Alcamo *et al.*, 1998; interview).

In the method used, the GCM output for changes in monthly temperature and precipitation in each GCM cell is first normalised (among others to account for differences in annual mean values between years). These normalised patterns are then multiplied by the global mean annual temperature change resulting from step 3 in the calculation chain. Next, the resulting monthly changes in temperature and precipitation are overlaid onto the grid cells. The changes in temperature and precipitation are 'added' to the temperature and precipitation data derived from a current climate database (Alcamo *et al.*, 1998; interview).

In this way global climate change was downscaled to changing patterns of temperature and precipitation on grid level. The downscaling also took into account the regional effects of sulphate aerosols.

5. Determining diurnal changes in temperature and precipitation on grid-scale

After step 4, for each grid cell the months of the year in which the average temperature is below 5 degrees Centigrade were no longer taken into account with regard to the growing season (interview).

For the other months, the monthly average temperature and precipitation values were downscaled to diurnal values. For both temperature and precipitation, a method was used that takes seasonal differences into account; for each month a different temperature and precipitation cycle was used (interview).

6. Determining soil moisture on grid-scale

The soil moisture was calculated in IMAGE 2.1.2 by subtracting evaporation from precipitation (interview). The texture, depth and storage-capacity of the soil were also taken into account. Per grid cell the dominant soil type was determined. The evaporation was mainly determined by the temperature.

7. Determining the change in the length of the growth season on grid-scale

The steps described above made it possible to determine for a certain year per grid cell, on which days (of the months in which the average temperature exceeded 5 degrees Centigrade)

---

[5] The grid refers to the discrete numerical representation of variables. In GCMs, for example, the horizontal grid size is typically a few hundred kilometers.

the temperature and soil moisture exceed the growth season minima. This cluster of days is referred to as the growing season.

The number of days that the requirements are met, were compared to the number of days in other years. The difference is the change in length of the growing season. In the figures of the indicator in the EO5 (see Figure 4.1 and 4.2) the average change per decennium is given.

Finally, some remarks should be made regarding the calculations. Calculating the length of the growth season is not a specific output goal of the IMAGE model. The length of the growth season is mainly calculated in the model in order to be able to calculate the feasible production of food crops.

For the EO5 existing model results were used. No extra calculations or adjustments took place.

4.2 Identified assumptions

In this study a list of 19 assumptions was drafted based on information from the interviews, information in the EO5 background reports and information in documentation on the IMAGE model. Some of the assumptions in the list were explicitly stated in the fore mentioned sources. Other, more implicit assumptions were logically deducted from information on the assessment. The list of assumptions was presented in the expert workshop. The participants were asked to check these assumptions, which was mainly intended to check our interpretation of assumptions in the list that were not explicitly stated in the information sources. Next, the participants were asked to add important assumptions they thought were missing.

After the changes and additions by the participants, the list of assumptions in the calculation chain of the indicator ‘change in length of the growth season’ – 23 in total – was as follows (grouped per step in the chain):

Step 1: Determining societal/demographical developments

1) Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis (based on information in the EO5 (RIVM, 2000)/adapted during the workshop)

Step 2: Determining global emissions of greenhouse gasses, sulphate aerosols and ozone precursors

2) Assumptions on intra-sectoral structural changes in economy (identified as key parameter in van der Sluijs *et al.*, 2002b)

3) Assumptions on learning curves for energy technologies (identified as key parameter in van der Sluijs *et al.*, 2002b)

4) Assumptions on stocks and depletion of fossil fuels and related 'supply curves' (identified as key parameter in van der Sluijs *et al.*, 2002b)

5) Assumptions on the autonomous and price-induced energy efficiency improvement (identified as key parameter in van der Sluijs *et al.*, 2002b)

6) Assumption implied by the fact that institutional factors are not taken into account in the models (identified in the workshop)

Step 3: Determining the changes in GHG concentrations, radiative forcing and the resulting change in global mean annual temperature

7) Assumption on the climate sensitivity (mentioned in interview)

8) Assumption that the natural variability of the climate (related to El Niño, volcano eruptions and solar activity) did not need to be considered (mentioned in interview)

9) Assumptions regarding the carbon cycle, including the feedback through land use (identified in workshop)

10) Assumption implied by the fact that a genuine (i.e. back and forth) coupling with the nitrogen cycle is left out (identified in workshop)

11) Assumptions regarding the atmospheric chemistry, including aerosols (identified in workshop)

12) Assumption regarding the delay in the climate system (identified in workshop)

Step 4: Determining changes in temperature and precipitation on grid-scale

13) Assumption that the cloud coverage is constant (mentioned in interview)

14) Assumptions implied by the choice for a GCM (mentioned in interview)

15) Assumptions on the interpolations of weather data in sites where the monitoring network is not dense enough (mentioned in interview)

16) Assumption that a grid cell is homogeneous regarding land use, soil properties, etcetera and the chosen size of the grid cell (the GCMs operate with larger grid cells) (identified in workshop)[6]

Step 5: Determining diurnal changes in temperature and precipitation on grid-scale

17) Assumption that the monthly temperature and precipitation cycles (based on data of the current time period) will not change in future time periods (mentioned in interview)
18) Assumption implied by the fact that the calculations take place on a monthly basis while the results are presented in days (identified in workshop)

Step 6: Determining soil moisture on grid-scale

19) Assumption that a simple model could be used to calculate the soil moisture for crops (based on interview information)
20) Assumption that one vegetation type (mixture of grass and forest) could be used for the evaporation on a global scale (identified in workshop)

Step 7: Determining the change in the length of the growth season on grid-scale

21) Assumption that the indicator 'change in length of the growth season' was relevant for the EO5 (based on information in the EO5 (RIVM, 2000))
22) Assumption that the change per decennium offered more relevant information than the change in length of growth season in 2050 in comparison to the length in 1990 (based on interview information)
23) Assumption that the applied definition for the growth season will also be relevant in the year 2050, despite potential adaptations of crops and agricultural technology (based on interview information)

4.3 Key assumptions

After the list of assumptions had been established in the workshop, the key-assumptions – the most important assumptions in the calculation chain of the change in length of the growth season – were identified by the workshop participants. The procedure devised beforehand was that each participant would select the 7 scoring cards with the – according to him or her – most important assumptions in the calculation chain. The participant would then sort the cards in order of importance and would note the results on a form. Based on this, a group

[6] This assumption also holds for steps 5 till 7.

ranking would be compiled, indicating the key assumptions in the chain according to the group.

However, in the growth season group, most participants found it hard to individually rank the assumptions based on the limited information they had on the assumptions in the IMAGE model. The group therefore decided to have a plenary discussion on the importance of the identified assumptions and to arrive in this way at the 7 most important assumptions. The discussion took place assumption by assumption. Every participant indicated to the group how important he or she thought the assumption to be and mentioned the arguments behind this evaluation. Importance was interpreted by the group as influence of the assumption on the outcome of the length of the growth season, or on the presentation of the results regarding length of the growth season. The scores of the participants on the importance – ranging from not at all important (--) to very important (++) – were written down on the whiteboard. Table 4.1 lists the results of this exercise. In the table, for each score (--, -, 0, + and ++) the number of participants that chose that score is indicated.

During the discussion the participants decided that the assumptions:

- Assumptions on intra-sectoral structural changes in economy (2)
- Assumptions on learning curves for energy technologies (3)
- Assumptions on stocks and depletion of fossil fuels and related ‘supply curves’ (4)
- Assumptions on the autonomous and price-induced energy efficiency improvement (5)
- Assumption implied by the fact that institutional factors are not taken into account in the models (6)

could be viewed as part of the ‘Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis’ (1). They were therefore not evaluated (and not included in Table 4.1).

Also, the ‘Assumption that the monthly temperature and precipitation cycles (based on data of the current time period) will not change in future time periods’ (17) was left out of the evaluation.

After the assumptions were discussed, and the scores for the assumptions were known, the participants together selected the 7 assumptions with the most plusses as the key-assumptions of the calculation chain.

The following 7 assumptions were thus identified as the key-assumptions of the growth-season calculation chain (in order of importance; starting with the most important one):

I. Assumptions implied by the choice for a GCM
II. Assumption on the climate sensitivity

		Importance of the assumption					
Step	**Assumption**	**--**	**-**	**0**	**+**	**++**	**cum. result**
4	Assumptions implied by the choice for a GCM	0	0	0	2	6	14
3	Assumption on the climate sensitivity	0	0	0	3	5	13
1	Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis	0	1	0	2	5	11
3	Assumptions regarding the atmospheric chemistry, including aerosols	0	0	1	4	3	10
3	Assumptions regarding the carbon cycle, including the feedback through land use	0	0	2	3	3	9
7	Assumption that the indicator 'change in length of the growth season' was relevant for the EO5	0	0	1	6	1	8
3	Assumption regarding the delay in the climate system	0	1	3	2	2	5
4	Assumption that a grid cell is homogeneous regarding land use, soil properties, etcetera and the chosen size of the grid cell (the GCMs operate with bigger grid cells)	0	1	4	3	0	2
6	Assumption that a simple model could be used to calculate the soil moisture for crops	0	3	1	4	0	1
3	Assumption implied by the fact that a genuine (i.e. back and forth) coupling with the nitrogen cyclus is left out	1	0	5	2	0	0
4	Assumption that the cloud coverage is constant	1	2	3	2		-2
5	Assumption implied by the fact that the calculations take place on a monthly basis while the results are presented in days	0	4	2	2	0	-2
6	Assumption that one vegetation type (mixture of grass and forest) could be used for the evaporation on a global scale	0	4	4	0	0	-4
3	Assumption that the natural variability of the climate (related to El Niño, volcano eruptions and solar activity) did not need to be considered	4	2	0	0	2	-6
4	Assumptions on the interpolations of weather data in sites where the monitoring network is not dense enough	2	4	0	2	0	-6
7	Assumption that the applied definition for the growth season will also be relevant in the year 2050, despite potential adaptations of crops and agricultural technology	3	5	0	0	0	-11
7	Assumption that the change per decennium offered more relevant information than the change in length of growth season in 2050 in comparison to the length in 1990	4	4	0	0	0	-12

Table 4.1. Evaluation of the workshop participants on the importance of the identified assumptions in the growth-season calculation chain. Per assumption the number of participants that chose the scores '--' '-', '0', '+' and '++' are indicated. The cumulative results were calculated in this table by counting every '--' as -2 points, every '-' as -1, every '0' as 0, every '+' as 1 and every '++' as 2 points.

III. Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis

IV. Assumptions regarding the atmospheric chemistry, including aerosols

V. Assumptions regarding the carbon cycle, including the feedback through land use

VI. Assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5

VII. Assumption regarding the delay in the climate system

4.4 Results of pedigree scores

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 (see section 3.2.3). The facilitators were instructed beforehand to facilitate a group discussion per criterion, after which the group members would write down their score individually on the discussed criterion. In case of the growth season group, however, the discussion on the criteria had for the larger part already taken place while ranking the assumptions. It was therefore decided in the group that all participants individually filled in the scoring cards of the 7 key-assumptions, without further group discussion on the criteria.

The scores that were filled in on the cards were processed after the workshop had finished. In Table 4.2 the average scores of the group members on the pedigree criteria are listed together with the standard deviations. Also, the mean scores averaged on all 6 pedigree criteria are given. All criteria were weighed equal.

Using a diagnostic diagram the weakest links in the chain of assumptions can be identified (see section 2.2). In Figure 4.3 the average pedigree score (averaged on the 6 pedigree criteria and averaged on all experts) per assumption is plotted against the average score on the ‘influence on results’. As can be seen in the graph, the differences between the assumptions are small. However, 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:

- Assumptions implied by the choice for a GCM
- Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis
- VI. Assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5

	Situational limitations		Plausibility		Choice space		Agreement peers		Agreement stakeholder		Sensitivity to views		All pedigree criteria		Influence on results	
Assumptions	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev	Avg.	St.dev
Assumptions implied by the choice for a GCM (I)	1.3	0.7	1.0	0.8	0.6	0.9	1.1	0.6	0.5	0.5	0.8	0.9	0.9	0.3	0.1	0.4
Assumption on the climate sensitivity (II)	1.8	0.7	1.3	0.7	1.0	0.9	1.4	0.5	0.9	0.6	1.1	0.8	1.2	0.3	0.1	0.4
Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis (III)	0.9	0.6	1.3	0.5	0.9	0.8	1.4	0.5	0.8	0.7	1.3	0.7	1.1	0.3	0.1	0.4
Assumptions regarding the atmospheric chemistry, including aerosols (IV)	1.1	0.8	1.0	0.0	0.6	0.7	1.4	0.5	1.3	0.5	1.3	0.7	1.1	0.3	0.4	0.5
Assumptions regarding the carbon cycle, including the feedback through land use (V)	1.3	0.5	1.4	0.7	0.8	0.7	1.0	0.7	1.0	0.8	0.6	0.5	1.0	0.3	0.4	0.5
Assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5 (IV)	1.3	0.5	1.5	0.5	0.3	0.5	1.0	0.0	1.3	0.5	1.0	0.5	1.0	0.4	0.3	0.5
Assumption regarding the delay in the climate system (VII)	1.1	0.8	1.5	0.5	1.0	0.5	1.3	0.5	1.3	0.5	1.1	0.8	1.2	0.2	1.0	0.5

Table 4.2. Average scores (average over the 8 participants) on the pedigree criteria and standard deviations.

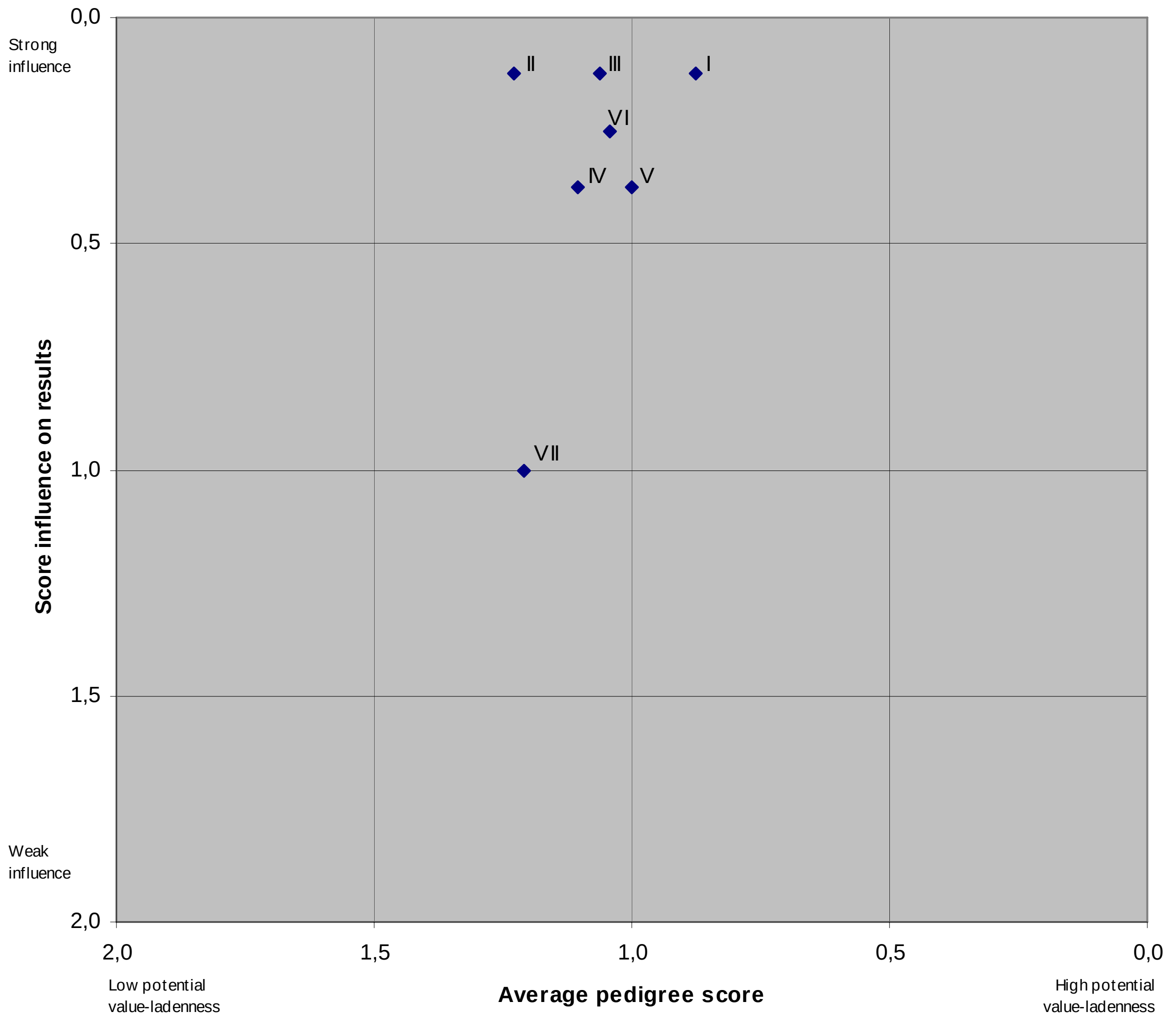


Figure 4.3: a diagnostic diagram of the assumptions of the indicator ‘change in length of the growth season’

I. Assumptions implied by the choice for a GCM
II. Assumption on the climate sensitivity
III. Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis
IV. Assumptions regarding the atmospheric chemistry, including aerosols
V. Assumptions regarding the carbon cycle, including the feedback through land use
VI. Assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5
VII. Assumption regarding the delay in the climate system

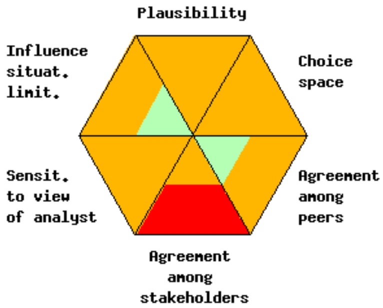
– VI. Assumption that the indicator ‘change in length of the growth season’ was relevant for the EO5

– V. Assumptions regarding the carbon cycle, including the feedback through land use

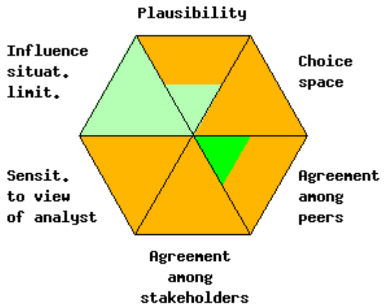
The average score on all pedigree criteria averaged on all experts is useful to find the weakest links in the calculation chain. However, in these average scores a considerable amount of information is lost: the average scores do not indicate what the nature and extent of the value-ladenness is, nor do they indicate the degree to which the experts agree on their scores.

For a further analysis of the potential value-ladenness of the assumptions, the original scores of the experts will have to be used. However, due to the large numbers of scores, it is difficult to obtain an overview and to draw conclusions from these data. An option for analysis would be to aggregate the data and calculate averages of the scores of the experts. However, in doing so, information on the extent to which there is agreement and disagreement between the participants of the workshop is lost. We therefore use diagrams, which visualise the scores using colours and provide information on the disagreement among the participants. The diagrams are explained in Box 1.

