

Variations on Reliability: Connecting Climate Predictions to Climate Policy

Leonard A. Smith^{1,2} and Arthur C. Petersen^{1,3,4}

¹ Centre for the Analysis of Time Series (CATS), London School of Economics and Political Science (LSE)

² Pembroke College, Oxford

³ Institute for Environmental Studies (IVM), VU University Amsterdam

⁴ PBL Netherlands Environmental Assessment Agency

1. Introduction

This chapter deals with the implications of uncertainty in the practice of climate modelling for communicating model-based findings to decision-makers, particularly high-resolution predictions¹ intended to inform decision-making on adaptation to climate change. Our general claim is that methodological reflections on uncertainty in scientific practices should provide guidance on how their results can be used more responsibly in decision support. In the case of decisions that need to be made to adapt to climate change, societal actors, both public and private, are confronted with deep uncertainty. In fact, it has been argued that some of the questions these actors may ask ‘cannot be answered by science’.² In this chapter, the notions of ‘reliability’ are examined critically, in particular the manner(s) in which the reliability of climate model findings pertaining to model-based high-resolution climate predictions is communicated. A broader discussion of these issues can be found in the chapter by Beck (this volume).

Findings can be considered ‘reliable’ in many different ways. Often only a statistical notion of reliability is implied, but in this chapter we consider wider variations in the meaning of ‘reliability’, some more relevant to decision support than the mere uncertainty in a particular calculation. We distinguish between three dimensions of ‘reliability’ (see Section 2) – *statistical reliability*, *methodological reliability* and *public reliability* – and we furthermore understand reliability as reliability *for a given purpose*, which is why we refer to the reliability of particular findings and not to the reliability of a model, or set of models, *per se*.

¹ The distinction between ‘prediction’ and ‘projection’ is arguably artificial if not simply false. As discussed below all probability forecasts, indeed all probability statements, are conditional on some information.

² Our focus in this chapter is on ‘today’s science’ and questions that might, or might not, be answered via a probability distribution. Weinberg (1972) used the term ‘trans-science’ to denote the situation of questions being asked to science that science cannot answer. Funtowicz and Ravetz (1993) refer to ‘post-normal science’ as the appropriate problem-solving strategy for this situation given that one still wants to make responsible use of the scientific knowledge available (see also Petersen et al. 2011 and Millner et al 2013).

At times, the statistical notion of reliability, or ‘statistical uncertainty’, dominates uncertainty communication. One must, however, seriously question³ whether the statistical uncertainty adequately captures the ‘relevant dominant uncertainty’ (RDU). The RDU can be thought of as the most likely known unknown limiting our ability to make a more informative (perhaps narrower, perhaps wider, perhaps displaced) scientific probability distribution on some outcome of interest; perhaps preventing even the provision of a robust statement of subjective probabilities altogether. Here we are particularly interested in the RDU in simulation studies, especially in cases where the phenomena contributing to that uncertainty are neither sampled explicitly nor reflected in the probability distributions provided to those who frame policy or those who make decisions. For the understanding, characterisation and communication of uncertainty to be ‘sufficient’ in the context of decision-making we argue that the RDU should be clearly noted. Ideally the probability that a given characterisation of uncertainty will prove misleading to decision-makers should be provided explicitly (Smith and Stern 2011).

Science tends to focus on uncertainties that can be quantified today, ideally reduced today. But a detailed probability density function (PDF) of the likely amount of fuel an aircraft would require to cross the Atlantic is of limited value to the pilot if in fact there is a good chance that metal fatigue will result in the wings separating from the fuselage. Indeed, focusing on the ability to carry enough fuel is a distraction when the integrity of the plane is thought to be at risk. The RDU⁴ is the uncertainty most-likely to alter the decision-maker’s conclusions given the evidence, while the scientist’s focus is understandably on some detailed component of the big picture. How can one motivate the scientist to communicate the extent to which her detailed contribution has both quantified the uncertainty under the assumption that the RDU is of no consequence, and also provided an idea of the timescales, impact and probability of the potential effects of the RDU? The discussion on the main case analysed in this chapter suggests that failure to communicate the relevant ‘weakest link’ is sometimes under-appreciated as a critical element of science-based policy-making.

Arguably the Intergovernmental Panel on Climate Change (IPCC) has paid too little attention to communicating the RDU in simulation studies, even though within climate science, and particularly within the IPCC, there has been increased attention on dealing with uncertainty in climate models over the last 15 years or so (see e.g. Randall and Wielicki 1997; van der Sluijs 1997; Moss and Schneider 2000; Petersen 2000, [2006] 2012; IPCC 2005; Risbey and Kandlikar 2007; Swart et al. 2009; Hulme and Mahony 2010; Mastrandrea et al. 2010). It remains questionable whether this increased attention has led to a sufficient understanding, characterisation and communication of uncertainty in model-based findings shared with decision-makers. Early model studies were often explicit that the quantitative results were not to be taken too seriously.

³ Statisticians have long recognised that confidence intervals on model-based forecasts are exceeded more often than theory suggests they should be. Reproducibility of a specific calculation, robustness of the statistics of a specific model, and belief in the fidelity of that model statistic (that it reflects reality) are three distinct things.

