

Opening up scientific assessments for policy: The importance of transparency in expert judgements

Bart Strengers, Leo Meyer, Arthur Petersen, Maarten Hajer, Detlef van Vuuren, and Peter Janssen

PBL Netherlands Environmental Assessment Agency, P.O. Box 303, 3720 AH Bilthoven, The Netherlands

Abstract

Scientific assessments have become an important tool to support decision-making. Such assessments are formal efforts to assemble selected knowledge with a view toward making it publicly available in a form intended to be useful for decision-making. Important criteria for producing scientific assessments that are able to influence policy are salience, credibility, and legitimacy. We emphasize that for an assessment to be credible and legitimate, at least the expert judgements constituting the core of the assessment need to be made transparent. We propose a method to evaluate the quality of scientific assessments in that respect. This method is based on the evaluation by the PBL Netherlands Environmental Assessment Agency of a part of one of the most well-known scientific assessments, the 2007 Assessment Report of Working Group II of the Intergovernmental Panel on Climate Change (IPCC). Among the most common weaknesses found were insufficiently transparent expert judgements. We argue that authors and assessment practitioners should become more aware of the inevitable role of 'expert judgement', in which experts make an assessment despite high degrees of uncertainty, and should make those judgements more transparent, i.e. readers need to be able to follow the arguments of the assessment team. Furthermore, in order to become more reflective of different views, assessment methodology should incorporate a procedure of 'open assessment', for example by inviting 'outsiders' to participate in the quality control process.

1. Introduction

Scientific assessments have become an increasingly common tool to support international and national policymaking, particularly in the environmental domain (Clark et al. 2006; van Vuuren et al. 2011). Key examples at the global level include the IPCC Assessment Reports (IPCC 2007), UNEP's Global Environmental Outlook (UNEP 2007), OECD's Environmental Outlook (OECD 2008), the International Assessment of Agricultural Scientific Technology and Development (Watson 2008), and the Millennium Ecosystem Assessment (MA 2005). At the national level, many governments regularly publish environmental assessments, often integrated in state-of-the-environment reports. The reason for the popularity of assessment reports can (at least partly) be understood on the basis of the nature of (global) environmental change problems. These problems are often complex and beset with uncertainty (some of which are irreducible in nature) (Ostrom 1990; van der Sluijs et al. 2003; Petersen 2012). This complexity involves, among others, the importance of time and

geographical scales, and the multiple relationships between physical and economic elements. Moreover, understanding these problems usually requires the expertise of different disciplines. With respect to the societal setting, often the risk perceptions regarding these problems differ among different societal groups (see for instance Hulme (2009) for climate change). Finally, different actors involved have often diverging interests in both the outcomes and solutions.

In principle, well-presented scientific information can help to suitably discuss uncertainty of environmental problems and its consequences to further decision-making. In that context, Pielke (2007) discussed the different roles that scientists may have in support of decision-making: that of a pure scientist, science advisor, advocate and honest broker of policy alternatives. Pielke argues that the appropriateness of these roles depends on the importance of uncertainty and value-differences regarding the issues at stake. In situations of large uncertainty and strong disagreement about the issues at stake, science would be most useful in the role of honest broker, i.e. integrating stakeholder concerns with available scientific knowledge without choosing sides in the policy debate. This role is what is often intended by large environmental assessments: assembling 'selected knowledge with a view toward making it publicly available in a form intended to be useful for decision making' (Mitchell et. al, 2006). Formulated differently, scientific assessments are the institutional vehicle for presenting policymakers with information that would help them define and decide on policy strategies (even for questions that are not directly addressed by science itself).

Historic evidence has shown that this role is far from easy to fulfil. Pielke warns for several possible risks, in particular the risk of not being completely impartial. Other problems are associated with, for instance, focus, readability of end-products, and cooperation among different scientists. Assessment bodies have tried to develop procedures that should ensure that assessments are performed in the best possible manner. This includes the selection of authors, elaborated review procedures, protocols for uncertainty assessment, and provision of information to authors. Most international assessments, for instance, constantly inform their authors that 'assessments should be policy-relevant, but not policy-prescriptive' (IPCC, 2001; IPCC, 2010a).

Evaluation of assessments with respect to the question of whether they adequately fulfil their role at the interface between science and policy—in order to support learning processes—is important. In this article, we focus on the importance of transparency and use the PBL Environmental Assessment Agency's *ex post* evaluation of a part of the Intergovernmental Panel on Climate Change (IPCC)'s Fourth Assessment Report (AR4) as a case study to learn about the difficulties involved in performing environmental assessment in a transparent manner and to discuss possible ways to deal with these difficulties. The case study is relevant as the work of the IPCC is perhaps the best-known international scientific assessment for policymaking. Compared to other assessments, the quality procedures of the IPCC should be regarded as the most developed, partly also as a result of the involvement of different governments. Another reason for its relevance is that for climate change, uncertainties, different interests, and different risk perceptions pertaining to all aspects of causes, consequences, and response options are inextricably linked to a strong demand for transparency in the assessments.