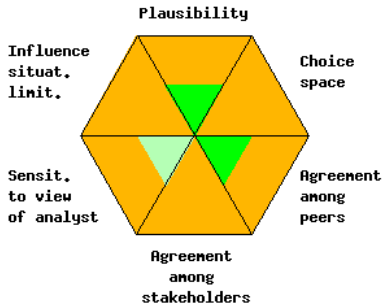
In Figure 4.5 the diagrams for the 7 key-assumptions are presented.



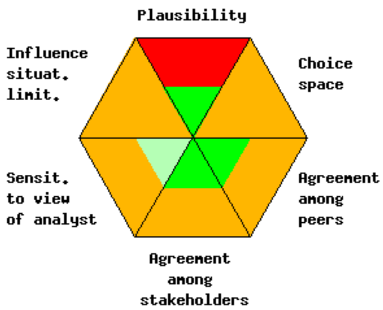
Assumptions implied by the choice for a GCM



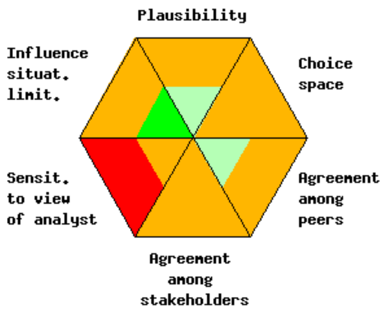
Assumption on the climate sensitivity



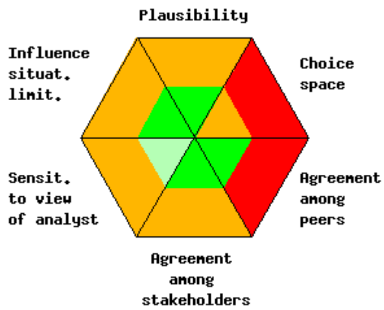
Assumption that the EC and GC scenarios were suitable for the EO5 analyses for the Netherlands and that the choice for the range in global emission scenarios was suitable for the global analysis



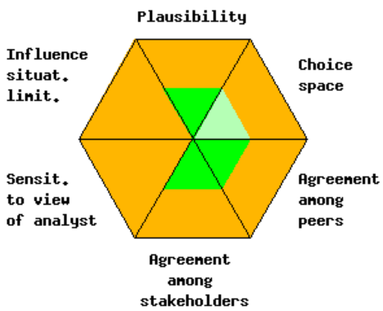
Assumptions regarding the atmospheric chemistry, including aerosols



Assumptions regarding the carbon cycle, including the feedback through land use



Assumption that the indicator 'change in length of the growth season' was relevant for the EO5



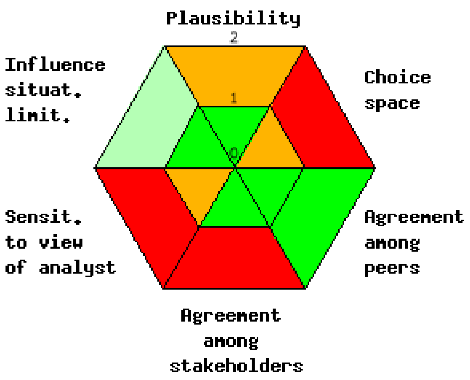
Assumption regarding the delay in the climate system

Figure 4.5. The diagrams with the pedigree score results indicating the potential value-ladenness of the key-assumptions of the indicator ‘change in length of the growth season’ (see for an explanation of the diagram Box 1).

Box 1. Diagrams for the pedigree score results.

The diagrams we use to present the pedigree score results were adapted from the so-called 'kite diagrams' that were introduced by Risbey *et al.* (2001). The diagrams offer a fast and intuitive overview, preserving the underlying information (van der Sluijs *et al.*, 2002b). The diagrams aggregate the scores of the individual experts without averaging them, and in such a way that expert disagreement on the scores is visualised.

One diagram is made for each assumption. The diagram is divided into 6 triangular segments, each segment representing one criterion (Figure 4.4). The scale in each segment is such that zero is in the center of the diagram and two on the border. For each criterion, the area of the corresponding segment from the center of the diagram up to the minimum score given in the group is colored green. If there is no consensus on the score for a given criterion, the area in each segment spanned up between the minimum and the maximum score in the group for that criterion is colored amber. The remaining area (from the maximum score to the outside border of the diagram) -if any- is colored red.



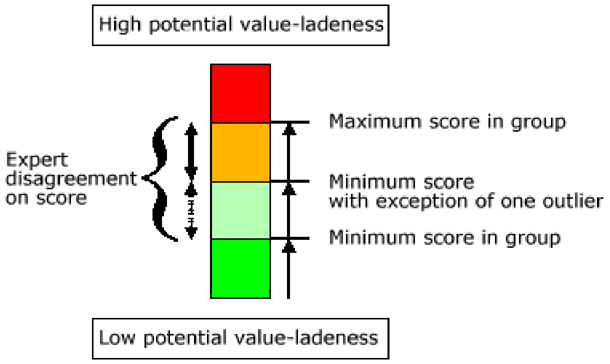


Figure 4.4. Diagrams for the visualisation of pedigree score results. Left: an example diagram. Right: explanation of the colours.

The convention follows a traffic-light analogy and is such that would an assumption on all criteria score 0 unanimously, the entire diagram will be red. If scores are better, more and more green comes into the diagram, whereas expert disagreement on scores is reflected in amber. On the other extreme, if an assumption scores 2 unanimously for all criteria, the entire diagram will be green. The scores for each criterion are such, that in all cases more green in the diagram corresponds to lower potential value-ladenness and more red to higher potential value-ladenness.

A further nuance has been made to account for outliers: in some cases a single outlier score in the group distorts the green area in the diagram. In these cases, a light-green area indicates what the green area would look like if that outlier were omitted.

By looking at the red areas, the extent to which the different types of value-ladenness may have played a role in the production process of the assumption can be assessed. Green areas indicate that the participants think value-ladenness with regard to the criteria at hand played a small role in the production process, red areas that they think value-ladenness played a large role. In case of amber areas it can be concluded that there is disagreement among the participants on these matters.

5 Indicator 2: deaths and hospital admittances due to exposure to ozone

A description of the calculation chain that was used in the EO5 assessment for the indicator 'deaths and hospital admittances due to exposure to ozone' is given in section 5.1. The assumptions in this chain are presented in section 5.2. Section 5.3 lists the assumptions that the workshop participants identified as most important in the chain. Finally, in section 5.4 the results of the pedigree scoring exercise in the workshop are presented.

5.1 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 formed ozone 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 Figure 5.1. In the figure the number of expected deaths and hospital admittances in several age categories in the year 2030 are presented for the scenarios European Coordination (EC) and Global Competition (GC). 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.

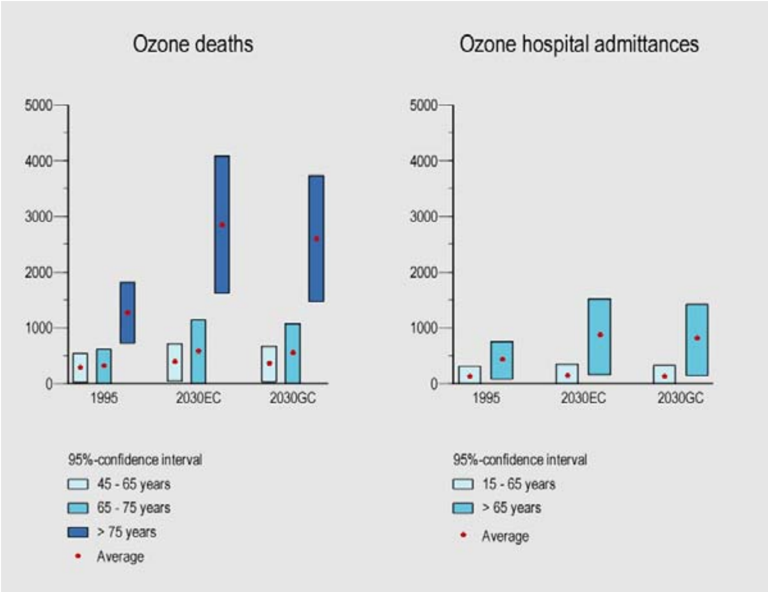


Figure 5.1. EO5-figure indicator 'deaths and hospital admittances due to exposure to ozone' (RIVM, 2000; translation PK).

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) till 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 and 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 5 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[7]). 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 hour maxima concentrations on a grid level of 50 x 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 *et al.*, 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 analyses yielded relative risk (RR) data on deaths and hospital admittances

---

[7] A description of the first version of this model can be found in van Loon (1997).

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).

5.2 Identified assumptions

A 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 25 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 to 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 en 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)
9) Assumption that the sectoral emissions abroad in 2010 to 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 harmonised 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 *et al.*, 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 *et al.*, 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 *et al.*, 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 hour average ozone concentrations (top ozone concentrations are therefore not considered)[8] (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)

5.3 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 following the procedure described in section 3.2.3.

The results of the ranking exercise are presented in Table 5.1. The ranking resulted in 14 key-assumptions, which we have here labelled I till XIV. The ten assumptions that are not mentioned in the table received no points at all: they were not mentioned as one of the 7 most important assumptions by any of the experts and are thus considered to be less important by the group.

5.4 Results pedigree scores

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 (see section 3.2.3). During the workshop there was enough time available to complete the scoring cards of 7 key-assumptions. The participants each individually completed the remaining 7 scoring cards lateron and sent them to the workshop organiser. 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 5.2 the average scores of the group members on the pedigree criteria are listed together with the standard deviations. Also, the mean scores averaged on all 6 pedigree criteria is given. All criteria were weighed equal.

---

[8] This assumption was applied by the epidemiological studies used for the EO5 assessment (Vonk and Schouten, 1998; Hoek *et al.*, 1997).