⁴ RDU will be used as both singular and plural. It may be thought of as a set of uncertainties where one dominates, as a many-headed Hydra where, as the science advances and removes one, another immediately replaces it.

While the IPCC has led the climate science community in codifying uncertainty characterisation, it has paid much less attention to specifying the RDU. The focus is at times more on ensuring reproducibility of computation than on relevance (fidelity) to the Earth's climate system, in fact it is not always easy to distinguish which of these two are being discussed. Instead, the attention has mainly been on increasing the transparency of the IPCC's characterisation of uncertainty. For instance, the latest IPCC guidance note for lead authors who are involved in the writing of the Fifth Assessment Report (Mastrandrea et al. 2010), which was endorsed by the IPCC governments at its 33rd session (10–13 May 2011, Abu Dhabi), like its two precursors emphasises that uncertainties need to be communicated carefully, using calibrated language for key findings, and that lead authors should provide traceable accounts describing their evaluations of evidence and agreement. While the basic message of the latest IPCC guidance thus does not differ significantly from the first (Moss and Schneider 2000), the IPCC has learned from past experience and from the recent evaluation by the InterAcademy Council (IAC 2010), and has made more clear when to use which calibrated uncertainty terminology. Also, given the turmoil surrounding the IPCC in 2010, it is understandable that its lead authors have been asked to put more effort in providing traceable accounts of the main findings and their uncertainty qualifications in the Fifth Assessment Report (AR5), which is due for publication in 2013 and 2014. Such accounts were largely lacking in the past assessment rounds (Petersen [2006] 2012; IAC 2010; Strengers et al. 2013). Still, being transparent, while certainly a good thing in itself, is not the same as communicating the RDU for the main findings.

The IPCC is not the only effort climate scientists are engaged in to assess and communicate findings on future climate change to decision-makers. In countries all over the world, ongoing projects aim to tailor information from climate models for use by decision-makers. These projects and their dissemination ultimately feed back into the IPCC assessment process, which periodically assesses the (peer-reviewed) literature in the context of climate science. In this chapter we critically reflect on one particular project which ran in the UK. This UK project produced the United Kingdom Climate Projections 2009 (UKCP09) and exemplifies perhaps the largest real-world case to date of climate decision support at very high-resolution (post-code or zip-code resolution through the end of this century) based upon climate models, within a so-called 'Bayesian' framework. A great deal has been learned about error, insight, and decision-making in the course of this ground-breaking project; one of the aims of this chapter is to explore how insights gained in this project can be used to minimise future confusion, misunderstanding and errors of misuse, avoiding spurious precision in the probabilistic products while maintaining engagement with user communities.

A central insight is to note that when the level of scientific understanding is low, ruling out aspects of uncertainty in a phenomenon without commenting on less well understood aspects of the same phenomenon can ultimately undermine the general trust decision-makers place in scientists (and thus lower the public reliability of their findings). Often epistemic uncertainty or mathematical intractability means that there is no strong evidence that a particular impact will occur; simultaneously there may be good scientific reason to believe the probability of some significant impact is nontrivial, say, greater than 1 in 200. How do we stimulate insightful discussions of things we can neither rule out nor rule in with precision, but which would have significant impact were they to occur? How do we avoid obscuring the importance of things we cannot rule out by placing (burying)

the warning on an agreed high impact RDU in a long list of standard, inescapable caveats? How can those throughout the evidence chain from science through modelling to analysis, consulting, and ultimately decision-making, be incentivised to stress the weakest links in their contributions and in the chain itself? Doing so would allow improved decision-making⁵ both in the present and in the future. This chapter addresses these questions and is structured as follows.

In Section 2 we introduce the three variations on the meaning of reliability. In Section 3 model-based probabilities are cast within a⁶ Bayesian framework, stressing the importance of the ‘information’ *I*, the evidence assumed to be effectively true in order to convey any decision-relevance to the probability distributions generated; the extent to which climate-model derived probability distributions can be considered methodologically reliable is put into question. Section 4 notes the need to resist shifting definitions from what counts to what can be counted today. The section focuses on the UKCP09 report, noting several passages and presentations which point to fundamental limitations in the decision-relevance of its outputs, but not in a manner likely to be accessed by users of its outputs. Examples are given where distinguishing obfuscations from clarifications of fact is nontrivial. Section 5 returns to the question of error in science, of how uncertainty and progress in climate science differ from other decision-relevant sciences, including the challenges posed by an active anti-science lobby. Climate science can likely be most quickly advanced and most usefully employed when its errors, shortcomings, and likely failures are laid as bare as can be. How are all actors in the chain - from science, through modelling, criticism, decision support to policy-making - best incentivised to use the information available today well and maximise the likely information available tomorrow? Recognising and clearly identifying the Relevant Dominant Uncertainty in each domain is a useful start. The Dutch approach towards climate scenarios is compared with the UK approach in passing, and argued to be more appropriate.

In short, this chapter addresses the question of whether the practice of climate modelling and communicating results to decision-makers suffers from shortcomings that may lead to problems at the interfaces between science, policy and society (cf. chapter 3, Beck).