Up until very recently, the IPCC mostly received praise for its activities, culminating in the award of the Nobel Peace Prize to the IPCC in 2007. Since 2009, however, the quality of the IPCC reports has been subject of fierce critique. In January 2010, the media reported

two errors in a specific part of the Working Group II (WGII) Report of the Fourth Assessment Report (AR4) of 2007 by the IPCC. The errors concerned an erroneously high rate of melting of the Himalayan glaciers (Sill 2010), and an erroneously high percentage of land area said to be lying below sea level in the Netherlands (PBL 2010a). Subsequently, several other problems were reported in the media, for example, on African agriculture (Leake 2010a) and damages from extreme weather events (Leake 2010b). The question of whether these latter issues were real 'errors' was heavily discussed (RealClimate 2010). However, since both the acknowledged and the alleged errors were relatively easy to understand by the general public *and* were pointed in the direction of alarmism, they affected the public perception of the scientific credibility of the IPCC, as well as its perceived legitimacy. Still, the errors as such did not influence the main conclusions that IPCC had drawn on projected regional climate impacts (PBL 2010b).

For the present article, the IPCC Fourth Assessment Report and the discussion around it represent a very useful case study to understand how difficult it is to do well in assessments. The case study is based on an evaluation project performed by the PBL Netherlands Environmental Assessment Agency, on request of the Dutch Minister for the Environment as a consequence of a resolution and debate in the Dutch parliament on 28 January 2010 concerning the reliability of the IPCC. The governmental agency was asked to conduct an assessment of the reliability of the regional chapters (9 to 16) of the IPCC WGII's contribution to the AR4. These chapters were selected as the errors discussed above were found in this part of the report. Arguably, the assessment of regional impacts is among the most difficult parts of IPCC assessments. Information on climate-change impacts is often only available in the form of local case studies and cannot always be found in peer-reviewed publications. Therefore, the IPCC has to make an assessment of the wider implications of such studies in order to answer the more general questions on potential impacts of climate change asked by the decision-makers. In the present paper, we are particularly interested to learn from PBL's evaluation of the IPCC assessment: how can assessments be made more credible and legitimate by increasing the transparency of expert judgements? In doing so we are aware that also other factors are of influence in the success of assessment to connect science and policy; and credibility and legitimacy are definitely not guaranteed by transparency. We do argue that a sufficient level of transparency is a necessary factor for success—even though it is hard to determine precisely what is the minimum level of transparency that would be required.

In the next section, a larger set of criteria is reviewed for producing scientific assessments that are able to influence policy. Section 3 provides a summary and analysis of the case study, which focuses on transparency and traceability. The final section contains conclusions that we derive from the case, which we claim to have wider validity for scientific assessments.

2. Criteria for creating influential scientific assessments

In the introduction, we provided a possible definition of scientific assessments; such assessments aim to assemble selected scientific knowledge with a view toward making it publicly available in a form intended to be useful for decision making. This focus on decision-making is the critical difference between assessments and normal scientific research. While

normal scientific methods often aim to reduce the role of uncertainty and divergent interpretations by further specifying the focal question and isolating the system at study, such uncertainty reduction is often not possible for answering policy-relevant questions on complex policy problems such as climate change (see also the discussion on post-normal science by Funtowicz et al. 1993).

In the scientific literature on assessments, some clear criteria for excellence in developing assessments have been derived by e.g. Mitchell et al. (2006), who focus on salience, credibility and legitimacy. Here we provide a brief elaboration of these three criteria, which does not aim to be complete:

- *Salience*: An assessment needs to address the relevant questions in the light of the decision-making problem. This may imply:
 - Active rephrasing of the original questions posed by the people involved in the assessment (the assessment team), based on their expertise.
 - Checking whether understandable and relevant answers to the original questions have been produced. This can be done in a dialogue between the authors and policymakers. Such a process is an element in the line-by-line approval sessions of the Summaries for Policymakers of IPCC reports.
 - Dealing with the questions in a comprehensive manner, that is, by explicitly incorporating all aspects that may be considered of relevance by potential users.

- *Credibility*: In order to be a useful element of a decision-making process, an assessment obviously needs to be scientifically credible to the main actors involved. From the many derived factors that can increase the credibility of an assessment we mention a few ones which especially relate to the scientific quality of the reported assessment findings:
 - A credible author team (competence, representative of different disciplines and geographical regions).
 - Based on high-quality input. This, for instance, involves the preference of using peer-reviewed scientific literature.
 - An adequate process of quality control of the assessment process (e.g. review procedures for the various draft versions of the assessment reports).
 - Transparency. As an assessment usually needs to interpret limited information to answer more generic questions, readers need—in principle—to be able to follow the arguments of the assessment team (in order to avoid a “just believe me” character). Even though many readers may not want to trace the underpinning of the assessment findings, in those cases where someone does want to know more about their underpinning, either by tracing it themselves or having expert advisers trace it for him or her, the assessments report should somewhere provide a traceable and verifiable account of the main assessment results.