Assumptions	**Total score ranking exercise**	**Ranking**
Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR	21	I
Assumption that emissions abroad will not differ for the EC and GC scenario	20	II
Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells	12	III
Assumption that the worst case meteorogical circumstances in the current time period will also be worst case meteorological circumstances in the future	10	IV
Assumption that the developments in emission factors and volume growth are harmonised in European context	9	V
Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone	8	VI
Assumption that in case of ozone there is a linear dose-effect relationship	7	VII
Assumption that the ozone concentration is representative for the exposure to ozone	6	VIII
Assumption that the global background concentration of ozone is constant	6	IX
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)	4	X
Assumption that the Dutch epidemiological data are adequate for the whole of the Netherlands	3	XI
Assumption that the sectoral emissions abroad in 2010 to 2030 will experience the same development as the sectoral emissions in the Netherlands in that time period	3	XII
Assumption that there is a direct causal relationship between ozone and death	2	XIII
Assumption that the CPB scenarios were suitable for the societal-demographical developments in the EO5	1	XIV

Table 5.1. Result of the ranking exercise in the workshop. For each assumption the total score of the participants' ranking is indicated and the ranking order of the assumptions is shown (I being the most important).

	Situational limitations		Plausibility		Choice space		Agreement peers		Agreement stakeholders		Sensitivity views analyst		All pedigree criteria		Influence on results	
Assumptions	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 meteorogical 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 to 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

Table 5.2. Average scores (average over the 5 participants) on the pedigree criteria and standard deviations.

Using a diagnostic diagram the weakest links in the chain of assumptions can be identified (see section 2.2). In Figure 5.2 the average pedigree score (averaged on the 6 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:

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

In Figure 5.3 the pedigree score results of the workshop are presented in diagrams.

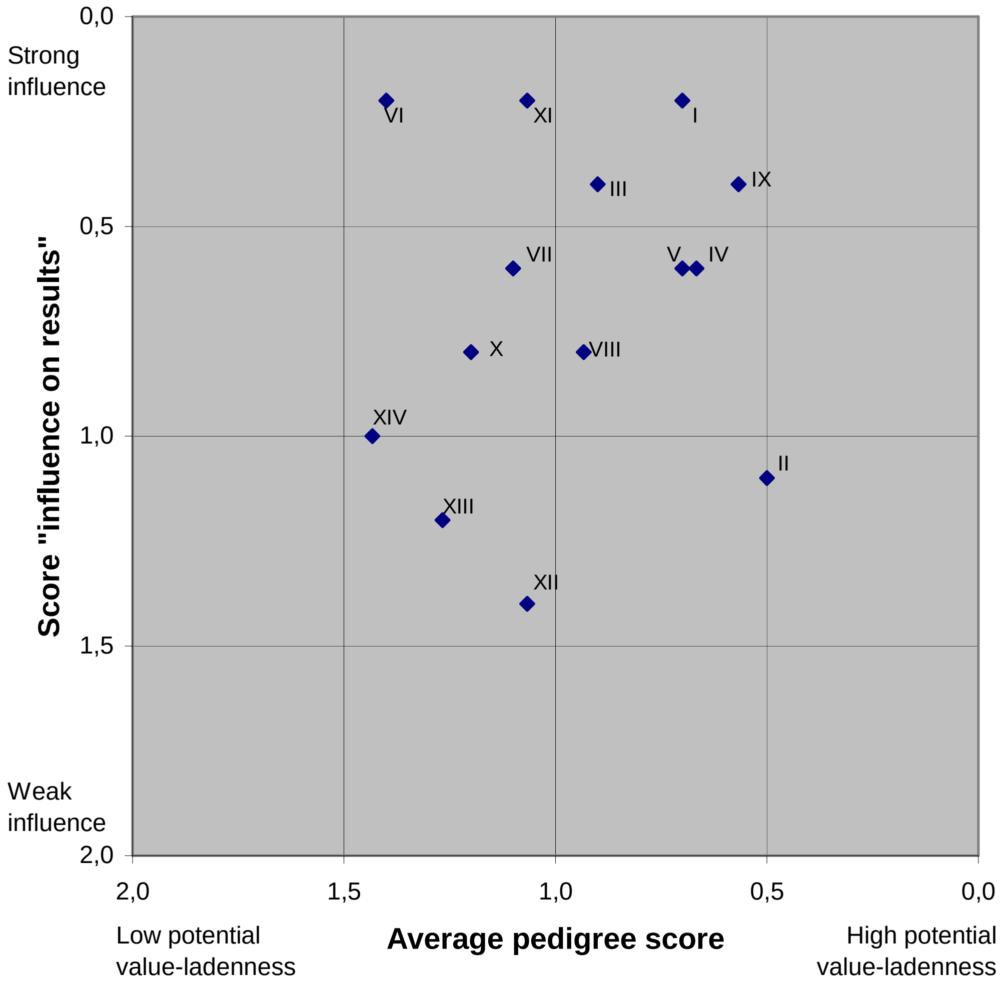
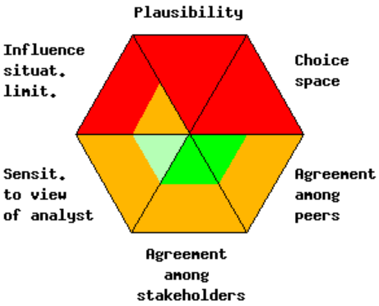
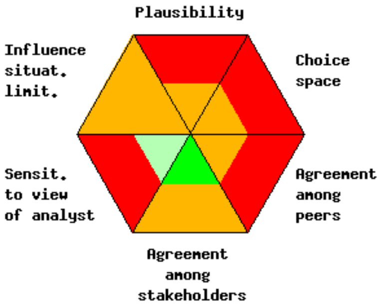


Figure 5.2: A diagnostic diagram of the assumptions of the indicator 'deaths and hospital admittances due to ozone'

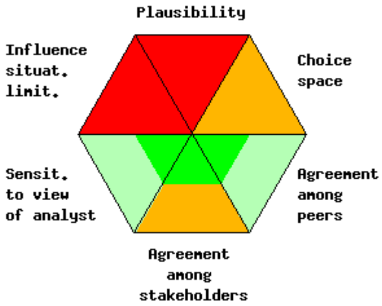
I. Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR
II. Assumption that emissions abroad will not differ for the EC and GC scenario
III. Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells
IV. Assumption that the worst case meteorological circumstances in the current time period will also be worst case meteorological circumstances in the future
V. Assumption that the developments in emission factors and volume growth are harmonised in European context
VI. Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone
VII. Assumption that in case of ozone there is a linear dose-effect relationship
VIII. Assumption that the ozone concentration is representative for the exposure to ozone
IX. Assumption that the global background concentration of ozone is constant
X. 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)
XI. Assumption that the Dutch epidemiological data are adequate for the whole of the Netherlands
XII. Assumption that the sectoral emissions abroad in 2010 to 2030 will experience the same development as the sectoral emissions in the Netherlands in that time period
XIII. Assumption that there is a direct causal relationship between ozone and death
XIV. Assumption that the CPB scenarios were suitable for the societal-demographical developments in the EO5



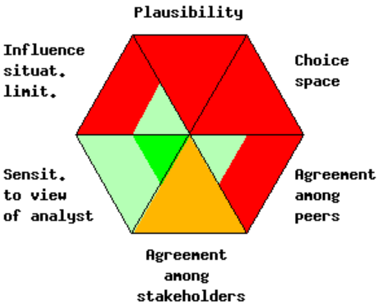
Assumption that the uncertainty in the step from concentration to effect is only determined by the uncertainty in the RR



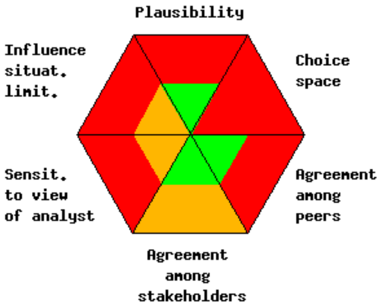
Assumption that emissions abroad will not differ for the EC and GC scenario



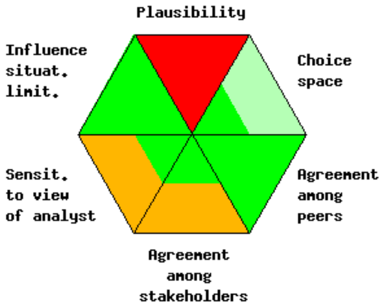
Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells



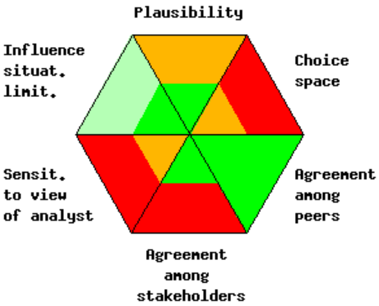
Assumption that the worst case meteorological circumstances in the current time period will also be worst case meteorological circumstances in the future



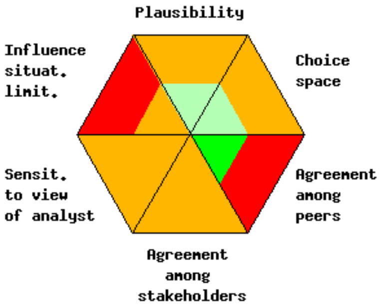
Assumption that the developments in emission factors and volume growth are harmonised in European context



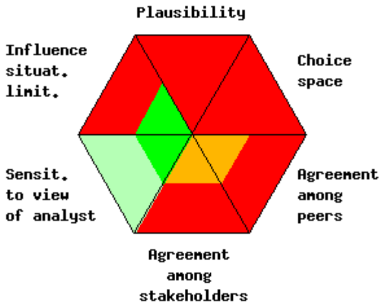
Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone



Assumption that in case of ozone there is a linear dose-effect relationship

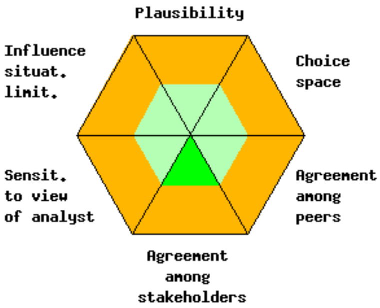


Assumption that the ozone concentration is representative for the exposure to ozone

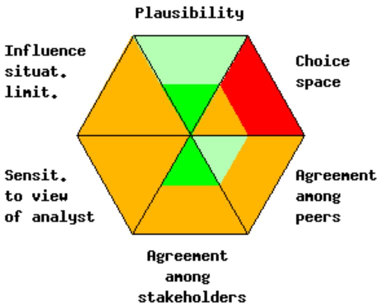


Assumption that the global background concentration of ozone is constant

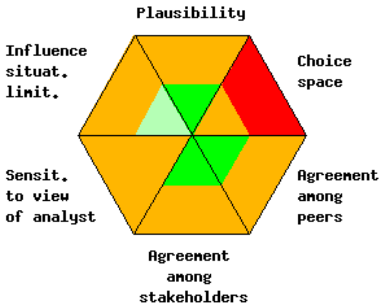
continued on next page



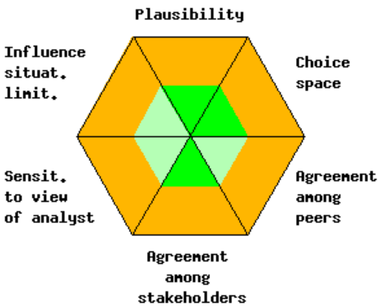
Assumption that the emissions abroad in 2010 will be equal to the national emission ceilings from that year on (established in the framework of the UNECE Convention on Long-range Transboundary Air Pollution)



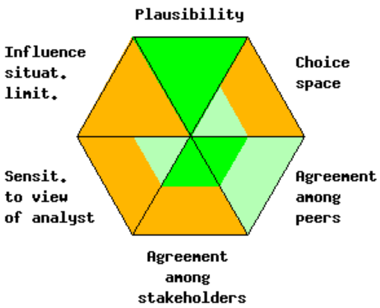
Assumption that the Dutch epidemiological data are adequate for the whole of the Netherlands



Assumption that the sectoral emissions abroad in 2010 to 2030 will experience the same development as the sectoral emissions in the Netherlands in that time period



Assumption that there is a direct causal relationship between ozone and death



Assumption that the CPB scenarios were suitable for the societal-demographical developments in the EO5

Figure 5.3. The diagrams with the pedigree score results indicating the potential value-ladenness of the key-assumptions of the indicator 'deaths and emergency hospital admittances due to exposure to ozone' (see for an explanation of the diagrams Box 1 in section 4.4).

6 Discussion of the test case results and procedures

When comparing the results of the analysis of both indicators presented in sections 4 and 5, two prominent differences can be noticed: the number of key-assumptions identified by the two groups (7 in the growth season group versus 14 in the ozone group), and the fact that there was more expert disagreement in the growth season group than in the ozone group. These, and other differences, can partly be related to differences in composition of the two groups and the procedure followed in both groups. It is therefore necessary to involve aspects of the workshop procedure when discussing and interpreting the results of the two cases. Reasoning the other way around, if different procedures are used, it can be explored what the consequences are of arranging the procedure in a certain way by analysing the differences in results. It must be mentioned here, however, that the number of participants and the number of assumptions processed in the workshop were limited. Therefore conclusions based on comparisons of averages and standard deviations across groups can only be seen as a preliminary indication of a possible effect.

In section 6.1 the results and procedures regarding the key-assumptions in the two calculation chains are discussed. The results and procedure-related issues regarding the pedigree scores are discussed in section 6.2. In section 6.3 results of the evaluation of the workshop are presented.

6.1 Key assumptions

Below, results and procedural aspects are discussed with respect to the number of key-assumptions that were identified in the workshop, and the scores regarding the estimated influence of assumptions on the results of the indicators.

Number of key assumptions

In the ozone group 14 key-assumptions were selected and reviewed, whereas the ranking exercise in the growth season group resulted in a selection of 7 key-assumptions. The difference is caused by a difference in procedure followed by the two groups. The growth season group selected the 7 most important assumptions from the list in a plenary discussion. After selecting the 7 most important ones, the ranking stopped. In case of the ozone group, each member picked 7 assumptions, after which a group ranking was established. For the group ranking, each assumption that was at least mentioned once by one of the experts was included in the list of key-assumptions. It seems preferable to not set a limit on the number of key-assumptions. This will offer better insight in the importance of assumptions throughout the calculation chain (as opposed to picking the *n* most important ones).

Scores regarding influence on results

Due to the ranking procedures followed in both groups, the aspect ‘influence on results’ was discussed more extensively in the growth season group. When looking at the average standard deviation of the scores on ‘influence on results’ (averaged over all key-assumptions), the standard deviation is higher in the ozone group then in the growth season group (0.6 and 0.4 respectively). There may be an effect that an extensive group discussion may lead to consensus formation in the group.

In case of the growth season indicator, the average scores on the influence of the key-assumptions filled in on the scoring cards is consistent with the scores given during the ranking exercise (see Table 4.2 and Table 4.1). This is less the case in the ozone group (see Table 5.2 and Table 5.1). This can be explained by the fact that the facilitator of the ozone group instructed the participants to choose the most important assumptions based on either influence on the indicator or on policy relevance or both. Because policy relevance was included as a selection criterion, some assumptions have a high ranking despite a relatively low estimated influence on the results (namely ‘Assumption that emissions abroad will not differ for the EC and GC’ and ‘Assumption that the ozone concentration is homogeneously distributed over the EUROS grid cells’) and some assumptions have a low ranking despite a relatively high estimated influence on the results (namely ‘Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR’ and ‘Assumption that the Dutch epidemiological data are adequate for the Dutch situation’). A second explanation for the differences is that the view of the participants on the influence may change during the workshop due to the discussions on the assumptions.

6.2 Pedigree scores

Spread in the scores

When comparing the results of the two groups it is evident that there is more spread in the scores of the members of the growth season group; the diagrams of the growth season assumptions in general reflect a larger degree of expert disagreement (see Table 4.2 and Figure 4.5). There are a few possible explanations for this. First, the growth season group had more participants than the ozone group. Hence, chances of finding diverse scores in the group are higher. A second explanation can be found in the procedure: in the ozone group the criteria were discussed one by one, after which the participants immediately filled in the score on the scoring card. In case of the growth season group, discussion took place earlier in the workshop when discussing the importance of each of the assumptions during ranking. The criteria in this case were not systematically and in some cases not explicitly addressed in the group. The pedigree cards were later on filled in individually, without further discussion. This second explanation is backed by the fact that this difference in procedure also leads to more or less differing scores in the ozone group. 7 Key-assumptions were discussed in the group;

the remaining 7 were scored individually without a group discussion, due to lack of time. The diagrams of the latter assumptions also show a larger degree of expert disagreement. It is preferable to let the discussion and the filling in of the cards be preceded by a focussed group discussion addressing each of the criteria; in this way all group members have the opportunity to argue why a certain criterion of an assumption should receive a high or a low score. After the discussion all participants can fill in the cards based on the same information. It is however important that the facilitator makes sure that the discussions among the participants do not slide off to a quick group consensus, but that there is an open discussion promoting critical review.

Diversity in scores can also be related to the scale used. If, for instance, to the participants' opinion the score for a certain criterion lies between 0 and 1, one participant may choose 0, and another may choose 1. While, had there been a score 0.5, both participants would have chosen 0.5. In retrospect, we therefore recommend to use a five point scale (0 – 0.5 – 1 – 1.5 – 2.0 – 2.5 – 3), instead of a three point scale.

Weighing of the criteria

In order to find the weakest links in the chain of assumptions, we plotted the average pedigree score against the score on influence on results. In determining the average pedigree score all 6 pedigree criteria were weighed equal. The 6 criteria bring forward different aspects of value-ladenness. Since to our opinion a hierarchy in these criteria is not the case, we did not apply different weighing factors.

6.3 Evaluation of the workshop

At the end of the workshop, the participants were asked to give a first reaction on the workshop, regarding both procedure and content. They were also asked to fill in an evaluation form with questions regarding the set up of the workshop, the information supply, their experiences with the identification of assumptions and the ranking of assumptions, the filling in of the pedigree cards and the usefulness/applicability of the method demonstrated in the workshop. The evaluation form was filled in by 9 participants (6 of the ozone and 3 of the growth season group). The main issues that arised from the evaluation are mentioned below.

The participants of the ozone group thought the group was too small, which is not surprising, since three participants of the ozone group at the last minute were unable to attend the workshop. Because of this, also expertise on some parts of the chain was missing. Several participants of the growth season group indicated that the information provided to them regarding the calculation chain was not sufficient.

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 opinion with respect to the difficulty of the ranking exercise was diverse in the growth season group. In the ozone group all respondents thought it was 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). In both groups, however, several times it was indicated that the criterion ‘agreement among stakeholders’ is hard to assess. Sometimes the participants had no idea what score to give, in other cases the answer seemed to depend significantly on what stakeholders one took in mind. The RIVM/MNP Guidance for Uncertainty Assessment and Communication (van der Sluijs *et al.*, 2003) has an extensive section with respect to stakeholder views. This could be used to identify stakeholders and to think about alternative problem definitions the different stakeholders may have with regard to the issue of the assessment. 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. According to several respondents the analysis yielded unexpected insights; 3 respondents of the ozone group thought the analysis did not find the results surprising.

In general, the method is thought to be suitable for application in the RIVM practice. The method is thought to be a useful extension of the tool catalogue that is part of the RIVM/MNP Guidance for Uncertainty Assessment and Communication (van der Sluijs *et al.*, 2004).