2. Three notions of reliability

Climate science is confronted with ‘deep uncertainty’ (cf. Kandlikar et al. 2005; called ‘ambiguity’ in Smith and Stern, 2011, and ‘Knightian uncertainty’ in economics). This is true both when one uses climate models to investigate the dynamics of the climate system in order to generate understanding and when one uses them to predict particular (aspects of) future states of the climate system. Deep uncertainty in findings based on climate models derives from two distinct sources:

⁵ How might a consultant who communicates the RDU openly maintain the attention and interest of decision-makers? How is one to avoid the false claim that fair and needed scientific uncertainty (exposed by discussion of the RDU) suggests that the science is too uncertain to justify action at this point in time?

⁶ Good ([1983] 2009) noted that there are 46,656 varieties of Bayesian; UKCP09 may well have introduced at least one more variety.

- A. There are fundamental limitations in our own cognitive and scientific abilities to understand and predict the climate system probabilistically, even when given the future pattern of emissions (which is a form of ‘epistemic uncertainty’).
- B. There are fundamental limitations to predictability inherent in the nature of the climate system itself and its surroundings (which refers to ‘ontic uncertainty’, cf. Petersen [2006] 2012).

A implies that:

- 1) The difference in the range of model outcomes under different models may be large.
- 2) One cannot justify the assignment of statistical meaning to a distribution of model outcomes in reference to the real world; this distribution can at best be regarded as a reflection of ‘model scenario uncertainty’ or a ‘non-discountable envelope’ (Stainforth et al. 2007a).
- 3) It is crucial to recognise and acknowledge ignorance and its implications: one knows that the models do not capture some dynamical features which are expected to become important in the context of climate-change studies. Additional features will play this role as a function of lead time. It is not clear what is the most reliable methodology for climate modelling (it is clear, though, that all climate models suffer from nontrivial methodological flaws, in terms of the limits of the technology of the day, including the favoured theoretical and empirical underpinning of assumptions, and similar hardware constraints).
- 4) Some model assumptions are strongly influenced by the particular epistemic values (general and discipline-bound) and non-epistemic values (socio-cultural and practical) which come into play in the design of both model and experiment.

And *B*, in combination with *A*, implies that:

- 5) Surprises are to be expected; unanticipated, perhaps inconceivable, transitions and events cannot be excluded.

The reader is referred to a book-length treatment of uncertainties in climate simulation for a more detailed explanation and underpinning of these claims and how they have played out in climate-policy advice to date (Petersen [2006] 2012).

Additional reflection is due both in the practice of climate science and in the practice of assessing and communicating that science to decision-makers. Ultimately, uncertainty assessment and risk management rely upon human expert judgment. This makes it even more important in assessments to provide a traceable account of how expert judgments on uncertainty have been reached. And since the reliability of findings are contingent on the reliability of the selected experts themselves, establishing the

overall reliability of these findings for decision-makers becomes an even more daunting task.

We here propose to distinguish three dimensions of reliability:

- 1) statistical reliability (reliability₁),
- 2) methodological reliability (reliability₂), and
- 3) public reliability (reliability₃).

Each of these dimensions plays a role in the public debate about ‘the’ reliability of our quantitative vision of the climate in the future, including model-based high-resolution climate predictions. They deserve to be consciously dealt with by the community of climate scientists.

2.1. *Statistical reliability (reliability₁)*

A statistical uncertainty distribution, or statistical reliability (denoted by reliability₁), can be given for findings when uncertainty can be adequately expressed in statistical terms, e.g., as a range with associated probability (for example, uncertainty associated with modelled internal climate variability). Statistical uncertainty ranges based on varying real numbers associated with models constitute a dominant mode of describing uncertainty in science. One cannot immediately assume that the model relations involved offer adequate descriptions of the real system under study (or even that one has the correct model class), or that the observational data employed are representative of the target situation. Statistical uncertainty ranges based on parameter variation are problematic in climate science, however, since the questions regarding the best model structure for climate models and the extent to which we can access it today remain open, as is argued in Section 3.

Both objective and subjective probabilities have been used in expressing reliability₁. Confusingly sometimes combinations also occur. For instance, the IPCC has used a combination of objective and subjective probability in arriving at its main findings on the attribution of climate change to human influences.⁷ Even more confusingly, different meanings of ‘subjective probability’ are often confounded, especially within competing Bayesian analyses. We point to the clear distinctions made by I.J. Good. Of particular importance is the distinction between the Bayesian’s target (which is the Subjective probability of an (infinite) rational org (Good [1983] 2009)) from the assortment of other types of probability in hand.

⁷ The finding that most of the observed warming over the last 50 years is attributable to the anthropogenic emissions of greenhouse gases was qualified in 2001 as ‘likely’ (defined as a ‘judgmental estimate’ of a 66–90% chance of being correct) and in 2007 as ‘very likely’ (a >90% chance). How objective and subjective probabilities were combined in these uncertainty qualifications is analysed in detail by Petersen ([2006] 2012).