- Scientifically correct. Errors in large volumes of text are difficult to fully prevent but require an extensive review process as they may undermine the credibility of the assessment.
- Adequately dealing with uncertainties. Assessments need to present all relevant uncertainties and provide information on how these uncertainties are evaluated and dealt with. Ideally it should provide also information on tacit and/or value-laden assumptions underlying the assessments, competing interpretations and controversies concerning scientific evidence and theories etc. The strength of the formulation of the assessment's main findings needs to reflect the most relevant uncertainties.

Many other aspects can be discerned that influence credibility to a large extent, and that are more directed towards the contextual, socio-cultural and institutional setting in which the assessment process takes place (see the literature on credibility from the field of social studies of science and technology, e.g. Gieryn 1999, Shapin 1995 and Jasanoff 1991, or from the field of risk assessment studies, e.g. Peters and Covello et al. 1997), which stress credibility features that have to do with the science/public interaction). Since the focus of the PBL evaluation was primarily to evaluate the reliability of the findings in (part of) the IPCC assessment reports, i.e. in the *product*, and not so much to evaluate the assessment *process*, we focus in this article on the above mentioned scientific credibility factors.

- *Legitimacy*: An assessment needs to be perceived as 'fair', that is, having considered the values, concerns, and perspectives of the main actors involved. One particularly important implication is that an assessment needs to be:
 - Impartial and unbiased. The risks assessed in environmental assessments, for instance, should not be understated or overstated.
 - Addressing issues of equity. Both the perspectives of those who stand to lose and of those who stand to gain from the developments assessed need to be reflected in the assessment.

As Mitchell et al. (2006) stress, the above mentioned attributions of excellence in assessments are typically judged in relational contexts (science to audience/public), and various audiences of an assessment can value it differently, depending on goals, interests, beliefs, expectations, strategies and involvement of the audience. Therefore it remains a (difficult) challenge to achieve salience, credibility and legitimacy, especially in situations with multiple audiences and where tactics to promote one attribution can undermine achieving the others. Mitchell et al. (2006) conclude that the way in which an assessment is set up, performed and communicated—i.e. the assessment process and its relational aspects—plays often a more decisive role in determining its influence than the assessment reports which result from the process. But, as the IPCC Fourth Assessment Report case shows, doubts on the quality and reliability of the findings in the delivered assessment reports can definitely undermine the assessment's influence as well. In this light it is of importance to evaluate the assessment from the perspective of its assessment reports as well, even though this does not comprise the complete story.

3. A method for evaluating scientific assessment reports: The case of the IPCC Fourth Assessment Report

3.1. Qualifications in evaluating scientific assessment reports

Scientific assessment reports have not been evaluated systematically on a regular basis before. Of course, the reports underwent peer review, but the criteria for such peer review have not been spelled out and the peer review has not systematically managed to make sure that every part receives the level of scrutiny it deserves. Therefore, for the PBL evaluation of a part of the IPCC report—an evaluation pertaining specifically to the underpinning of the main conclusions of the assessment product—a new evaluation framework needed to be developed, focusing in particular on errors and transparency (PBL 2010b). This has been done by taking into account the credibility criteria discussed in the previous section and operationalizing them into a small number of types of qualifications of assessment statements. PBL distinguished between two main types of evaluation categories: ‘errors’ (how the question was originally framed by the Dutch parliament) and ‘comments’ (the new frame proposed by PBL as to where the real issues lie):

- *Errors.* The term ‘error’ was reserved for obvious (unambiguous) scientific errors. In the PBL evaluation, several of these were identified in the report, including the incorrect formulation regarding the area of the Netherlands below sea level. Within this category, PBL distinguished between inaccuracy (e.g. misrepresenting a range, but without any real content implications) and real error (a faulty statement that leads to incorrect conclusions). For the first, a simple erratum could correct the error (type E1a, see Table 1) while for the second, formally a new assessment of the issue at stake would be needed (type E1b). Finally, as part of this error category one could also include inadequate references (E2).
- *Comments.* For most of the credibility criteria described in Section 2, the word ‘error’ seemed inadequate, however. One reason is that an assessment does, to some degree, contain subjective elements. Lack of transparency is therefore not (necessarily) an error, but it still reduces the credibility of an assessment. PBL introduced the qualification of ‘comments’ as a complement to ‘errors’ and distinguished seven categories of comments:
 - C1. Insufficiently substantiated attribution
 - C2. Insufficiently founded generalisation
 - C3. Insufficiently transparent expert judgement
 - C4. Inconsistency of messages
 - C5. Untraceable reference
 - C6. Unnecessary reliance on grey referencing
 - C7. Statement unavailable for review

See Table 1 for a further explanation of the different types of qualification.