7 A method for the analysis of assumptions

In the previous sections, we presented a method for identifying the key-assumptions in calculation/modelling chains and for assessing the potential value-ladenness of these assumptions.

Based on the results of this value-ladenness assessment, further action can be taken. The potential value-ladenness of the key-assumptions can be examined in more detail. Next, assumptions in the chain can be adjusted based on the findings or a sensitivity analysis can be added to the assessment. Finally, it is important to communicate the key-assumptions and their background as well as their implications in terms of robustness of results to the users of the assessment.

Adding these extra steps, our method for the analysis of assumptions in assessments contains 7 steps:

ANALYSIS

1. Identify explicit and implicit assumptions in the calculation chain
2. Identify and prioritise key-assumptions in the chain
3. Assess the potential value-ladenness of the key-assumptions
4. Identify 'weak' links in the calculation chain
5. Further analyse potential value-ladenness of the key-assumptions

REVISION

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

COMMUNI-CATION

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

This method can be applied by the analysts carrying out the assessment. However, each analyst has limited knowledge and perspectives with regard to the assessment topic, and in consequence will have some 'blind spots'. Therefore preferably other analysts (peers) are involved in the method as well, as in the two cases was done in the form of a workshop. Stakeholders, with their specific views and knowledge, can be involved as well. The persons taking part in the review of the assumptions in an assessment (analysts, peers and/or stakeholders) are referred to below as ‘participants’.

In the two test-cases analysis of the pedigree scoring results of the expert workshop was done without the participants. If facilities are available to calculate and present results immediately after the participants have filled in their pedigree score cards (and if enough time is available) the group can continue with the rest of the steps. Honingh (2004) applied the method for review of assumptions in her study on uncertainties in a policy oriented model for the calculation of agricultural ammonia emissions in the Netherlands. The workshop took place in the Policy Lab of Utrecht University, which is a meeting room with a number of linked computers. The data that were entered by the participants were immediately subjected to calculations, and the results were presented and discussed on the spot.

All steps of the method will be elaborated on below.

Step 1: Identify explicit and implicit assumptions in the calculation chain

In the first step explicit and implicit assumptions in the calculation chain are identified. To identify implicit assumptions, it is important to constantly consider whether a certain assumption implies other assumptions.

Preferably, analysts construct a list of assumptions while doing the assessment. If done *ex post*, analysts can identify assumptions by systematically mapping and reconstructing the calculation chain based on documentation analysis, interviews and critical review. Next, the list of assumptions composed by the analysts can be checked by peers and/or stakeholders. In our two cases the list was checked and completed during an expert workshop.

From the calculation chains of the two indicators analysed in this study, it is clear that assumptions can be quite diverse with respect to aggregation level. An assumption can refer to a specific detail in the chain (e.g., ‘Assumption on the climate sensitivity’), as well as to a cluster of assumptions on a part of the chain (‘Assumptions regarding the carbon cycle’). The test cases show that assumptions formulated on both levels can be usefully analysed, provided that the participants share the same meaning with respect to the assumptions as formulated in the list. Therefore a good description of the chain is needed, next to a good facilitated discussion on the assumptions.

Step 2: Identify and prioritise key-assumptions in the chain

Due to the fact that the time for an analysis of potential value-ladenness is always limited and due to the fact that not all value-laden 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.

The assumptions identified in step 1 are prioritised by taking into account the influence of the assumptions on the outcomes of interest of the assessment. Ideally, this selection is based on a comprehensive sensitivity analysis. Since such an analysis will often not be attainable, the participants can be asked to indicate the estimated influence of the assumptions on outcomes of interest of the assessment. An expert elicitation technique can be used in which the participants bring forward their opinions and argumentation on whether an assumption is of high or low influence on the outcome. 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. This procedure was demonstrated in the cases.

Step 3: Assess the potential value-ladenness of the key-assumptions

In this study we developed and used a pedigree matrix to assess the potential value-ladenness of assumptions. The matrix was presented in section 2.2 (see Table 2.2). For all key-assumptions pedigree scoring cards are filled in individually by the participants, informed by a group discussion (see Appendix B for an example of a scoring card (in Dutch)). The order in which the key-assumptions are discussed is determined by the ranking established in step 2 of the method, starting with the assumption with the highest rank.

Here, again a group discussion takes place first, in order for the participants to remedy each others blind spots and 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.

The facilitator should stress that the scores 0 and 2 should not be seen as extremes or ideal types. In the description of the modes an attempt has already been made to make sure that the formulation is not too extreme (e.g., 'few would have made the same assumption' was used, in stead of 'nobody would have made the same assumption'). It is important that the participants feel confident to use all three modes of the scale. If the 0 and 2 score are presented as two extremes, the tendency will be to only use the '1' score. If this is the case, the differences in scores between criteria of different assumptions will be lost. As discussed in section 6.2, in retrospect we recommend using a five point scale instead of a three point scale (0 – 0.5 – 1 – 1.5 – 2.0 – 2.5 – 3).

For the presentation of the results of the pedigree scoring exercise, diagrams can be used that combine information on the scores that were given by the participants and the degree to which there was group (dis)agreement (see section 4.4). The set-up of the diagrams enables a quick intuitive overview of the potential value-ladenness of assumptions, following a traffic light analogy.

Step 4: Identify 'weak' links in the calculation chain

The pedigree matrix 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 influence on the results of the assessment can be viewed as problematic weak links in the calculation chain. As described in section 2.2 a diagnostic diagram can be used to identify the weak links. In a diagnostic diagram the average score on the pedigree criteria is plotted against the score on influence of results. The key-assumptions can be prioritized according to their position in the spectrum from danger zone to safe zone, starting in the upper right corner. The assumptions that lie most in or towards the danger zone can be viewed as the most problematic assumptions in the calculation chain.

Preferably, error bars are used in the diagnostic diagram for each assumption, as was done in van der Sluijs *et al.* (2002b). In our two cases, however, the number of participants was too limited.

Step 5: Further analyse potential value-ladenness of the key-assumptions

In step 5, the nature and extent of the potential value-ladenness of the individual key-assumptions is further explored. Based on inspection of the diagrams visualising the pedigree scores (or based on the table of pedigree scores) it can be analysed:

- what types of value-ladenness possibly play a role and to what extent
- to what extent there is disagreement on the pedigree scores among the participants
- whether changing assumptions is feasible and desirable

<u>Types of value-ladenness</u>

When analysing the types of value-ladenness possibly playing a role in the assumptions, primarily the red areas are of interest. These areas indicate that all participants agree that there is no higher pedigree score. Red areas can therefore be interpreted as flags for potential value-ladenness. The larger the red area, the higher the potential value-ladenness of the assumption on the pedigree criterion at hand, according to the participants. The pedigree criteria that contain red areas offer insight in why value-ladenness potentially comes into play in the assumption at hand: the analyst had to revert to an implausible assumption, peers and stakeholders disagree on what the assumption should be, there is a large choice space, situational limitations largely determine the choice for the assumption and/or the choice for the assumption is sensitive to the view and interests of the analyst. The pedigree matrix (Table 2.2) shows what types of value-ladenness are related to the pedigree criteria.

Green areas indicate that the participants think value-ladenness with regard to the criteria at hand played a small role in the production process.

Extent of disagreement

Amber areas in the diagrams indicate that the participants do not agree on the score on the pedigree criterion at hand. They therefore do not agree on the extent of potential value-ladenness of the assumption.

If there are many participants, chances are higher that more diverging scores are used then in a small group. The diagrams composed of the results of a large group therefore tend to show amber surfaces. In case of disagreement in a group, histograms can provide extra information. They show the general ‘direction’ of the group; whether the group is evenly distributed over the 3 scores, or whether there is a general tendency to score low or to score high. In this way for the amber areas as well, it can be explored how the group feels about the potential value-ladenness of these criteria.

Changes in assumptions that are feasible and desirable

Potential motives for reconsidering assumptions in the assessment are:

- the assumption is fictive or speculative (score 0/red area in the criterion on plausibility)
- few peers would have made the same assumption (score 0/red area in the criterion on agreement among peers)
- few stakeholders would have made the same assumption (score 0/red area in the criterion on agreement among stakeholders)

If there are ‘hardly any alternative assumptions available’ (score 2 on the criterion choice space), the possibilities for changes in the assumptions are very limited. If not, it is useful to reconsider the existing alternatives for the assumption and possibly generate new alternatives.

If changes seem impossible due to a score 0 on the criterion influence on situational limitations (‘totally different assumption had there not been limitations’) it can be explored whether these limitations can be lifted by investing, for instance, in new tools, extending the deadline for the assessment, etcetera.

If the choice for an assumption is sensitive to the view and interests of the analyst (score 0 on the criterion sensitivity view and interests analyst) one can ‘check’ whether the choice that was made is well underpinned. If not, here as well changes in the assumption can be considered.

The key-assumptions that are positioned most in or near the danger zone of the diagnostic diagram should receive the most attention in the analysis in this step of the method.

Step 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.

Besides a sensitivity analysis, specific assumptions can be revised or diversified. In the case of revising an assumption, the assumption is replaced by a different assumption. In some cases however, it will be difficult or undesirable 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 several assessments are formed, with differing outcomes, depending on what assumptions are chosen.

If several assumptions in the chain are diversified, 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 outcomes of interest in the assessment.

In section 2.1 we presented four phases that can be distinguished in choice processes:

1. Determining that an issue requires choices to be taken
2. Making an inventory of options to choose from
3. Choosing an option
4. Checking/evaluating the choice

If it is decided, based on the analysis in step 5, that assumptions in the assessment might have to be changed, the phases of this choice process will be applied (again). In case of implicit assumptions, it was not determined previously that the issue required choices to be made. Hence, the other phases were not entered. Therefore, if an implicit assumption is revised, an inventory of options will have to be made, from which one (or, in the case of diversification, several) alternatives are chosen.