2.2. *Methodological reliability (reliability₂)*

There are both epistemological and practical problems with maintaining a strong focus on the statistical reliability (reliability₁) of model findings. We know that models are not perfect and never will be perfect. Especially when extrapolating models into the unknown, we wish ‘both to use the most reliable model available and to have an idea of how reliable that model is’ (Smith 2002, Smith 2006), but the reliability of a model as a forecast of the real world in extrapolation cannot be established. There is no statistical fix here; one should not confuse the range of diverse outcomes across an ensemble of model simulations (projections), such as used by the IPCC, with a statistical measure of uncertainty in the behaviour of the Earth. This does not remotely suggest that there is no information in the ensemble or that the model(s) is worthless, but it does imply that each dimension of ‘reliability’ needs to be assessed.⁸

And so another precise definition of ‘reliability’ results from consideration of the question: What is the referent and what is the purpose of a particular modelling exercise? If the referent is the real world (and not some universe of mathematical models) and the purpose is to generate findings about properties of the climate system or prediction of particular quantities, then ‘reliability₁’ is uninformative: one can have a reproducible, well-conditioned model distribution which is reliable₁ without reliable being read as informative regarding the real world.

Two distinct varieties of statistical uncertainty ranges are often estimated, one from comparing the simulation results with measurements – provided that accurate and sufficient measurements are available – another from sensitivity analysis within the model – provided that the accuracy of the different components in simulation practice (e.g. model parameters) are well-defined and known. Of course in climate-change modelling, one does not have verification of outcomes as one has in weather forecasting. Realisations of weather-like simulations can continuously be compared with measurements and predictive skill can be determined, which is not possible in climate simulation.

Models, when they do apply, will hold only in certain circumstances. We may, however, be able to identify shortcomings of our model even within the known circumstances and thereby increase our understanding (Smith 2002). As was observed in the previous subsection, a major limitation of the statistical definition of reliability is that it is often not possible to establish the accuracy of the results of a simulation or to quantitatively assess the impacts of different sources of uncertainty. Furthermore, disagreement (in distribution) between different modelling strategies would argue against the reliability of some, if not all, of them. An alternative is therefore to define the reliability of findings based on climate models in more pragmatic terms. As Parker (2009) has shown, in order to establish the ‘adequacy-for-purpose’ of a model in a particular context, scientists rely not simply on the statistical proximity of model results to, for instance, an historical dataset of the quantity of interest, but use a much more elaborate argumentation. This argumentation includes an assessment of the ‘methodological reliability’ of the model, for instance of the quality (fidelity) of the

⁸Indeed, this applies to all simulation models as well as ‘modes of thought’ (cf. Whitehead’s ‘fallacy of misplaced concreteness’).

representation of a particular dynamic process that is thought to be of importance, for example, in modelling particular future changes.

So, one also holds *qualitative* judgments of the relevant procedures to derive findings on the basis of climate models, that is, of the methodological quality of the modelling exercise. A methodological definition of reliability, denoted by reliability₂, indicates the extent to which a given output of a simulation is expected to reflect its namesake (target) in reality. The nature of this reflection may be even further restricted in terms of the particular purpose for which those outputs are being employed, specifically its methodological quality. The methodological quality of a simulation exercise derives from the methodological quality of the different elements in simulation practice, *given the particular purpose to which the models under consideration are being put*. It depends, for example, not only on how adequately the theoretical understanding of the phenomena of interest is reflected in the model structure, but also, for instance, on the empirical basis of the model, the numerical algorithms, the procedures used for implementing the model in software, the statistical analysis of the output data, and so on.

It is rarely, if ever, a straightforward affair to determine the methodological quality of a finding, the extent to which the outcome of a scientific simulation will reflect future events in the real world. Ultimately, it may not be possible to do more than agree on the probability that the results of simulation will prove to be mis-informative (see Smith and Stern 2011 and Smith 2013). Reliability₂ has a qualitative dimension, and the (variable) judgment and best practice of the scientific community provides only a reference for its extraction. It depends, for instance, on how broadly one construes the relevant scientific community and what one perceives as the purpose of the model. The broader the community, the more likely it is that the different epistemic values held by different groups of experts could influence the assessment of methodological quality. Criteria such as (1) theoretical basis, (2) empirical basis, (3) comparison with other simulations, (4) adequacy of the computational technology and (5) acceptance/support within and outside the direct peer community can be used for assessing and expressing the level of reliability₂ (see Petersen [2006] 2012).

2.3. *Public reliability (reliability₃)*

In addition to the qualitative evaluation of the reliability of a model, increasingly also the reliability of the modellers⁹ is taken into account in the internal and external evaluation of model results in climate science. We therefore introduce the notion of reliability₃ of findings based on climate models, which reflects the extent to which scientists in general and the modellers in particular are trusted by others.

The situation for climate scientists and climate modellers¹⁰ concerning the use of high-resolution model-based predictions is different from that of, for instance, doctors and the ‘placebo effect’: in the latter situation an over-confident doctor may increase the chances that a placebo works. In a similar situation, climate scientists might choose to be

⁹ And that of the scientists.

¹⁰ There are, of course, climate scientists who model and climate modellers who are scientists, the distinction here is meant to reflect those whose primary direct focus of interest is the Earth System from those whose primary focus is on models themselves. Each of these groups is heavily populated as well.

