

Equipped to deal with uncertainty in climate and impacts predictions: lessons from internal peer review

Anna Wesselink · Andrew Juan Challinor · James Watson · Keith Beven · Icarus Allen · Helen Hanlon · Ana Lopez · Susanne Lorenz · Friederike Otto · Andy Morse · Cameron Rye · Stephane Saux-Picard · David Stainforth · Emma Suckling

Received: 30 September 2013 / Accepted: 14 July 2014 / Published online: 12 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The quantification of uncertainty is an increasingly popular topic, with clear importance for climate change policy. However, uncertainty assessments are open to a range of interpretations, each of which may lead to a different policy recommendation. In the EQUIP project researchers from the UK climate modelling, statistical modelling, and impacts communities worked together on ‘end-to-end’ uncertainty assessments of climate change and its impacts. Here, we use an experiment in peer review amongst project members to assess variation in the assessment of uncertainties between EQUIP researchers. We find overall agreement on key sources of uncertainty but a large variation in the assessment of the methods used for uncertainty assessment. Results show that communication aimed at specialists makes the methods used harder to assess. There is also evidence of individual bias, which is partially attributable to disciplinary backgrounds. However, varying views on the methods used to quantify uncertainty did not preclude consensus on the consequential results produced using those methods. Based on our analysis, we make recommendations for developing and presenting statements on climate and its impacts. These include the use of a common uncertainty reporting format in order to make assumptions clear; presentation of results in terms of processes and trade-offs rather than only numerical ranges; and reporting multiple assessments of uncertainty in order to elucidate a more complete picture of impacts and their uncertainties. This in turn implies research should be done by teams of people with a range of backgrounds and time for interaction and discussion, with fewer but more comprehensive outputs in which the range of opinions is recorded.

This article is part of a Special Issue on "Managing Uncertainty in Predictions of Climate and Its Impacts" edited by Andrew Challinor and Chris Ferro.

Electronic supplementary material The online version of this article (doi:10.1007/s10584-014-1213-1) contains supplementary material, which is available to authorized users.

A. Wesselink (✉) · A. J. Challinor · J. Watson · K. Beven · I. Allen · H. Hanlon · A. Lopez · S. Lorenz · F. Otto · A. Morse · C. Rye · S. Saux-Picard · D. Stainforth · E. Suckling
University of Twente, Enschede, The Netherlands
e-mail: a.j.wesselink@utwente.nl

1 Introduction

Predictions of climate change and its impacts form the basis of many policy decisions. Perhaps most notably, scientific knowledge compiled by the Intergovernmental Panel on Climate Change (IPCC) has been at the heart of attempts to build a global policy regime centred on the UN Framework Convention on Climate Change (UNFCCC), and especially its Kyoto protocol (Hulme and Mahony 2010). In spite of the expanding knowledge base, global and national climate change policy making remains fraught with controversies that often focus on uncertainty in the scientific input. Trust in this science was further shaken by interpretations of unauthorised release of emails from the Climate Research Unit at the University of East Anglia, which became known as Climategate. Around the same time a few minor errors were discovered in the IPCC 4th assessment (Grundmann 2013; Skrydstrup 2013). Although from an academic point of view these events were of minor importance, politically they were highly significant because they were used by climate change sceptics and deniers to undermine scientific credibility and thus give ammunition to arguments that no action is needed with respect to climate change (Leiserowitz et al. 2012). However, this phenomenon is not new. Already in 1992 Brown concluded that ‘scientific uncertainty has become an operational synonym for inaction on global environmental issues’ (Brown 1992 p.19). On the other hand, when scientific results are used as arguments for climate action they are often taken at face value, without due consideration of uncertainties. In the latter situation uncertainties are politically inconvenient because they undermine the strength of the arguments for taking action (Brunner 1996; Webster 2003). Thus, there is a tendency in society to either not trust climate science at all, as in the case of climate sceptics who insist on perfect knowledge, or to trust it too much, when uncertainties are brushed under the carpet (Shackley et al. 1999). In summary, society often has trouble dealing with scientific uncertainty.

At the same time, climate scientists are continuously looking for better ways to assess and communicate uncertainties. Assessing uncertainties is part of quality assurance in scientific peer review, and uncertainty is also a topic in its own right that is discussed extensively in scientific publications (e.g. Stainforth et al. 2005; Knutti et al. 2008; Knutti and Sedláček 2012). Concerns for policy application have resulted in detailed guidelines for communication of uncertainties (Morgan et al. 2009; Mastrandrea et al. 2010), resulting for example in judgements of uncertainty in all IPCC assessment reports that have been gradually refined and better defined over time (Kandlikar et al. 2005; Risbey and Kandlikar 2007; Curry 2011; Jones 2011). These strategies can contribute to preventing over-confidence in both academia and society and the inevitable backlash this causes when it is undermined by events such as Climategate.

Being able to communicate uncertainties presupposes knowing what they are and this is no simple matter in the domain of climate change. The climate models that are used to produce projections of climate change are extremely complex and associated with many different uncertainties (Petersen 2012). In addition, ‘there are significant differences in opinion amongst modellers, indeed what could be termed different cultures of doing climate modelling [...] These different cultures result in different sets of standards by which climate change science is evaluated. What is a good piece of research according to those following one style, is not viewed so favourably by those working in a different style. The existence of different styles raises issues concerning the assessment of GCM modelling for policy purposes’ (Shackley et al. 1999). However, policy makers are also interested in the likely impacts of projected climate change, not just in changes in climate itself. In the estimates of climate change impacts the uncertainties associated with impact models are combined with the uncertainties associated with climate projections to yield a complex array of uncertainties (Moss and Schneider 2000; Challinor et al. 2013). The resulting elements of uncertainty can be thought of as a chain, or

cascade (Schneider 2001), consisting of some or all of: observations of climate, projections of future CO₂, climate model output, climate indices derived from climate model output, and projected impacts (Challinor et al. 2009).