If the assumption in the initial assessment did go through the choice process phases, an inventory has already been made of alternatives. However, the knowledge and perspectives of the analysts who made the 'list' may again have influenced which alternatives were explored. Involving peers and stakeholders in a dialogue process of extended peer review will help to come to a broader exploration of options.

The same goes for the next phase: choosing an option. In this phase the pros and cons of the different alternatives are evaluated. Based on a comparison, one or more alternatives are chosen. Finally, the choice can be checked and evaluated. If the consequences of the chosen alternative are not satisfactory (for instance, because the assumption requires an implicit assumption that is unacceptable) a new option can be chosen.

When assumptions are changed the pedigree scores on the criteria plausibility, agreement among peers, agreement among stakeholders and influence of situational limitations may change. However, the goal of changing assumptions is not to increase the pedigree scores. Nor is the goal to reach consensus among the analysts, peers and stakeholders involved in this process. What this method intends to do is pinpoint important assumptions in the chain where value-ladenness may play a role and to assess the relevance of these assumptions in view of the outcomes of the assessment. Because of their importance for the outcome of the assessment they are analysed in detail, after which the choice for the assumption may be reconsidered. The assumption analysis enables the analysts to make a conscious, well-underpinned, transparent choice, and pinpoints the issues in the chain that are important to communicate to the audience of the assessment report (see step 7).

Step 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 step 1 to 6 form the basis for this. The key-assumptions were identified and prioritised in step 2. The weak links in the chain were identified and analysed in step 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. One can, for instance, be explicit about the fact that limitations played an important role in setting an assumption or be explicit about why this assumption was chosen, despite the fact that few stakeholders and peers would have made the same choice.

Often, the limited space available for the presentation of assessment results will not allow for a detailed account on the issues mentioned above. In this case the principle of progressive disclosure of information (Guimarães Pereira and Corral Quintana, 2002) can be applied: a ‘layered’ approach, in which interested peers and stakeholders are offered the possibility to access more information in, for instance, background reports or on the internet.

Appendix C offers a condensed overview of the seven steps of the method.

8 Discussion and conclusions

In this study we developed a method to systematically identify, prioritise and review assumptions to assess the potential value-ladenness of important assumptions and to deal with these potentially value-laden assumptions in an explicit and transparent manner. We demonstrated and tested part of the proposed method by applying it to the assumptions in the calculation chains of two EO5 indicators.

The following research questions were central to the design of the method:

- How can assumptions in a calculation chain be mapped and prioritised?
- How can potential value-ladenness of assumptions be analysed?
- How can assumptions be handled, based on the inventory and prioritisation of assumptions and the analysis of potential value-ladenness?
- How can the potential value-ladenness of an assessment be communicated?

The method we developed starts with an inventory of explicit and implicit assumptions which are collected by examining all steps that are taken in a calculation chain. These assumptions are then ranked according to their estimated importance for the assessment. Core of the method is the review of the potential for practical, epistemic, disciplinary and socio-political value-ladenness in the key-assumptions. Using the pedigree criteria plausibility, agreement among peers, agreement among stakeholders, choice space, influence situational limitations, and sensitivity to view and interests of the analyst the room for value-ladenness of an assumption is 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. Changes can be made to the assessment based on this analysis, by extending the assessment with an sensitivity analysis regarding crucial assumptions, by replacing assumptions by other assumptions or by diversifying assumptions. Next, based on the analysis, important key-assumptions and their implications for the robustness of the outcomes of interest of the assessment can be communicated to peers and to users of the assessment results. The method thus contains an analysis part, a revision part and a communication part. Appendix C gives an overview of the seven steps of the method.

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 revisions based on the analysis will no longer be 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. Application during the assessment is preferable, since an iterative treatment of assumptions can improve the assessment.

The time required for this method is variable. Firstly, it depends on the number of calculation chains in the assessment that are analysed and on the complexity of the (models in) the chain.

Secondly, the method can be applied by the analysts carrying out the assessment alone or can be applied by the analysts together with peers and/or stakeholders. Involving the latter is preferable, since each analyst has limited knowledge and perspectives with regard to the assessment topic, and in consequence will have some ‘blind spots’. Involving peers and/or stakeholders will, however, require more time and resources.

The cases showed that the results of the method can be sensitive to the composition of the group of participants (both the number of persons and the persons’ backgrounds).

The results can also be sensitive to procedure details as determined by the group facilitator. Further research and testing of the method can provide more insight into procedural changes and its effect on the method’s results, into how to involve stakeholders in the application of this method, and into how the method can be applied in case of complex assessments with many calculation chains.

This method we designed incorporates the main elements for dealing with value-laden assumptions that are mentioned in literature (among others by Keepin and Wynne, 1984; Saltelli *et al.*, 2000; Schneider, 1997): transparency, insight in the robustness and sensitivity of results, and diversification of assumptions. When stakeholders are involved in the analysis of the assumptions, the method can also be part of an extended peer review process (Funtowicz and Ravetz, 1993). However, due to lack of expertise it may be difficult for stakeholders to score on the pedigree criteria of assumptions of a highly technical nature. A recommendation to deal with this problem was made during a workshop assessing scientific studies on possible health effects from waste incineration using the pedigree concept (Craye *et al.*, 2005): stakeholders can select an expert that they trust. This expert can participate in the process, and represent the stakeholders’ views.

We would like to stress that the method deals with *potential* value-ladenness, which should not be confused with actual value-ladenness. Assessing the actual value-ladenness of assumptions is impossible, since it would require exact and detailed knowledge on what factors contributed to what extent to the analysts’ choices. This method is also not intended for decreasing the (potential) value-ladenness of assumptions, nor for reaching consensus among the analysts, peers and stakeholders involved in this process. The proposed method helps raise awareness on assumptions that are made in assessments, helps to systematically reflect on them, to reconsider choices that were made, as well as shape communication on crucial assumptions. It thus enables a well-structured discussion on potentially value-laden assumptions among scientists and stakeholders. In this discussion not only the politically controversial assumptions are addressed (as is often the case when assessment results are discussed in public), but also other assumptions that turn out to be important for assessment results. Another novel aspect of this approach is, that it acknowledges that also pragmatic factors may play a role in the colouring of assumptions. The method has been added to the tool catalogue of the RIVM/MNP Guidance for Uncertainty Assessment and Communication (van der Sluijs *et al.*, 2004).

Acknowledgements

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 report.

References

Alcamo J, Leemans R and Kreileman E (eds.) (1998). Global Change Scenarios of the 21st Century. Results from the IMAGE 2.1 Model. London, United Kingdom: Pergamon.

Asselt MBA van (2000). Perspectives on uncertainty and risk: the PRIMA approach to decision support. Dordrecht: Kluwer.

Asselt MBA van, Langendonck R, Asten F van, Giessen A van der, Janssen PHM, Heuberger PSC and Geuskens I (2001). Uncertainty and RIVM's environmental outlooks: documenting a learning process. Maastricht, The Netherlands: ICIS/RIVM.

Blom JG and Roemer MGM (1997). Description of the 3D LOTOS model. Part I: Dynamics. Amsterdam, The Netherlands: CWI. Report number MAS-N9701.

CBS and CPB (1997). Bevolking en arbeidsaanbod: drie scenario's tot 2020'. The Hague, The Netherlands: Sdu uitgevers.

CPB (1997). Economie en fysieke omgeving. The Hague, The Netherlands: Sdu uitgevers.

CPB (1999). Worldscan, the core version. The Hague, The Netherlands:CPB.

Craye M, Sluijs J van der and Funtowicz S (2005). A reflexive approach to dealing with uncertainties in environmental health risk science and policy. International Journal for Risk Assessment and Management 5 (2): 216-236

Eerens HC and Dam JD van (eds.) (2001). Grootschalige luchtverontreiniging en depositie in de Nationale Milieuverkenning 5. Bilthoven, The Netherlands: RIVM. Report number 408129016.

Funtowicz SO and Ravetz JR (1990). Uncertainty and quality in science for policy. The Netherlands: Kluwer Academic Publishers.

Funtowicz SO and Ravetz JR (1993). Science for the Post-Normal Age. Futures 25: 735-755.

Guimarães Pereira A and Corral Quintana S (2002). From technocratic to participatory decision support systems: responding to the new governance initiatives. Journal of Geographic Information and Decision Analysis 6(2):95-107.

Harremoës P, Gee D, MacGarvin M, Stirling A, Keys J, Wynne B and Guedes Vaz S (eds.) (2002). The Precautionary Principle in the 20th century. Late Lessons from early warnings. London: EARTHSCAN.

Hoek G, Verhoeff A and Fischer P (1997). Daily mortality and air pollution in the Netherlands, 1986-1994. Wageningen, The Netherlands: Landbouwuniversiteit Wageningen. Report number 1997-481.

Honingh N (2004). Onzekerheden in het Mest- en Ammoniakmodel. Utrecht: Copernicus Institute for Sustainable Development and Innovation, Universiteit Utrecht. Report number NW&S-I-2004-20.

Hornberger GM and Spear RC (1981). An approach to the preliminary analysis of environmental systems. Journal of Environmental management 12: 7-18.

Huesemann MH (2002). The inherent biases in environmental research and their effects on public policy. Futures 34: 621-633.

Janssen PHM, Petersen AC, Sluijs JP van der, Risbey JS and Ravetz JR (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Quickscan Hints & Actions List. Bilthoven, The Netherlands: RIVM/MNP. ISBN 90-6960-105-2.

Jasanoff S (1990). The Fifth Branch, Science Advisers as Policy Makers. Harvard: Harvard University Press.

Jasanoff S and Wynne B (1998). Science and decisionmaking. In: Rayner S and Malone EL (eds.). Human Choice and Climate Change 1: 1-87. Columbus, OH: Battelle Press.