‘overly humble’ just as other physicists still tend to take care when discussing ‘the discovery’ of a Higgs-like Boson. As we argue below, climate scientists can indicate the shortcomings of the details of their modelling results, while making clear that solid basic science implies that significant risks exist. If climate scientists are seen as ‘hiding’ uncertainties, however, the public reliability (reliability₃) of their findings may decrease, and with it the reliability₃ of solid basic physical insight.

Assessment of climate-model simulation for decision-making aims for consensus on the risks faced along with a rough quantitative estimate of the likelihood of their impacts. Providing such assessment is in practice the fundamental aim of the IPCC¹¹ and does not imply agreement on the action or level of investment to make.

3. Reliability₂ and climate-model based probability distributions

‘A good Bayesian does better than a non-Bayesian,
but a bad Bayesian gets clobbered’
Herman Rubin (1970)
quoted in I.J. Good ([1983] 2009)

Each and every probability distribution is conditioned on something. This claim led Sivia (2000) to argue that one should always write (or at least think) of every probability P as the probability of x conditioned on I , $P(x|I)$, where I is all of ‘the relevant background information at hand’. In a forecast situation x is the value of interest, say the temperature of the hottest day in Oxford in the summer of 2080, for instance. Sivia, with many other Bayesians, argues that absolute probabilities do not exist. Following Good ([1983] 2009) we would argue that it is often useful to speak as if absolute (physical) probabilities did exist, regardless of whether or not they do exist. Without omniscience, such probabilities may well not exist; with omniscience they may or they may not; as physicists we would note, following Mach (1856), that we will never be able to tell.

The challenges that face scientists, climate modellers, and policy-makers are not nearly so deep. Questions surrounding what is intended by I are nontrivial. UKCP09, for instance, appears to reinvent the expression ‘evidence considered’¹² leaving unanswered the question of whether there was enlightening ‘evidence suppressed’. It also provides no clear answer to the critical judgement call as to whether or not the current state of the scientific modelling is adequate for the purpose of forecasting x in the far future, full stop. We return to what was intended by the writers of UKCP09 below, the general point here is that if we are aware from the science that an RDU is not included within the evidence

¹¹ Principle 2 governing the work of the IPCC reads: ‘The role of the IPCC is to assess on a comprehensive, objective, open and transparent basis the scientific, technical and socio-economic information relevant to understanding the scientific basis of risk of human-induced climate change, its potential impacts and options for adaptation and mitigation. IPCC reports should be neutral with respect to policy, although they may need to deal objectively with scientific, technical and socio-economic factors relevant to the application of particular policies.’ See <http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles.pdf>.

¹² See UKCP09 web pages and presentations; for instance <http://ukclimateprojections.defra.gov.uk/22612> (in web page on UKCP09 customisable maps in detail) or http://www.ukcip.org.uk/wordpress/wp-content/UKCP09/UP_0609_presentation.pdf (in slide 23 in presentation to User Panel on June 3rd, 2009).

considered when constructing probability forecasts, then decision-makers have no rational foundation for using those probabilities as such. How can we motivate contractors, in academia and in industry, to lay bare the limitations of their analysis without the fear that (by respecting the limitations of today's science and simulations) they will be accused of failing to deliver on the contract?

The UKCP presentation introducing UKCP09 includes the statements 'The problem is we do not know which model version to believe.' and 'The only way we have of assessing the quality of a climate model is to see how well they simulate the observed climate.'¹³ The first statement might be taken to indicate that there is one model version which we might believe; no measure of how well these models simulate the current climate is given, but it is clear that some do better than others.

The decision-relevant climate simulations of any epoch must run significantly faster than real-time on the computers of that epoch. It is widely suggested within the UKCP09 guidance materials that relevant climate simulations of 2009 are available at postcode level resolution over the remainder of this century. Some climate modellers suggest that this is indeed the case although no clear statement supporting the claims of the UKCP09 methodology appeared in the peer reviewed literature before this year (2013). For an alternative view see Held (2005); for the reasons why we should expect difficulty in forming decision-relevant probability forecasts for systems best simulated by nonlinear models see Smith (2002). The importance of stressing the limitations on our ability to quantify uncertainty is stressed by Murphy et al. (2007) and Stainforth et al. (2007b).

Consider a parallel case. Many of us could code up Newton's Laws of gravity and simulate the motion of the planets of our solar system. Suppose each of us did so. The models would use different integration schemes, computer chips, integration time steps, perhaps some of us would include Pluto and others would not. Some of us would make coding errors; others might not have access to an accurate snapshot of the positions of each planet, some of us might have accidentally tweaked our models in-sample so that we account for noise in the observations as well as signal... Nevertheless, assessing the quality of each run based on the observed motions of the past would allow one rank-ordering of the relative quality of each run. And looking at the diversity of forecasts from the 'better' models would, we believe, provide a reasonable distribution for the likely location of the Earth for quite a long time into the future; for the planet Earth, but not for the planet Mercury.