To assess the uncertainties in predicted climate change impacts an ‘end-to-end’ uncertainty assessment is therefore required which addresses each of the elements described above. This special issue presents some of the results of the EQUIP project, whose aim was exactly that (<http://www.equip.leeds.ac.uk/>). This three-year consortium project brought together researchers from the UK climate modelling, statistical modelling, and impacts communities to work closely together on developing risk-based prediction for decision making in the face of climate variability and change. EQUIP also aimed to develop new methods for evaluating climate and impacts predictions, with a particular focus on marine ecosystems and crop production and on probabilistic predictions of changes in heat waves and drought.

Ultimately, uncertainties in climate projections are unknowable since they can only be verified in the future, so it is dangerous to take only one assessment of uncertainty at face value. Preferably, different assessments of uncertainties are therefore taken into account. We assert that this can be done by involving researchers from different scientific communities in producing uncertainty assessments. In this paper we therefore focus on the variation in the assessment of uncertainties between researchers. We start with the observation that different disciplinary or peer communities have different approaches to assessing uncertainties. A comprehensive assessment of uncertainties would therefore benefit from combining knowledge from these different communities. To achieve this, we conducted an experiment in internal peer review amongst the EQUIP members for the papers submitted to this special issue.¹ This experiment permits us to assess and report different views on uncertainty. We begin by describing EQUIP research and explaining further the rationale for this experiment (Section 2). Subsequently, we give an overview of the experimental design (Section 3) and the results (Section 4); details are presented in the Electronic Supplementary Material. We then discuss how such extended peer review can provide a better understanding of uncertainties (Section 5) and we conclude with the implications of our findings for the conduct of policy relevant research (Sections 6 and 7).

2 Rationale: capturing the range of uncertainties in climate and impact assessments

Four papers in this special issue present quantifications of different elements of the cascade of uncertainty (Section 1). Hanlon et al. (2013) (abbreviated to H2013) assess how well extreme heat events in the near future can be predicted using an ensemble of 9 climate model runs. They evaluate model skill through hind casting and then quantify some of the uncertainties through bootstrapping with replacement. Otto et al. (2013a) (abbreviated to O2013) also look at extreme weather events, in this case precipitation, but they explore the extent to which these can be attributed to human influence. Their method is probabilistic event attribution based on large ensembles of hind cast weather simulations. Their uncertainty analysis uses bootstrapping and focuses on the effect of initial model conditions. Saux Picart et al. (2013) (abbreviated to SP2013) use a statistical method to overcome observation scarceness in oceanography that provides confidence intervals on the statistics associated with the observational distributions. This results in a combined dataset of observations and model-derived data that enables the assessment of ecosystem indices for the current climate and under different

¹ The design and analysis presented here was done by the first two authors with input from the third author. The other co-authors participated in the reviews and commented on this paper.

climate change scenarios. Their uncertainty assessment focuses on the compiled input data set. Watson et al. (2014) (abbreviated to W2014) investigate the relevance of key weather input characteristics to crop model skill, which is assessed through perturbing observed weather and observing the changes in predicted crop yield; this constitutes an uncertainty assessment of crop model input data. They compare a process-based and a statistical crop model with the aim to combine the strengths of both approaches.

Together, these papers address different elements of climate impacts assessment and cover different disciplines. In Table 1 the respective foci are summarized. It should be read as follows: ‘H2013 mainly addresses uncertainties in climate models/scenarios as they affect climate indices, but it also pays attention to uncertainties in input data/observations’, etc.

To capture the range of opinions amongst the experts involved in EQUIP we used questionnaires to:

- 1) identify the opinions of EQUIP members regarding the importance of different sources of uncertainty;
- 2) make explicit the implicit decisions taken by authors on methods and the role of expert judgement in interpreting results;
- 3) identify disagreement between EQUIP researchers;
- 4) identify what types of uncertainty are addressed by EQUIP members and which are not.

As a benchmark for assessing the latter, we refer to Petersen’s (2012) typology of uncertainty in climate simulations:

- ontological, e.g. limits in our current understanding of climate and impact processes, or processes that have not been included in the current generation of models;
- epistemological, e.g. intrinsic impossibility to measure variability in climate parameters at sufficient time and spatial scales together with limits to predictability resulting from chaotic processes;
- methodological, e.g. using perturbed-parameter and multi-model ensembles to produce climate projections;
- axiological, e.g. regarding the purpose and underlying worldview of the research.