Table 1 *Typology of errors (referred to with an E) and comments (referred to with a C). In parentheses is indicated how many instances of each error or comment were identified in the investigation, see Section 3 and Table 2.*

Type	Description	Remarks/Explanation
E1	Inaccurate statement (7)	Two subcategories were defined for this type of error.
E1a	Errors that can be corrected by an erratum (5)	For example, typographical errors, incorrect phrasing of part of a sentence, wrong dimensions, and wrong reference years.
E1b	Errors that require a redoing of the assessment of the issue at hand (2)	Such as establishing a new range of numbers by revised calculations from the reference sources available during the assessment period, and/or rephrasing of the expert judgement including its uncertainty labelling.
E2	Inaccurate referencing (3)	A reference to a wrong source, or source not correctly cited. In all cases, an erratum would be needed.
C1	Insufficiently substantiated attribution (1)	The presence of multiple stresses is not sufficiently signalled, or there is a one-sided attribution of impacts to climate change, while other factors also would have been expected to play a critical role (e.g. population growth, industrialisation, migration, and changes in land use and land cover).
C2	Insufficiently founded generalisation (2)	A proper argumentation is lacking or the evidence in the references does not justify a generalisation or extrapolation of impacts in one country or sector to include entire regions and/or additional sectors.
C3	Insufficiently transparent expert judgement (10)	The reasoning behind an expert judgement, including the reasoning behind its level of likelihood and/or confidence, is not accessible to a non-expert reviewer. It does <i>not</i> imply the judgement is wrong, since the authors may have had their reasons, and may have considered additional information or knowledge that was not explicitly referred to.
C4	Inconsistency of messages (2)	A message's content and/or confidence level change when going from the main text to a summary. The IPCC procedures require that all summary texts are consistent with the main text or lower level summaries.
C5	Untraceable reference (3)	A reference in a statement cannot be found at all.
C6	Unnecessary reliance on grey referencing (2)	A reference to a grey publication, although strong peer-reviewed journal references were available at the time of writing the report. Notice that grey literature is an indispensable part of many assessments since not all relevant literature is published in peer-reviewed scientific journals.
C7	Statement unavailable for review (1)	If a new statement contained new literature findings that were not clearly derived from a content issue raised in the Second Order Draft ¹ (SOD) review, then this new material would have been kept out of the review process. This is not supposed to happen; expert reviewers or governments must have access to all findings that will be part of the published report.

¹ The first round in the review process of IPCC reports comments on the First Order Draft (FOD) by expert reviewers. Subsequently, the findings of these reviewers must be dealt with, which results in a Second Order Draft (SOD) that is again reviewed by the same expert reviewers *and* the governments of the countries that are members of the IPCC. The SOD and its review form the basic material for finalizing the report.

3.2. Evaluation approach

Regarding the evaluation of the IPCC report, PBL decided to base it on the so-called 'information pyramid' of the IPCC which reflects the various levels within assessment reports (summaries and synthesis reports) as well as the basic literature (peer-reviewed literature and other publications) underlying the IPCC assessment reports (see Figure 1). The base of the pyramid consists of peer-reviewed literature and other publications (designated as level 0 in Figure 1)², which are referred to at level 1: the main chapter texts. The next level (level 2) is formed by the Executive Summaries of the individual chapter texts. Part of the information in these summaries was used in the Technical Summary (level 3). Level 4 consists of the Summary for Policymakers (a short document, approved line-by-line by the governments involved in the IPCC). The information in the Synthesis Report (level 5) has a higher level of aggregation than that in the Working Group Reports, which, together, consist of over 2,800 pages. The information in the Summary for Policymakers (SPM) of the Synthesis Report represents the highest level of aggregation (level 6).

By tracing the statements at this highest level back to their original sources, the transparency can be assessed with which scientific facts on climate-change impacts had 'travelled' from their sources (the peer-reviewed literature) to their destination (the SPM of the Synthesis Report), cf. Morgan (2010). The IPCC process should deliver useful and reliable knowledge to policymakers, and it is expected that transparency of the reasoning leading to the IPCC summary statements has positive influence on their credibility.

PBL, in its evaluation, started with the thirty two summary conclusions on regional impacts as presented in Table SPM.2 of the Synthesis Report (IPCC 2007). These conclusions feature examples of projections of climate-change impacts on food, water, ecosystems, coastal regions, and health, for all the Earth's continents. A number of findings were formulated referring to the foundations for statements on regional impacts in addition in the Technical Summary and Executive Summaries of the regional chapters. In each case, the information was traced back to the main text of the report.

These findings were presented to and discussed with the responsible IPCC authors in several review rounds, and the draft of the PBL report was subject to external review. It should be noted that in most cases the analysts of PBL and the IPCC-authors agreed upon the assessment of the existing text (although the final judgement remained the responsibility of PBL).