Mercury does not 'obey' Newton's Laws, at least not without adding non-existent planet(s) between Mercury and the Sun.¹⁴ General relativity is required to model Mercury's orbit accurately. The fact that Mercury's orbit differed from expectations was known long before general relativity was available to allow realistic simulation. There are parallels here with climate, which include not knowing which calculation to make, and not being capable of accounting for effects we know are of importance in the calculation.

¹³ See <http://ukclimateprojections.defra.gov.uk/22530> (in web page on UKCP09 reports and guidance).

¹⁴ In the same way LeVerrier and Adams hypothesised a planet beyond Uranus to account for its motion, leading to the discovery of Neptune. In fact, LeVerrier did just this, leading to the discovery of the planet Vulcan in 1859.

We could be unaware of the theory of general relativity; alternatively one might find that many fewer of us could code up the relativistic simulation than could code the Newtonian simulation. If all we have access to is our ensemble of Newtonian simulations, then the diversity of our simulations do not reflect our uncertainty in the future location of Mercury.¹⁵ We can know that this is the case without being able to ‘fix’ it: the RDU in this case was structural model error demonstrated by the lack of fidelity in Newton’s Laws applied to Mercury’s orbit (we might, of course, think it to be some other unknown). Our model output for Mercury is not decision relevant regardless of how principled the statistical post-processing methods applied to the model ensemble may be. When the model class in hand is inadequate, and believed to be inadequate, the meaning of $P(x|I)$ is cast into doubt, as I is known to be incomplete, perhaps known to be false.

Let x be the location of Vulcan, which does not exist; let I be Newton’s Laws and let $\{\text{data}\}$ be all observations of the planets up to the year 2000. Would the Bayesian want to say that the integral of $P(x|\{\text{data}\}, I)$ over all space is equal to one? to zero? This integral is the oft suppressed ‘normalisation constant’ in the denominator of Bayes’ Theorem. Its importance is noted by Sivia (2000) in his Equations 1.10 and 2.2. If I is effectively inconsistent with the data, as observed when our model class is inadequate due to structural model error, and we either do not know how or cannot afford a better model class, the decision relevance of any model based probability statements is thrown into doubt, and with them the machinery of the Bayesians. It is not that Bayes’ Theorem fails in any sense, it is that arguably the probability calculus (of which it is a fundamental part) does not apply.

And so it is with climate. We know of phenomena that are critical to climate impact in the UK which are not simulated realistically in the models that went into the UKCP09 probability distributions. Not merely ephemeral phenomena like clouds, but rather more solid phenomena like mountain ranges: the Andes for example are not simulated realistically. Thus to the extent that these phenomena would affect the derived probability distributions, these distributions should be considered unreliable (*reliability*₂), which is a qualitative judgment (see Section 2). For instance, failure to realistically simulate ‘blocking’ was noted explicitly by the official Hoskins Review sponsored by the funders of UKCP09.¹⁶ Due to this acknowledged shortcoming, there are many circumstances in which interpreting UKCP09 probabilities as decision-relevant probabilities would be a mistake with potentially maladaptive consequences. But how is information on this propagated along the chain from an international review complete with external reviewers’ commentary to the decision-maker interested in how many consecutive clear days, or rainy days, or days above/below a given temperature are likely to be encountered in a typical year of the 2080s? Not well. We return to this below after noting that definitions might better be formed based on the underlying science than on what our current models might simulate.

¹⁵ Although applying an empirically determined statistical adjustment might lead to a model better than (more adequate than) Newton’s Laws on time scales which are much less than those on which Newton’s laws (on their own) provide adequate probability forecasts for other planets.

¹⁶See <http://ukclimateprojections.defra.gov.uk/23173> (UKCP09 science review).

4. A close look at UKCP09: Definitions and scope

Climate was sometimes wrongly defined in the past as just “average weather”.
H.H. Lamb (1982)

The definition of climate matters. Smith and Stainforth (in preparation) argue that the temptation to take the limitations of today’s models (or today’s RDU) into account when defining terms should be resisted. Current definitions provided by the IPCC Glossary of 2007¹⁷ and the American Meteorological Society Glossary of 2002 focus on averages, noting both variability and changes *in these averages*. Many older definitions, including the American Meteorological Society Glossary of 1959, discuss the ‘synthesis of weather’ however observed, embracing the entire distribution of higher dimensional weather states, and their time evolution, however expressed.

Defining climate as the distribution of such weather states allows a knowledge of climate to imply whatever averages may prove of interest to a scientist or a decision-maker. Defining climate in terms of average values and variances may make it more straightforward to evaluate climate models, while in reality removing the policy relevance of ‘climate’ by definition. Average values reflect neither the phenomena of interest to the physicist nor the impacts on individuals which determine policy (Smith and Stern 2011).

The Earth’s climate system appears sufficiently nonlinear that if state-of-the-art models are unable to realistically simulate the weather of a given epoch, then it is difficult to argue persuasively that they will get even the climatological averages of the following epoch correct. From either a scientific standpoint or a decision-making standpoint, the traditional definition of climate as a distribution is called for. Such a definition may imply that we cannot simulate even today’s climate very well. In that case determining (1) why this is the case, identifying whether RDU is due to computational constraints, to incomplete theory, or to a lack of observations; (2) estimating the magnitude of that uncertainty as a function of spatial scales and time; and (3) providing an estimate of when we can expect to address that dominant uncertainty, are each of immediate value to the policy process. Knowing the time required before the current RDU is likely to be reduced to the extent that another head of RDU will take its place is of great value in weighing up the advantages of delaying action against its costs.