Other classifications have been used (e.g. Van Asselt and Rotmans 2002) but they do not usually cover the whole range of paradigmatic uncertainty. Petersen (2012) asserts that each type of uncertainty can occur in five locations: in the conceptual model, the mathematical model (structure and parameters), the technical model, model inputs and in output data and interpretation. Only some of these uncertainties are statistically quantifiable; most can only be assessed through qualitative judgement for which ‘the (variable) judgement and best practice of a scientific community provides a reference’ (Petersen 2012 p.58). This means that ‘the

Table 1 End-to-end uncertainty in predicted climate impacts

	Input data/ observations	Climate model/ scenarios	Climate indices	Predicted impacts
H2013	u	U	Q	
O2013	u	U	Q	
SP2013	U			Q
W2014	U			Q

Quantities calculated (Q) and uncertainties assessed (U = focus; u = some attention) in each of the four papers

broader the community, the more likely it is that the different epistemic values held by different groups of experts could influence the assessment' (*ibid.*).

In EQUIP we had just such a (relatively) broad community. The disciplinary backgrounds of EQUIP scientists include statistics, (atmospheric) physics, philosophy, computer science, ecosystem modelling, biochemistry and political science. Two groups could conceivably be distinguished: those who work on climate modelling and those who work on climate impact modelling (and a political scientist who does neither), but within either group individual foci are different. Some scientists had worked together previously, others had not, and they publish partly in the same and partly in different journals. These scientists do not, therefore, belong to one well-defined peer community, nor do they belong to very different communities (except the political scientist); what unites them is working on the same questions of climate and impact modelling. Robinson (2007) labelled this 'issue-driven interdisciplinarity'. In addition, most of them combine multiple disciplines in their research. Overall, this means that in this case boundaries between peer communities are difficult to draw: they are fluid and indistinct because of varying individual backgrounds.

3 Experimental design and data

The experiment consisted of two steps: a review of proposed methods and a review of achieved results, both through a questionnaire. This paper discusses the outline of the experiment and the main results; details are presented as Electronic Supplementary Material. In each step EQUIP researchers reviewed two aspects of the papers:

1. the sources of uncertainty in methods (step 1, EMS 1) and results (step 2, EMS 2)
2. the uncertainty assessment methods used.

They also stated their confidence in their own review. The four papers in this special issue that present quantifications of different elements of the cascade of uncertainty were assessed in detail (Hanlon et al. 2013; Otto et al. 2003a; Saux Picart et al. 2013; Watson et al. 2014). The remaining four papers aim to inform uncertainty assessments by examining methodologies for assessment (Allen et al. 2013; Lorenz et al. 2013; Otto et al. 2013b; Calel et al. 2013). These papers have no quantitative results, so only a limited number of questions in the questionnaires were applicable. We invited all EQUIP project members to participate in the reviews; we made it a requirement of entry to this special issue that at least one response per submitted paper was received for each review. The authors themselves also completed the questionnaires for their own paper. Both reviews therefore came from three sources: the authors themselves (self-assessment), authors from other papers in this special issue, and EQUIP members who did not contribute to this special issue. We call the latter two categories 'reviewers'; when we use 'respondents' this includes both authors and reviewers. In total we had 11 respondents. The fact that the authors filled in the questionnaires for their own paper enables us to compare their specialist opinions with those of researchers from different backgrounds. Answers were anonymised for publication using an internal EQUIP numbering system.

For the first step, a project meeting was organized in August 2012 at which the draft papers were presented and discussed amongst authors and other EQUIP members. In addition to the scientific contents, in their presentation authors were asked to define the outcome variables that would form the bases for those consequence statements. We define outcome variables as 'the variable quantified in the study that is used, to assess impacts'. Authors were also asked to highlight what tangible environmental or societal consequences might be predicted from their

work. We define a consequence statement as ‘a quantitative or qualitative assessment of the consequences of changes in the outcome variables for human society’. The second step took place in October/November 2012, when more advanced draft papers were distributed. In addition, authors were asked to formulate consequence statement(s) based on the outcome variable(s) in their paper. At the same time Results Questionnaires (see EMS 2) were sent out to all respondents. We used many open questions in order to capture as much of the variety in opinions as possible and not limit the answers based on our own preconceptions. The more structured data collection through lists and tables were designed to cover as much of the research presented in the papers as possible.

4 Results

4.1 Review of methods

We wanted to know whether respondents agreed on:

1. sources of uncertainty in the outcome variable(s) and their relative contribution to total uncertainty (irrespective of the method used in the paper to quantify uncertainty).
2. the degree to which these sources of uncertainty had been adequately quantified in the paper
3. the merit of the method used to assess uncertainty (irrespective of whether it quantifies uncertainty)
4. whether more than one method had been used to assess uncertainties.

We designed the Methods Questionnaire accordingly (EMS 1). The results from this questionnaire are discussed below.

4.1.1 Sources of uncertainty

We find that the reviewers agree more on the sources of uncertainty for the two papers that develop and test climate simulation methods (H2013 and O2013) than for the two papers that assess impacts, and hence cover a larger part of the cascade of uncertainty (SP2013 and W2014). This confirms that the assessment of uncertainties in research methods becomes more difficult when uncertainties are located in more sources, which points at the need to involve more diversified expertise. We wanted to know whether the higher variation in the assessments in SP2013 and W2013 could be explained partly by differences in the degree of the respondents’ confidence in their assessment. We found that SP2013 has the lowest reviewers’ confidence score. This is explained by comments from many reviewers that ‘I cannot assess this because I do not know marine ecosystem modelling’. The marine ecosystem model used by SP2013 is apparently so far removed from the expertise held by most reviewers, and the climate elements of that paper so small, that they are not confident they can assess how the paper deals with uncertainty. In contrast, the other impact paper W2014 uses a relatively simple crop model and places greater influence on effect of input climate data, so the reviewers are more confident they can assess this paper.