3.3. Results from the evaluation

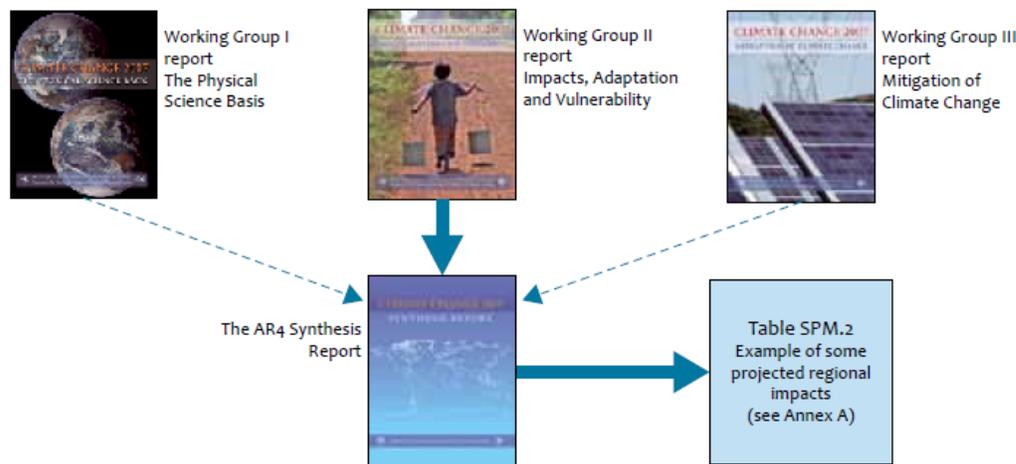
The framework described above turned out to work very effectively in the PBL evaluation. Table 2 provides an overview of the outcome of the evaluation.³ Overall, in eight of the thirty two summary conclusions in Table SPM.2 of the Synthesis Report, PBL identified two statements qualified as errors and made a total of nine comments. Additionally, in twenty one conclusions at other summary levels or in the main text of the regional chapters, nine errors and thirteen comments were identified, with one error and four comments being

² One could, theoretically, also check the references in the cited sources ('level -1').

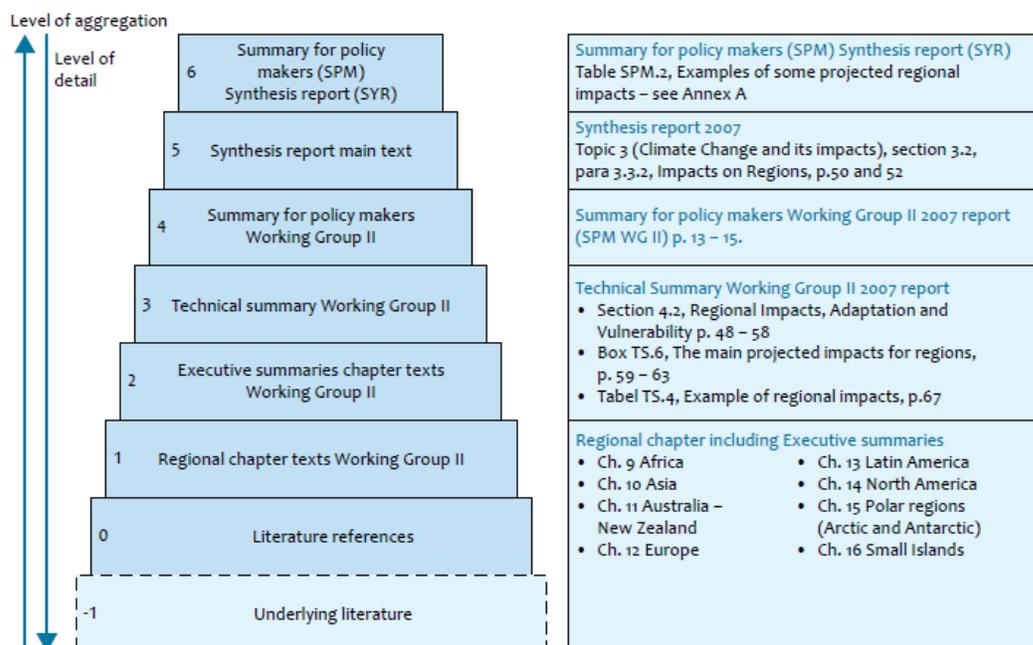
³ This table has been produced specifically for this paper and was not part of the PBL (2010b) report.

Figure 1 Structure of IPCC's Fourth Assessment Report, focusing on findings on projected regional climate-change impacts

A. Structure of the Fourth Assessment Reports



B. Regional chapter analysis pyramid



classified as 'major'. A detailed discussion can be found in the PBL evaluation report (PBL 2010b); here we only discuss some of the most salient issues.

Interestingly, an arguably small number of real errors were found. In addition to the errors already identified in media—the area of the Netherlands below sea level and the statement on Himalayan glaciers—nine further minor errors were found. The far majority of these were qualified as inaccuracies. Examples of inaccuracies included incorrect range statements in Table SPM.2 (not 75 to 250 million people, but 90 to 220 million people are projected to be exposed to increased water stress due to climate change in Africa) (E1b) and

Table 2 Overview of the errors and comments with respect to the regional chapters of the Working Group II Report of the IPCC Fourth Assessment Report. Explanation of codes is given in Table 1. Bold underlined numbers indicate that the error or comment was corrected in the erratum list on the IPCC website. The PBL analysis resulted in an erratum for nine of the eleven errors found. For the 23 comments, only a minor one resulted in an erratum (see also Section 4.1). Numbers between brackets refer to the number of instances that the particular type of error or comment was found. The column #S refers to the number of statements related to these comments and errors. #E = number of errors and #C = number of comments.