In discussion of simulations in the context of reality, it is of value to clearly separate the properties of ‘model quantities’ from the ‘quantities’ themselves. Manabe and Wetherald (1975), along with many other climate modelling papers of that time, repeatedly distinguish ‘the state of the model atmosphere’ from that of the atmosphere. Similarly they often say the ‘model troposphere’, the ‘model stratosphere’ and so on. They note ‘a mistake in the programming’ and quantify it, suggesting its impact. And they conclude noting boldly that ‘because of the various simplifications of the model described above, it is not advisable to take too seriously the quantitative aspect of the results obtained in this study.’ Manabe, Wetherald and other climate modellers took pains to clarify aspects of reliability₂ in their results, the focus being on understanding phenomena rather than quantifying events in the world.

¹⁷ See http://www.ipcc.ch/publications_and_data/ar4/wg1/en/annexes.html.

Modern studies are rarely so blunt nor so clear in terms of distinguishing model-variables from their physical counterparts. They often also lack clarity in terms of stating spatial and temporal scales (if any) on which the quantitative aspects of model simulations are not to be taken seriously, or on how those change as the simulations are run farther into the future. The disclaimers tend to state the obvious: that today's climate science cannot provide a complete picture of the future. It is of course doubtful that climate science ever will provide a complete picture inasmuch as climate prediction always remains extrapolation. Catch-all disclosures would be more valuable if accompanied by clear statements of specific known limitations of current insights regarding particular applications.

The UKCP09 Briefing Report (Jenkins et al. 2009) does contain some clear statements. For example in a bullet point list on page 6 on the power of Bayesian statistical procedures:

Errors in global climate model projections cannot be compensated by statistical procedures no matter how complex, and will be reflected in uncertainties at all scales.¹⁸

This quote is taken almost verbatim from the Hoskins Review, although it might be misread to suggest that model errors will be reflected in the UKCP09 probability distributions ('uncertainties at all scales') which the Hoskins Review held that they will not. An online version of a similar bullet point list contains the statement:

Models will never be able to exactly reproduce the real climate system; nevertheless there is enough similarity between current climate models and the real world to give us confidence that they provide plausible projections of future changes in climate (Annex 3).¹⁹

It is as true as it is uninformative to claim that models will never be able to 'exactly reproduce' their target systems. Do the authors of the report believe that the probabilities provided can be interpreted as the probabilities of events in the real world? The hyperlink provided in the report to 'confidence' above does not address this question, and in Annex 3 we find:

We have no positive evidence that such factors would, if included, provide sources of uncertainty comparable with those included in UKCP09 (at least for projection time scales of a century or less), but this remains an issue for future research.²⁰

This is followed by a discussion of assumptions 'imposed by limitations in computational resource' and that 'non-linear interactions between uncertainties in different components

¹⁸See <http://ukclimateprojections.defra.gov.uk/22549> (Briefing report downloads).

¹⁹See <http://ukclimateprojections.defra.gov.uk/22769> (Online climate change projections report Purpose & design of UKCP09).

²⁰See <http://ukclimateprojections.defra.gov.uk/22783> (Online climate change projections report 3.3 Interpretation).

of the Earth system are important at the global scale, but not at the regional scale, because our finite computing resources were not able...’ and then:

we believe that the UKCP09 methodology represents the most systematic and comprehensive attempt yet to provide climate projections which combine the effects of key sources of uncertainty, are constrained by a set of observational metrics representative of widely-accepted tests of climate model performance, and provide a state-of-the-art basis for the assessment of risk, within limits of feasibility imposed by current modelling capability and computing facilities.²¹

We agree with this quote from Annex 3 in terms of what the UKCP09 probabilities are, but these clear caveats do not suggest to us that the UKCP09 probabilities can be taken to reflect what we expect in the 2080s, or as ‘plausible projections of future changes in climate’ at 25 km length scales across the UK.

In the original print version of the bullet points on page 6, the last point reads:

The method developed by UKCP09 to convert climate model simulations into probabilistic estimate of future change necessitates a number of expert choices and assumptions, with the result that the probabilities we specify are themselves uncertain. We do know that our probabilistic estimates are robust to reasonable variations within these assumptions.²²

This last sentence is perhaps key. While it is useful to know that the UKCP09 probabilistic estimates are robust to reasonable variations within these assumptions, it seems remiss not to acknowledge that those same estimates are known not to be robust to variations in (alternative) assumptions which are consistent with today’s science, if that is indeed the case.

Annex 3 includes long discussion of small uncertainties and things that can be sampled and controlled for. It concludes:

In summary, the UKCP09 projections should be seen as a comprehensive summary of possible climate futures consistent with understanding, models and resources available at present, but users should be aware that the projections could change in future, as the basis for climate prediction evolves over time.²³

What appears to be missing in the entire (evolving) document is a statement as to whether or not *the projections are expected to change*. We would argue that they are not mature probabilities; we expect this probabilities to change without additional empirical evidence just as the Newtonian forecasts of Mercury were expected to change. Not due to new science, but due to steps taken to address known shortcomings in the UKCP09 projections. Were that the case, the phrase ‘evidence considered’ takes on a Machiavellian hue.