We also found a difference in the self-assessment when two authors review their own paper. This suggests that there is an inherent bias associated with individual perspectives. We noticed that this bias can be systematic across multiple studies. For example, reviewer 3A listed the same sources of uncertainty for papers SP2013 and W2014. The extent to which this variation

is due to scientific training (reviewers 3A and 3B have backgrounds in physics and computer science respectively) compared to research history and/or personal preference is impossible to determine. However, the existence of prior beliefs regarding sources of uncertainty can be assessed, at least to some degree. To this end, reviewers 3A and 3B were invited to verbally explain the reasoning behind their scores for their own paper. This confirmed that fundamental points of view can lie underneath the reviews. In this case, confidence in model structure was seen as a significant source of uncertainty by the more experienced researcher who had been involved in its development, and therefore knew of its flaws and assumptions of the model, and judged less important by the more junior researcher. The two senior reviewers (3A and 9) indeed consistently rank models as a primary source of uncertainty across all papers, but several of the more recently-qualified reviewers do as well. In these limited data the researcher's experience alone does not therefore explain the differences in assessment.

4.1.2 *Methods of uncertainty assessment*

From the answers given to the Methods Questionnaire (1 and 2a/2b), it is clear that the interpretation of the methods of uncertainty assessment varies significantly between respondents. For example, when assessing whether the methods used quantify uncertainty, for one paper one respondent indicated that all important sources were majority quantified and another respondent indicated that none were. For all four papers there was significant variation between respondents in this answer. Variation also exists in self-assessment between authors of the same paper. We find, then, that reviewers and authors not only disagree on ranking of sources of uncertainty (as shown in above in Section 4.1.1), they also disagree on what was actually done. Two reasons may be behind the diversity of responses: reviewers have different opinions, or they understand the methods used differently. We do not have data to investigate the first possibility; this would require a round table discussion which we did not have time to organise. However, there is evidence of a structural difference of opinion: the two statisticians amongst the respondents did not enter one single 'majority quantified' whilst for the remaining respondents this answer formed between 20 and 63 % of entries.

4.2 Review of results

We wanted to know whether respondents agreed on:

1. the outcomes of the research, as formulated by the authors using consequence statements
2. whether the reviewed paper communicated well the uncertainty underlying the consequence statements
3. the relative importance of sources of uncertainty in the consequence statements
4. what decisions the outcomes may inform
5. the relative importance of sources of uncertainty in the decisions
6. which lessons could be learnt from the research, esp. concerning the treatment of uncertainty.

We designed the Results Questionnaire accordingly (EMS 2). Below we discuss the results and try to explain why we obtained these results.

We asked the authors to formulate consequence statements from their study, and the respondents to assess the veracity of those statements. Model inputs and model structure are assessed as the dominant sources of uncertainty across the four papers. Intrinsic and non-measurable stochastic variability is assessed as the least important source is When asked to

identify decisions that could be affected by the studies, respondents independently came up with similar answers: EU level assessment for agriculture & health adaptation policies (H2013), flood defence & insurance decisions (O2013), agricultural & marine management & legislation (SP2013), selection of model given weather input data characteristics (W2014). Again, respondents generally ranked model inputs and model structure highest amongst the sources of uncertainty in eventual decisions based on the studies. There are no obvious patterns in the scores for each respondent (this would mean that a reviewer gives the same score for the same source across all papers). It is remarkable that respondents rank the type of uncertainty that they have least possibility to assess quantitatively, or even know qualitatively, i.e. intrinsic and non-measurable stochastic variability, as the least important source. A perfect model would still have limits to predictability from chaos, as that is an intrinsic property of the system. Current models do include chaotic processes and chaotic limits to predictability, but that doesn't overcome the limitation.

The Results Questionnaires show that authors and reviewers alike generally have high confidence in the consequence statements. We hypothesised that this high confidence would be associated with high confidence in the methods for uncertainty assessment that were used, which had been the subject of the Methods Questionnaire. As a proxy for the latter we use the response that sources of uncertainty are adequately accounted for ('majority quantified,' in the language of the questionnaire). Figure 1 plots these two quantities, showing that low confidence in consequence statements is indeed associated with low confidence in uncertainty assessment (albeit for only 2 data points). However, high confidence in consequence