		Table SPM.2					Additional				
		Major		Minor		#S	Major		Minor		#S
Africa		C3,C5,C7		E1b,C3		3	<u>E1a</u> ,C4		E1b		3
Asia		C2,C3				1	C3,C6		<u>E2</u>		2
Aust & NZ				<u>E2</u> ,C3		1	C1		C4, <u>E1a</u>		3
Europe				C3		1			<u>E1a</u> ,C3(3), <u>C4</u>		5
L America		C2				1			E1a, <u>E1a</u> , <u>E2</u> ,C5(2),C6		6
N America				C3		1					
Poles											
Islands									<u>E2</u> ,C3		2
Total	#E			E1b, E2	2	2	E1a	1	E1a(4),E1b,E2(3)	8	8
Total	#C	C2(2),C3(2) C5, C7	6	C3(4)	4	6	C3,C4,C6,C1	4	C3(4),C4(2),C5(2),C6	9	13

incorrect referencing in the main text of Chapter 11 on Australia and New Zealand, which underpins the first summary statement in Table SPM.2 for this region (E2). One major new error was found, in Chapter 9 on Africa. A projected decrease by 50 to 60% in *extreme wind and turbulence* over fishing grounds was mistakenly represented as a 50 to 60% decrease in *productivity* as a result of changes in wind and turbulence. However, this error and the other remaining errors did not travel to the level of the summaries, and therefore did not affect the summary conclusions.

Here are two examples of the comments that were made by PBL:

- For Latin America, in the summary it is stated that ‘livestock productivity is projected to decline’. In the underlying material, ‘livestock’ was found to be limited to ‘cattle’ and is underpinned by references to studies on Bolivia and central Argentina only. There is no text on how the limited material was interpreted at the more general level of Latin America and in terms of livestock. This makes the assessment conclusion an insufficiently founded generalisation (C2).
- Three major comments (C3, C5, and C7) were made on a summary conclusion with respect to Africa: ‘By 2020, in some countries, yields from rain-fed agriculture could be reduced by up to 50%’. The statement was based on a report (Agoumi 2003) that referred to untraceable studies (C5) and to the Initial National Communications (INCs) of Morocco, Algeria, and Tunisia. However, only the INC of Morocco (Kingdom of Morocco 2001) reported a decline in yields for 2020: ‘a 50% reduction in cereal

yields in Morocco might occur in dry years and 10% in normal years'. The IPCC authors explained that present-day climates and projected future climate change in the three countries are very similar and also that only cereals are grown without irrigation. Furthermore, using information from EUROSTAT (2010), the authors made plausible that, due to current climate variability, the yields in Algeria, Morocco and Tunisia have been varying annually, including yield reductions of nearly 70% in individual years, in the period between 2000 and 2006. In hindsight, these additional explanations could have provided further foundations for the statement (C3). Also, the INC of Morocco indicated that this statement was based on 'The Study of Morocco's Vulnerability to CC Effects' that could, however, not be traced nor could it be provided by the IPCC authors (C5). Although the IPCC does not formally require that authors check references of references of references, such in-depth checks would make sense for statements that reach the SPM and the Synthesis Report. Finally, the statement seemed to have been added to the main text of Chapter 9 after the Second Order Draft had been reviewed, but we were unable to find the particular substantive comment(s) made in this review that could have led to the addition of this statement (C7).

The comment 'insufficiently transparent expert judgement' (C3) clearly dominated the results. For six out of the thirty two key statements in Table SPM.2, the regional chapters in PBL's view insufficiently convey the IPCC authors' reasoning behind their weighing of the evidence that was available from the literature. For another four important statements in the regional chapters, information on the reasoning behind them is lacking in the report.

3.4. Recommendations for scientific assessments identified in the evaluation

Based on the findings in the PBL (2010b) report, summarized in Section 3.3, some recommendations for producing credible and legitimate assessment reports were made. It should be noted here that several important improvement steps have been taken with respect to the IPCC. Most of these steps were in explicit response to the 'Review of the Processes and Procedures of the IPCC', which was published by the InterAcademy Council in August 2010 (IAC 2010), and which had taken up some of the PBL (2010b) recommendations. The improvement steps were initiated during the 32nd Plenary Session of the IPCC in Busan, South Korea (October 2010, IPCC 2010b), and led to decisions taken at the 32nd and 33rd IPCC Plenaries in May and November 2011. The PBL (2010b) recommendations, however, are still relevant, not in the least in the context of other assessments. Here we briefly discuss four of the most important recommendations.

Raising assessment quality of the authors. In Section 2, we indicated that an assessment is different from normal scientific work. Given the inability to reduce uncertainty by specifying and/or isolating the research topic, the criteria worked out in Section 3 under 'comments' (C1–C7) are of key importance to uphold transparency, and hence increase credibility and legitimacy. In our view, many of the authors involved in assessments have too little awareness of these key issues. Communication of the essence of assessment, the methodologies, the possible traps, etc. to authors could help to seriously increase the quality

of the assessment team (beyond the quality of knowledge on the issue at stake, for which the authors were selected). Such communication is currently still lacking in all (global environmental) scientific assessments.