Keepin B and Wynne B (1984). Technical analysis of IIASA energy scenarios. Nature 312: 691-695.

Kloprogge P and Sluijs JP van der (2002). Choice processes in Modelling. In: Rizzoli, AE and Jakeman AJ (eds.). Integrated Assessment and Decision Support, Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society. Manno, Switzerland: iEMSs, 96-101.

Lindfors LG, Christiansen K, Hoffmann L, Virtanen Y, Juntilla V, Hanssen OJ, Rønning A, Ekvall T and Finnveden G (1995). Copenhagen: Nordic Council of Ministers. LCA-Nordic Technical Reports No. 1 – 9, TemaNord 1995:502.

Loon M van (1996). Numerical Methods in Smog Prediction. PhD thesis. Amsterdam, The Netherlands: University of Amsterdam.

Nakicenovic N. and Swart R (eds.) (2000). Special Report on Emission Scenarios. Special Report of the Intergovernmental Panel on Climate Change. Cambridge, United Kingdom: Cambridge University Press.

Oreskes N, Shrader-Frechette K and Belitz K (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263: 641-646.

Petersen AC, Janssen PHM, Sluijs JP van der, Risbey JS and Ravetz JR (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Mini-Checklist & Quickscan Questionairre. Bilthoven, The Netherlands: RIVM/MNP. ISBN 90-6960-105-1.

Refsgaard, JC (ed.) (2002). State-of-the-Art Report on Quality Assurance in modelling related to river basin management, HarmoniQuA-report, D-WP1-1 (http://harmoniqua.wau.nl/public/Reports/SOA%20chapters/SOA.pdf)

Risbey JS, Sluijs JP van der and Ravetz JR (2001). A Protocol for Assessment of Uncertainty and Strength of Emissions Data. Utrecht, The Netherlands: Department of Science, Technology and Society, Utrecht University. Report number NWS-E-2001-10.

Risbey JS, Sluijs JP van der, Ravetz JR and Janssen PHM (2001). A Checklist for Quality Assistance in Environmental Modelling. Utrecht, The Netherlands: Department of Science, Technology and Society, Utrecht University. Report number NWS-E-2001-11.

RIVM (2000). Nationale Milieuverkennig 5 2000-2030. Bilthoven, The Netherlands: Rijksinstituut voor Volksgezondheid en Milieu.

Saltelli A, Chan K and Scott EM (eds.) (2000). Sensitivity Analysis. John Wiley & Sons Publishers. Probability and Statistics Series.

Saltelli, A (2002). Sensitivity analysis for importance assessment. Risk Analysis 22 (3): 579-590.

Schneider SH (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.

Sluijs JP van der (1997). Anchoring amid uncertainty; On the management of uncertainties in risk assessment of anthropogenic climate change. PhD thesis. Utrecht, The Netherlands: Utrecht University.

Sluijs JP van der (2002). A way out of the credibility crisis around model-use in Integrated Environmental Assessment. Futures 34: 133-146.

Sluijs JP van der, Eijndhoven JCM van, Shackley S and Wynne B (1998). Anchoring devices in science for policy, the case of consensus around climate sensitivity. Social Studies of Science 28 (2): 291-324.

Sluijs JP van der, Risbey JS and Ravetz JR (2002a). Uncertainty Assessment of VOC emissions from Paint in the Netherlands. Utrecht, The Netherlands: Utrecht University, Copernicus Institute/Dept. of Science, Technology and Society.

Sluijs J van der, Potting J, Risbey J, Vuuren D van, Vries B de, Beusen A, Heuberger P, Corral Quintana S, Funtowicz S, Kloprogge P, Nuijten D, Petersen A and Ravetz R (2002b). Uncertainty assessment of the IMAGE/TIMER B1 CO2 emissions scenario, using the NUSAP method. The Netherlands: Dutch National Research Program on Climate Change. Report number 410200104.

Sluijs JP van der, Risbey JS, Kloprogge P, Ravetz JR, Funtowicz SO, Corral Quintana S, Guimarães Pereira A, De Marchi B, Petersen AC, Janssen PHM, Hoppe R and Huijs SWF (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Detailed Guidance. Utrecht, The Netherlands: Utrecht University, Copernicus Institute/Dept. of Science, Technology and Society and RIVM.

Sluijs JP van der, Janssen PHM, Petersen AC, Kloprogge P, Risbey JS, Tuinstra W, Asselt MBA van and Ravetz JR (2004). RIVM/MNP Guidance for Uncertainty Assessment and Communication: Tool Catalogue for Uncertainty Assessment. Utrecht, The

Netherlands: Utrecht University, Copernicus Institute/Dept. of Science, Technology and Society and RIVM.

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

Stirling A (1999). On science and precaution in the management of technological risk: Volume I - a synthesis report of case studies. Seville: European Commission Institute for Prospective Technological Studies. Report number EUR 19056 EN.

Stirling A (2001). Inclusive deliberation and scientific expertise: precaution, diversity and transparency in the governance of risk. PLA Notes 40: 67-71.

STOWA/RIZA (1999). Good Modelling Practice Handbook. The Netherlands: STOWA and Dutch Department of Public Works, Institute for Inland Water Management and Waste Water Treatment. Report number 99-05 and 99.036.

UNESCO COMEST (2005). The Precautionary Principle. Paris: UNESCO.

Vonk JM and JP Schouten (1998). Daily emergency hospital admissions and air pollution in the Netherlands, 1982-1986 and 1986-1995. Groningen, The Netherlands: Rijksuniversiteit Groningen. Report number VROM 96.140072.

Walker WE, Harremoës P, Rotmans J, Sluijs J P van der, Asselt MBA van, Janssen P and Krayer von Krauss MP (2003). Defining Uncertainty; A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. Integrated Assessment 4(1): 5-17.

Wee GP van, Kuijpers-Linde MAJ and Gerwen OJ van (2001). Emissies en kosten tot 2030 bij het vastgesteld milieubeleid. Achtergronddocument bij de Nationale Milieuverkenning 5. Bilthoven, The Netherlands: RIVM. Report number 408129013

Appendix A: List of interviewees and workshop participants

Interviewees

Length of growth season case study:

Jan Bakkes (MNP/LED)

Johannes Bollen (MNP/KMD)

Lex Bouwman (MNP/KMD)

Bas Eickhout (MNP/KMD)

Rik Leemans (MNP/KMD)

Ozone case study:

Jeannette Beck (MNP/LED)

Guus de Hollander (RIVM/VGZ/VTV)

Marten Marra (RIVM/MEV/MGO)

Workshop participants

Length of growth season case study:

Arthur Beusen (MNP/IMP; notetaker)

Lex Bouwman (MNP/KMD)

Rob van Dorland (KNMI)

Rik Leemans (MNP/KMD)

Bert de Vries (MNP/KMD)

Arthur Petersen (MNP/IMP; facilitator)

Koos Verbeek (KNMI)

Hans Visser (MNP/IMP)

Ozone case study:

Leendert van Bree (MNP/LOK)

Peter Builtjes (TNO-MEP)

Paul Fischer (RIVM/MEV/MGO)

Peter Janssen (MNP/IMP; notetaker)

Penny Kloprogge (Utrecht University; notetaker)

Jeroen van der Sluijs (Utrecht University; facilitator)

Appendix B: Example of a pedigree scoring card

Assumption: Assumption that changes in the composition of the air pollution mix will not lead to changes in the RR for ozone.

Professional expertise with respect to the assumption: little – some – considerable

Involvement in this step of the EO5 analysis: little – some – considerable

		0	1	2		Explanation/argumentation
Plausibility	*fictive*				*plausible*	
Agreement among peers	*few*				*many*	
Agreement among stakeholders	*few*				*many*	
Choice space	*ample choice*				*hardly any alternatives*	
Influence situational limitations	*a lot of influence*				*hardly any influence*	
Sensitivity to view and interests of the analyst	*sensitive*				*hardly sensitive*	
Estimated influence on the results	*determines results indicator to a large extent*				*local influence*	

Appendix C: Overview of the method for the analysis of assumptions in assessments

ANALYSIS

1. Identify explicit and implicit assumptions in the calculation chain
   - *systematically map the explicit and implicit assumptions*
   - *the list is preferably checked and completed by peers and/or stakeholders*
2. Identify and prioritise key-assumptions in the chain
   - *estimate the influence of each assumption on outcomes of interest of the assessment*
   - *rank the assumptions according to the estimated influence*
   - *if done by more than 1 person, establish a group ranking*
3. Assess the potential value-ladenness of the key-assumptions
   - *for each assumption (or for the ones ranked highest in step 2) fill in the scores on the pedigree criteria*
   - *if done by more than 1 person, calculate averages for the scores on the pedigree criteria*
4. Identify 'weak' links in the calculation chain
   - *position the assumptions in a diagnostic diagram (average score on the pedigree criteria plotted against the score on influence on results)*
   - *identify the assumptions positioned in the upper right corner of the diagram*
5. Further analyse potential value-ladenness of the key-assumptions
   - *if the pedigree criteria were scored by more than 1 person, visualise the results in diagrams*
   - *analyse what types of value-ladenness possibly play a role and to what extent*
   - *if the pedigree criteria were scored by more than 1 person, analyse to what extent there is disagreement on the scores among the participants*
   - *analyse whether changing assumptions is feasible and desirable*

REVISION

6. Revise/extend assessment
   - *perform a sensitivity analysis for the weak links*
   - *diversify assumptions*
   - *replace assumptions by other assumptions*

COMMUNICATION

7. Communication
   - *indicate in the assessment documentation:*
     - *the key-assumptions in the calculation chain*
     - *the weak links in the chain*
     - *the alternatives and the underpinning of the choices regarding assumptions made*
     - *the robustness of the outcomes of interest in view of the key-assumptions*