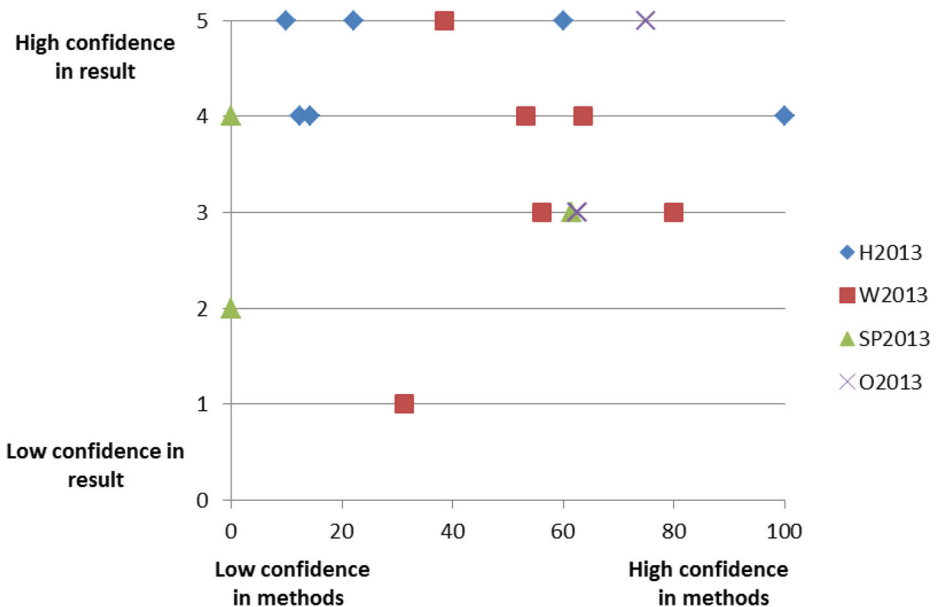


Fig. 1 Relationship between respondents' confidence in method and confidence in result for each assessed paper, Legend: The y-axis shows the mean respondents' confidence score (see EMS 2: Results Questionnaire question 1 and Table 2) for the consequence statement made by authors. The x-axis shows the confidence in the methods used to assess uncertainties in the outcomes. This was measured using the percentage of responses that indicated that the majority of uncertainties had been accounted for in the study (see EMS 1: Methods Questionnaire question 2a). Data are plotted for each unique combination of respondent and paper. Where multiple consequence statements were made for one paper, the first on the list was used

statements is not uniquely associated with high confidence in methods used – low confidence in methods can coincide with high confidence in results. This is particularly evident for H2013. Thus low confidence in uncertainty assessment does not preclude high confidence in, or indeed consensus about, consequence statements.

We also hypothesised that where authors had low confidence in their *own* assessment of their methods for uncertainty assessment, they would be unlikely to have high confidence in consequence statements. We found that in six of the eight cases respondents had high confidence in the consequence statement. However, in four of these six cases the respondents had a low confidence in their own assessment of uncertainty. Surprisingly, then, low confidence in the respondents' own assessment of their uncertainty assessment did not preclude confidence in consequence statements. Apparently their confidence in their methods to calculate the outcomes overrules their low confidence in the methods used to assess the uncertainties in those outcomes. This highlights the ability – or tendency – of an expert to use their training and judgement to assess whether or not a given result is correct, even if it is not provably correct.

4.3 Reviewers' comments and recommendations

Finally, we invited respondents to give written comments on the papers. A brief synthesis of common responses follows. The comment that re-occurs most frequently across the four papers discussed so far is 'limited number of climate model runs'. This applies especially, but not only, to the climate modelling papers H2013 and O2013. For the impacts papers SP2013 and W2014 other methods are core to the uncertainty assessment, such as Bayesian statistics, experimental design and model interpretation, and one method by itself does not suffice. More relevant to the theme of this paper, lessons learnt also include the possible transfer of the methods that were developed to other domains. This and other issues arising from the written responses are discussed in ESM 3.

Two of the three papers not discussed so far (Otto et al. 2013b; Allen et al. 2013) focus on the development of such generic methods. Allen et al. (2013) review a number of approaches to exploring uncertainty in climate model simulations and suggest a simple approach as a first step to making more explicit which approach authors are taking to make studies more comparable. All reviewers agree with these authors that it would be beneficial to have an agreed classification system to categorise uncertainties so papers would be more comparable, with one reviewer suggesting that this raises the question whether ensembles should be designed in a standardized way too to make the results of different research more comparable. Otto et al. (2013b) propose a new bounding methodology that helps to make and justify judgement on credibility of climate projections based on past performance of substantially dissimilar models with similar difficulty. The reviewers agree that this is a proposal worth exploring, but they also point out the problems with, and judgment required for, determining whether one task is similar in difficulty to another one. The main finding in Lorenz et al. (2013) is that the extent and explicitness of uncertainty reporting is not necessarily linked to the state of the knowledge base for a particular policy strategy but rather to the different policy styles in different national contexts. The reviewers appreciate this paper as potentially being useful for their own communications to policy makers, but stress that more openness about uncertainty would be good in all cases.

The respondents' recommendations for future work emphasise that quantifying uncertainties better (i.e. to use the same models but understand them better) is currently more important than reducing uncertainty in outcomes (i.e. to develop the models further). This would eventually result in recommendations for the area(s) in which reducing uncertainty through further modelling would be most beneficial. Paradoxically, this is not the objective of

many funders, who seek to develop new models and produce new results, rather than exploring existing models and results better. It would seem that scientists need to set the agenda on this issue, which was indeed one of EQUIP's objectives and one to which we hope this paper will contribute.

5 Discussion: assessing consensus about uncertainties

When we developed the EQUIP review process we expected to find that researchers with different disciplinary backgrounds and experience would have different understandings of uncertainty. The analysis has shown that there is indeed a spread in opinions between EQUIP members about how to assess uncertainty and where to locate major sources of uncertainty. We found that reviewers generally agree with authors on the ranking of overall sources of uncertainty, but disagree amongst each other and with the authors on the exact source, particularly when more elements of the uncertainty cascade are included in the research. More surprisingly, we also found that reviewers do not agree on the interpretation of the methods for assessing uncertainty used in each paper. This disagreement appears most clearly from their assessment of whether a particular method fully quantifies uncertainty, which can range amongst reviewers from 'not at all' to 'fully' for the same paper. This finding suggests both difficulties in interpreting what has been done and individual bias.