Error handling. Assessments should make sure that they have the option to respond to error. Most assessments (including the IPCC, before the recent procedural changes) do not have formal procedures for error correction. This is rather odd at a time when it is easily possible to publish errata on websites. It should be expected and accepted that in a large report there are likely to be small errors and that therefore, procedures need to be developed for publishing errata.

Foundations of summary conclusions. 'Insufficiently transparent expert judgement' was the dominant comment with respect to the investigated statements reported on by PBL (2010b). Clearly, expert judgement is a key practice in assessments. Yet, it is essential that more attention is paid to the transparency of reasoning in such situations to enhance verifiability and thus potential credibility of expert assessments. This definitely contributes to scientific credibility, but, as Jasanoff (2010) stipulates, still does not serve to fully address accountability of the science to the public. In a significant number of cases, the IPCC authors, when asked, were able to reconstruct their reasoning with hindsight, and conveyed them to the PBL. Nevertheless, we would argue that deliberative reasoning must be published in the report itself, in order to enhance the credibility of an assessment. Another type of comments deserving wider attention pertains to the use of grey literature. In our view, grey literature should be used very carefully (even if we acknowledge that the peer-reviewed literature has significant gaps). Use of grey literature should always require a critical evaluation of scientific status.

Balancing the value orientations behind the assessment. The PBL evaluation revealed that IPCC authors considered negative impacts to be the most relevant to policymakers. IPCC had adopted a risk-oriented approach. This may explain why all comments and errors listed in the PBL assessment highlighted a risk-averse assessment of facts. Obviously, the IPCC may choose that its mandate is to take a risk-oriented approach. However, most readers may not be aware of this and thus, the assessment becomes one-sided. For communication, it seems better to clearly separate the most likely development from the possible risks. In fact, important positive impacts of (in this case) climate change should also be mentioned.

Opening up scientific assessments. Besides making expert judgements more transparent (and thus also opening them up for scrutiny), there are other and more direct ways to open up scientific assessments, for instance, by involving wider groups of experts. In producing assessment reports, expert reviewers should be invited to contribute comments in an open process. There should also be a public request for submission of possible errors that should be considered for errata or reassessment.⁴

⁴ The PBL practised such a process by launching a public website as part of its investigation into the IPCC report. The website was available for the course of one month and gave all experts in the Netherlands the opportunity to contribute to our investigation. We asked for submissions of possible errors found in regional chapters of the WGII Report. By the end of that month, the PBL had registered forty submissions; however, most of them were about issues related to the Working Group I Report. Two submissions qualified to be addressed in our report. All submissions and PBL's responses have been published on the PBL website (<http://www.pbl.nl/>).

4. Conclusion

The methodology presented in this paper to evaluate scientific assessments can be used in evaluation and review processes for other (global) assessments, either within or outside the environmental domain. This methodology has been applied to summary statements about projected regional impacts from the IPCC Working Group II Report on climate-change impacts, adaptation, and vulnerability. We claim that this approach could also be applied to other scientific assessments, for instance, the assessments that will be performed by the Intergovernmental Science–Policy Platform on Biodiversity and Ecosystem Services (IPBES) or the assessments performed by the PBL Netherlands Environmental Assessment Agency itself.

The most important problem in the evaluated part of the IPCC report was found to reside in insufficiently transparent expert judgements. Our evaluation has helped us and others to understand some of the difficulties involved in writing assessment reports. It must be added here that the PBL (2010b) results showed that, despite our critical review, the far majority of the conclusions turned out to be well founded (after we had received more information from the authors than was available in the report)—and that very few real errors could be found.

There are several options to further improve the quality of scientific assessments and its reporting. In the PBL report (2010b), a set of possible improvements for the IPCC assessment process was derived, which have a much broader validity, beyond the IPCC. Thus, most of PBL's recommendations are also relevant for other assessments. Measures include: raising assessment quality of the authors, error handling, foundations of summary conclusions, balancing the value-orientations behind the assessment, and opening up the scientific assessment in several ways. All in all, we argue that assessors should become more aware of the inevitable role of 'expert judgements', in which experts conduct assessments despite high degrees of uncertainty and should make these judgements more transparent. In an open assessment procedure, the logic of the reasoning is made public and many 'outsiders' are invited to participate in the quality control process.

Acknowledgements

We would like to thank Annemieke Righart and Hannah Martinson Hughes for correcting and improving the English text. Moreover, we would like to thank all PBL authors involved in the evaluation of parts of the IPCC WGII report. Finally, we would like to thank the IPCC authors for their help in performing the evaluation, which provided the case for this paper.