We explored the reasons for this disparity and found that there is indeed some evidence of individual bias, with reviewers ranking the same sources of uncertainty consistently between papers. This bias was to some extent related to research experience (more experienced researchers doubt models more) and to discipline (the two statisticians are most strict about the meaning of 'quantification of uncertainty'). This result begs an important question: given that near-peers did not always understand or agree on the meaning of a particular uncertainty assessment, how can we expect understanding or agreement amongst those who are further removed? The emphasis placed by respondents on improving input data and number of model runs in future research, and the low importance they assign to intrinsic and non-measurable stochastic variability, indicates that EQUIP members focus on reducing methodological uncertainty (cf. the typologies in Section 2). In this reasoning, it is ultimately possible to arrive at models that represent reality. Knowledge derived in this way is typically represented by one value with error bands, for example 'yields decrease by 10–70 %', and a focus on calculating probabilities. We advocate another approach where models (and also data) are treated as tools from which information is extracted, rather than as competing attempts to represent reality (cf. Challinor et al. 2013). In this view consequence statements that describe processes and trade-offs are better way to report results (see also Beven 2006, 2012). For example: 'higher temperatures will reduce the time to maturity of crops, thus reducing yield. Model results suggest that increases in rainfall will compensate for this in 40–60 % of cases'. The IPCC 5th Assessment Report guidance for treatment of uncertainties seems to choose this approach. It suggests that independently evaluating the degrees of certainty in both causes and effects should be considered for findings (effects) that are conditional on other findings (causes) (Mastrandrea et al. 2010 p.1).

Different points of view are, as we have shown here, a significant source of variation in assessments of uncertainty. There is sufficient spread in the assessments to suggest that both specialist knowledge and individual variation in opinion are important in determining researchers' assessment of key uncertainties. 'If we could elicit probability distributions for parameters from each expert without any bias, we would still find that different experts disagree. Disagreements arise from different underlying knowledge sets and from different

beliefs about fundamental properties of the system being described (economic system, climate system, etc.). Unfortunately, there is no commonly accepted methodology for combining multiple expert judgments (Webster 2003 p.4). The challenge, then, is to see whether these differences can complement each other and, especially when they contradict, how further discussions can yield better insight in overall uncertainties. In EQUIP we appreciated from the start the importance of being explicit about the uncertainties that are quantified in any given study, and about the methods used for that assessment. We designed and circulated a common uncertainty reporting format (ESM 3) in order to prompt thinking about the implication of methodological decisions for the results produced. It has the advantage of providing a structure for framing the various assumptions needed in making an assessment of uncertainties and in doing so provides an audit trail for later analysis (Beven et al. 2010; Beven et al. 2011). We recommend that similar methods to these are used in future studies, and present our general recommendations in Section 7.

6 Drawing on science and technology studies

Issues surrounding the generation, interpretation and use of uncertain information are not unique to the topic of climate change. Science and Technology Studies (STS) has a long history of analysing the ways in which scientists produce knowledge for policy, including how they deal with uncertainty, on which we now draw to put out findings in a wider perspective.

As we have seen above, the quantification of uncertainty is a fairly specialised topic in which misinterpretation of techniques and terminology is easy. Terms such as ‘forcing’, ‘bias correction’, ‘transformation function’ are easily misconstrued by lay people (Kerr 1994). If the specialists disagree then this causes a problem, since few are qualified to arbitrate, manage or synthesise the various views. Moreover, the presence of the contrarians who are ready to make the most of uncertainty and disagreement may have inhibited the expression of such intra-peer community differences (Shackley et al. 1999). They argue that openness about, and explanation of, uncertainties could potentially reduce some of the political effectiveness of the contrarians, because uncertainty and limitations in climate knowledge would cease to be a prime cause for disbelief or policy inaction (Shackley et al. 1999 p.448). Similarly, Webster (2003) argues that whilst policy-oriented science based on consensus has been highly effective in the IPCC case for describing the areas that are known relatively well, this is less well suited to treating uncertainty because much of the disagreement and uncertainty comes directly from the current lack of scientific consensus (Webster 2003). These findings also suggest that decisions about adaptation to the impacts of climate change could be made in a different way (Wilby and Dessai 2010; Prudhomme et al. 2010; Beven 2011).

We are therefore not the first to recognise that a single quantitative measure for uncertainty, such as probability, is not always appropriate in climate change policy-oriented research. Stirling advocates to ‘keep it complex’ (Stirling 2010) and shows that probability is only appropriate when knowledge about management possibilities and about probabilities are both unproblematic. The IPCC guidelines also recognise that probabilities may not always be appropriate: depending on the nature of the evidence evaluated, teams have the option to quantify the uncertainty in the finding probabilistically or present an assigned level of confidence in the validity of a finding, based on the type, amount, quality, and consistency of evidence and the degree of agreement (Mastrandrea et al. 2010 p.1). What then does ‘keeping it complex’ mean for the conduct of policy-relevant science? We address this question in our conclusions.

7 Conclusion: the characteristics of post-normal policy-oriented science

Our experiment supports the argument that that policy-relevant science, including uncertainty assessment and IPCC-style assessments, needs to be done differently from science aimed at an academic audience only. This ‘new’ kind of knowledge production has been labelled post-normal (Funtowicz and Ravetz 1993); Mode-2 (Gibbons et al. 1994), and inter- or transdisciplinary (Klein et al. 2001; Robinson 2007). It is different from ‘ordinary’ science because it is problem-oriented and requires ‘integration, interactivity and emergence, reflexivity, and strong forms of collaboration and partnership’ (Robinson 2007 p.70). This science results in policy advice that incorporates input from different disciplines that need to mutually adjust and work towards this one goal. From the EQUIP experience it is clear that one of the challenges of this kind of research will be to deal with different interpretations of uncertainty, in addition to other challenges identified e.g. by Robinson (2007).

While a lot of effort is made to develop uncertainty communication protocols for policy makers and society as a whole (Van der Sluijs et al. 2003; Faulkner et al. 2007; Morgan et al. 2009; Mastrandrea et al. 2010; Beven et al. 2011), less attention is paid to who is participating in the assessment and how they arrive at their conclusion. We need research in which teams with people from different backgrounds work together on policy-relevant questions. They should spend enough time discussing uncertainties and document ranges, reasons and contingencies so trade-offs can be better assessed. This research will take longer and produce fewer, but more comprehensive outputs, which has implications for academic assessment procedures. As part of this process, uncertainties will be dealt with as recommended above, making it easier to assess (dis)agreement on sources, magnitudes, and types. It differs from prevailing approaches in that it makes these influences more rigorously explicit and thereby more democratically accountable (Stirling 2010). However, some caution is warranted, since ‘a move towards plural and conditional expert advice is not a panacea’ (ibid.). It cannot promise escape from the deep intractabilities of uncertainty or the power political character of decision making (Wesselink and Hoppe 2011). At the least, it will go some way to present a balanced view that recognizes the limitations of ‘the predictive capabilities of science, the importance of predictions in rational decision processes, and the potential for political consensus on comprehensive policy’ (Brunner 1996 p.124–125).

Based on our analysis, we formulate the following recommendations for quantifying uncertainty. They contribute to making explicit the inherent limitations of a given piece of research, and thus reducing false confidence and increasing utility. Some of these echo and develop the analysis of Challinor et al. (2013).

1. Consequence statements that describe processes and indicate a range of sources and assessments of errors are a way to represent the outcomes of this approach. For example: ‘warmer temperatures will reduce the time to maturity of crops, thus reducing yield. Model results suggest that increases in rainfall will compensate for this in 40–60 % of cases’
2. Conditionalisation of projections in order to identify the assumptions upon which the results of a study are contingent. A common uncertainty reporting format can be used to make explicit the identified conditions. For example, the framework in EMS 3 was designed to prompt thinking about the implication of methodological decisions for the results produced.
3. Reporting multiple, rather than single, assessments of the confidence placed by experts in particular predictions. Such reporting would make explicit a range of views, which we have demonstrated can be significant.

Acknowledgements This study was part of the EQUIP project <http://www.equip.leeds.ac.uk> which was financially supported by UK Natural Environment Research Council (grant NE/H003525/1).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Allen M, Lopez A, Otto F, Suckling E, Rowlands D (2013) Using observations to constrain climate models forecasts. *Climatic Change* this issue
- Beven, K J (2006) A manifesto for the equifinality thesis. *J. Hydrology* 320:18–36.
- Beven KJ (2011) I believe in climate change but how precautionary do we need to be in planning for the future? *Hydrol Process* 25:1517–1520
- Beven, K J (2012) Causal models as multiple working hypotheses about environmental processes, *Comptes Rendus Geoscience, Académie de Sciences, Paris*, doi:10.1016/j.crte.2012.01.005
- Beven KJ, Leedal DT, Alcock R (2010) Uncertainty and good practice in hydrological prediction. *Vatten* 66:159–163
- Beven, K J, Leedal, D T, McCarthy, S (2011) Framework for Assessing Uncertainty in Fluvial Flood Risk Mapping, UK FRMRC Research Report SWP1.7 http://web.sbe.hw.ac.uk/frmrc/downloads/Framework_for_Assessing_Uncertainty_Fluvial_Flood_Risk_Mapping_WP1_7.pdf
- Brown G E Jr. (1992) Global Change and the New Definition of Progress. *Geotimes* (June) 19–21
- Brunner RD (1996) Policy and global change research: a modest proposal. *Clim Chang* 32:121–147
- Calel R, Stainforth D, Dietz S (2013) Tall tales and Fat tails: The science and economics of extreme warming. *Climatic Change* this issue
- Challinor AJ, Osborne T, Morse A, Shaffrey L, Wheeler T, Weller H, Vidale PL (2009) Methods and resources for climate impacts research: achieving synergy. *Bull Am Meteorol Soc* 9:836–848
- Challinor AJ, Stafford Smith M, Thornton P (2013) Use of agro-climate ensembles for quantifying uncertainty and informing adaptation. *Agric For Meteorol* 170:2–7
- Curry J (2011) Reasoning about climate uncertainty. *Clim Chang* 108:723–732
- Faulkner H, Parker D, Green C, Beven K (2007) Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner. *Ambio* 16(7):692–703
- Funtowicz SO, Ravetz JR (1993) Science for the post-normal age. *Futures* 7:735–755
- Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, Trow M (1994) *The New production of knowledge: the dynamics of science and research in contemporary societies*. Sage, London
- Grundmann R (2013) ‘Climategate’ and the scientific ethos. *Science Technology and Human Values* 38(1):67–93
- Hanlon H M, Hegerl G C, Tett S F B and DM Smith (2013) Near-term prediction of impact-relevant heatwave indices. *Climatic Change* this issue
- Hulme M, Mahony M (2010) Climate change: what do we know about IPCC? *Prog Phys Geogr* 34:705–718
- Jones RN (2011) The latest iteration of IPCC uncertainty guidance - an author perspective. *Clim Chang* 108:733–743
- Kandlikar M, Risbey J, Dessai S (2005) Representing and communicating deep uncertainty in climate-change assessments. *Compt Rendus Geosci* 337:443–455
- Kerr R (1994) Climate modeling’s fudge factor comes under fire. *Science* 265:1528
- Klein JT, Grossenbacher-Mansuy W, Häberli R, Bill A, Scholz RW, Welti M (eds) (2001) *Transdisciplinarity: joint problem solving among science, technology, and society*. Birkhäuser Verlag AG, Basel
- Knutti R, Sedláček J (2012) Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change* doi:10.1038/nclimate1716
- Knutti R, Allen MR, Friedlingstein P, Gregory JM, Hegerl GC, Meehl GA, Meinshausen M, Murphy JM, Plattner G-K, Raper SCB, Stocker TF, Stott PA, Teng H, Wigley TML (2008) A review of uncertainties in global temperature projections over the twenty-first century. *J Clim* 21:2651–2663
- Leiserowitz A A, Maibach E W, Roser-Renouf C, Smith N, Dawson E (2012) *Climategate, Public Opinion, and the Loss of Trust*. *American Behavioral Scientist* doi:10.1177/0002764212458272
- Lorenz S, Dessai S, Paavola J, Forster P M (2013) The communication of scientific uncertainty in European National Adaptation Strategies. *Climatic Change* this issue
- Mastrandrea MD, Field CB, Stocker TF, Edenhofer O, Ebi KL, Frame DJ, Held H, Kriegler E, Mach KJ, Matschoss PR, Plattner G-K, Yohe GW, Zwiers FW (2010) Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. IPCC, Geneva