References

- Agoumi (2003). *Vulnerability of North African countries to climatic changes: adaptation and implementation strategies for climatic change. Developing Perspectives on Climate Change: Issues and Analysis from Developing Countries and Countries with Economies in Transition*. IISD/Climate Change Knowledge Network. <http://www.cckn.net/pdf/north_africa.pdf> accessed 22 June 2012.
- Funtowicz, S. and Ravetz, J. (1993). "Science for the post-normal age." *Futures*, 25: 739-755.
- Gieryn, T. F. (1999). *Cultural Boundaries of Science: Credibility on the Line*. University of Chicago Press: Chicago..
- Hulme, M. (2009). *Why We Disagree about Climate Change : Understanding Controversy, Inaction and Opportunity*, Cambridge University Press: Cambridge.
- IAC (2010). *Climate Change Assessments: Review of the Processes and Procedures of the IPCC*. InterAcademy Council (IAC): Amsterdam.
<<http://reviewipcc.interacademycouncil.net/report.html>> accessed 22 June 2012
- IPCC (2001). *Climate Change 2001: Synthesis report. Summary for policy makers. An assessment of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge.
<http://www.grida.no/climate/ipcc_tar/vol4/english/pdf/spm.pdf> accessed 22 June 2012
- IPCC (2007). *Climate Change 2007 - Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge.
- IPCC (2010a). *Statement on IPCC principles and procedures*. 2 February 2010.
<<http://www.ipcc.ch/pdf/press/ipcc-statement-principles-procedures-02-2010.pdf>> accessed 22 June 2012
- IPCC (2010b). *Decisions taken by the Panel at its 32nd session. With regards to the Recommendations resulting from the Review of the IPCC Processes and Procedures by the InterAcademy Council (IAC)*. 11-14 October 2010. Busan, Republic of Korea.
<http://www.ipcc.ch/meetings/session32/ipcc_IACreview_decisions.pdf> accessed 22 June 2012
- Jasanoff, S., 1991. Acceptable evidence in a pluralistic society. In *Acceptable evidence*, edited by D. Mayo and R. Hollander. Oxford, UK: Oxford University Press.
- Jasanoff, S., 2010. Testing time for climate science. *Science* 328: 695–696.
- Kingdom of Morocco (2001). *Initial National Communication on the United Nations Framework Convention on Climate Change*. Ministry of Land-use Planning, Housing, and the Environment, Morocco. <<http://unfccc.int/resource/docs/natc/mornc1e.pdf>> accessed 22 June 2012
- Leake, J (2010a). Africagate: top British scientist says UN panel is losing credibility. *The Sunday Times*, 7 February 2010
<<http://www.timesonline.co.uk/tol/news/environment/article7017907.ece>>.
- Leake, J, (2010a). UN wrongly linked global warming to natural disasters. *The Sunday Times*, 24 January 2010
- MA (2005). *Millennium Ecosystem Assessment Synthesis Report*. Island Press: Washington DC.
- Mitchell, R. B., Clark, W. C., Cash, D. W., and Dickson, N.M. (2006). Evaluating the Influence of Global Environmental Assessments. *Global Environmental Assessments: Information and Influence*, MIT press: Cambridge.
- Morgan, M. S. and Howlett, P. (2010). *Traveling facts. How Well Do Facts Travel?* Cambridge University Press: Cambridge.
- OECD (2008). *OECD Environmental Outlook to 2030*. OECD: Paris.

- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press: Cambridge.
- PBL (2010a). *Correctie formulering over overstromingsrisico Nederland in IPCC-rapport*. (Correction of the statement on the risk of flooding for the Netherlands in IPCC-report), Netherlands Environmental Assessment Agency (PBL): The Hague/Bilthoven, the Netherlands.
- PBL (2010b). *Assessing an IPCC Assessment: An Analysis of Statements on Projected Regional Impacts in the 2007 Report*. . Netherlands Environmental Assessment Agency (PBL): The Hague/Bilthoven, the Netherlands.
- Peters, R. G., Covello, V.T., McCallum D.B. (1997). The determinants of trust and credibility in environmental risk communication: an empirical study. *Risk Analysis* 17(1): 43-54, 1997.
- Petersen, A. C. (2012 [2006]). *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*, Second Edition. FL: CRC Press: Boca Raton
- Pielke, R. (2007). *The Honest Broker*, Cambridge University Press: Cambridge.
- RealClimate (2010). *Leakegate: a retraction*.
<<http://www.realclimate.org/index.php/archives/2010/06/leakegate-a-retraction/>> accessed.
- Shapin, S. (1995). Cordelia's love: credibility and the social studies of science. *Perspectives on Science* 3(3): 255-275.
- Sill, J. (2010). "Tracking the Source of Glacier Misinformation." *Science* 327: 522-523.
- UNEP (2007). *Global Environment Outlook 4*. UNEP: Nairobi.
- van der Sluijs, J. P., J. S. Risbey, et al. (2003). RIVM/MNP Guidance for Uncertainty Assessment and Communication. *Detailed Guidance*. Utrecht University: Utrecht.
- van Vuuren, D. P., M. Kok, et al. (2011). "Scenarios in Global Environmental Assessments: Key characteristics and lessons for future use." submitted.
- Vorobyov, Y. (2004). *Climate change and disasters in Russia*. World Climate Change Conference, Moscow, Institute of Global Climate and Ecology.
- Watson, B., Ed. (2008). *International Assessment of Agricultural Science and Technology Development*.
<<http://www.agassessment.org/index.cfm?Page=IAASTD%20Reports&ItemID=2713>> accessed 22 June 2012