- Morgan G M, Dowlatabadi H, Henrion M, Keith D, Lempert R, McBride S, Small M, Wilbanks T (2009) Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decision Making. U.S. Climate Change Science Program and the Subcommittee on Global Change Research
- Moss RH, Schneider SH (2000) Uncertainties in the IPCC TAR: recommendations to lead authors for more consistent assessment and reporting. IPCC, Geneva
- Otto F, Rosier S M, Allen M R, Massey N R, Rye C J, Imbers Quintana J (2003a) Attribution of high precipitation events in summer in England and Wales over the last decade. *Climatic Change* this issue
- Otto F, Ferro C, Fricker T, Suckling E (2013b) On judging the credibility of climate projections, *Climatic Change* this issue
- Petersen A (2012) *Simulating nature: a philosophical study of computer-model uncertainties and their role in climate science and policy advice*, 2nd edn. CRC Press, Abingdon
- Prudhomme C, Wilby RL, Crooks S, Kay AL, Reynard NS (2010) Scenario-neutral approach to climate change impact studies: application to flood risk. *J Hydrol* 390(3):198–209
- Risbey JS, Kandlikar M (2007) Expressions of likelihood and confidence in the IPCC uncertainty assessment process. *Clim Chang* 85:19–31
- Robinson J (2007) Being undisciplined: transgressions and intersections in academia and beyond. *Futures* 40:70–86
- Saux Picart S, Allen J I, Butenschon M, Artoli Y, de Mora A L, Wakelin S, Holt J (2013) What can ecosystem models tell us about the risk of eutrophication in the North Sea? *Climatic Change* this issue.
- Schneider SH (2001) What is ‘dangerous’ climate change? *Nature* 411:17–19
- Shackley S, Risbey J, Stone P, Wynne B (1999) Adjusting to policy expectations in climate change modelling. An interdisciplinary study of flux adjustments in coupled atmosphere–ocean general circulation models. *Clim Chang* 43:413–454
- Skrydstrup M (2013) Tricked or troubled natures? How to make sense of “climategate”. *Environmental Science and Policy* 28:92–99
- Stainforth DA, Aina T, Christensen C, Collins M, Faull N, Frame DJ, Kettleborough JA, Knight S, Martin A, Murphy JM, Piani C, Sexton D, Smith LA, Spicer RA, Thorpe AJ, Allen MR (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* 433:403–406
- Stirling A (2010) Keep it complex. *Nature* 468:1029–1030
- Van Asselt MBA, Rotmans J (2002) Uncertainty in integrated assessment modelling: from positivism to pluralism. *Clim Chang* 54:75–105
- Van der Sluijs J, Risbey J, Klopogge P, Ravetz J, Funtowicz S, Corral Quintana S, Guimaraes Pereira A, De Marchi B, Petersen A, Janssen P, Hoppe R, Huijs S (2003) RIVM/MNP Guidance for Uncertainty Assessment and Communication Detailed guidance. RIVM/MNP De Bilt
- Watson J, Challinor A J, Fricker T, Smith D, Ferro C (2014) Identifying and using crop-relevant weather information for regional scale crop projections: A case study of maize yields in France. *Climatic Change* this issue
- Webster M (2003) Communicating climate change uncertainty to policy-makers and the public. *Clim Chang* 61: 1–8
- Wesselink AJ, Hoppe R (2011) If post-normal science is the solution, what is the problem? The politics of activist science. *Science, Technology and Human Values* 36(3):389–412
- Wilby RL, Dessai S (2010) Robust adaptation to climate change. *Weather* 65(7):180–185