²¹See footnote 20.

²²See footnote 18.

²³See footnote 20.

To our knowledge, the UK Met Office Hadley Centre (MOHC) has made no claim that the UKCP09 probabilities are robust in terms of providing well-calibrated forecasts of future weather of the 2080s; UKMO support is notably lacking for the output of the weather generator which is central to many headline claims.²⁴ At the 2009 UKCP science meeting in the Royal Society, members of the UKMO repeatedly stated that they would not know how to improve them today, when asked directly if the probabilities were adequate for decision-making. Climate researchers are in a difficult position, particularly given the political climate into which they have been drawn. How might one incentivise scientists and consultants to open the can of worms regarding where a given climate report is known not to supply well-calibrated decision-relevant probabilities, rather than obscure the fact of the matter with discussions of ‘best available’ or our inability to do better in the near future? How do we communicate our uncertainty when probabilities conditioned on ‘evidence considered’ are expected to give a rosier picture of uncertainty than those conditioned on all the available evidence? Does the failure to do so not risk a backlash from numerate users when future reports make this obfuscation clear?

A scientist reading through the report will find what appears to be clear evidence that the UKCP09 probabilities should not be taken at face value as probability statements about the real world; red flags to a scientist might not be so obvious to non-scientists. A few are collected below; they suggest the user look at the detailed regional model results if ‘variability is important to the individual user’ (not that the output is adequate for purpose); they defend the Hadley Centre model as competitive with other models (not adequate for purpose); they claim to represent ‘some aspects’ of blocking ‘with reasonable fidelity’ (not claiming to be adequate either for the purpose of extracting weather impacts, nor for the purpose of guiding reasonable simulation of the evolution of the UK’s geography as it changes over the six decades still to pass before 2080):

Careful evaluation of such diagnostics from the RCM simulations and the weather generators is recommended in cases where such variability is important to the individual user.²⁵

It should be recalled from Annex 3, Figure A3.6, that current positions and strengths of the modelled storm track do not always agree well with observations, and this should be taken into account when assessing the credibility of their future projections. The HadCM3 ensemble shows a better agreement in present day location than most other climate models, and a reasonable agreement in strength.²⁶

The mechanisms for atmospheric blocking are only partially understood, but it is clear that there are complex motions, involving meso-scale atmospheric turbulence, and interactions that climate-resolution models may not be able to represent fully. The prediction of the intensity and duration of blocking events is one of the most difficult weather forecasting situations. The HadCM3 model does represent, with reasonable fidelity, some aspects of present-day atmospheric blocking in the N.

²⁴See <http://ukclimateprojections.defra.gov.uk/23261> (Weather Generator).

²⁵ See <http://ukclimateprojections.defra.gov.uk/23080> (Online climate change projections report Annex 3.4.2).

²⁶See <http://ukclimateprojections.defra.gov.uk/22980>(Online climate change projections report Annex 6.2).

Atlantic region (see Figure A3.7) with the performance in summer better than that in winter. At other longitudes the model shows less fidelity, in particular in the Pacific sector. (An additional complication is that it is not clear that simply doubling the resolution of a climate model automatically produces a better simulation of blocking — in the case of one Met Office Hadley Centre model, this results in a degradation).²⁷

The inconsistency of the three diagnostics makes it difficult to make a clear statement about the ability of the perturbed physics ensemble to simulate anticyclones, but in general the HadCM3 ensemble is competitive with other climate models.²⁸

The insights gained in the construction of UKCP09 hold significant value for governments and decision-makers beyond its targeted areas in the United Kingdom. It is extremely valuable to learn that some things cannot be done now, nor in the near future. Knowing that some aspects of guidance will not be available in the near future is a great aid to decision-making. Making it clear that some of the known limits of today's models 'might' severely limit the relevance of the precise probabilities provided by UKCP09, would certainly be of value. But sometimes the report fails to give a clear 'Yes' even to such questions:

The role of atmospheric blocking under climate change is currently a major topic of research. Might current model errors severely limit the reliability of climate change projections (e.g. Palmer *et al.* 2008; Scaife *et al.* 2008)? Might large changes in blocking, that current models cannot simulate, cause large changes in the frequency of occurrence of summer heat waves for example?²⁹

Given this positive framing of the outputs of the project, how are other nations to learn of the severe limitations UKCP09 imposes, subtly, on the application of its outputs in practice? How would other nations debating similar studies evaluate the value of such a study, where it might have decision relevance useful within their borders, and where decisions must be made under deep uncertainty, that is, in ambiguity, without robust statements of probability? Failing to clarify the limitations of the exercise are likely to lead to misapplications within the UK, and potentially misguided attempts to obtain in other countries what the quotes above show has *not* been obtained for the UK.

How can those civil servants and politicians promoting UKCP09 from inside, as well as professional scientists, be incentivised to communicate their clear insights into its limitations? How can they resolve the conflict between attracting and engaging their target audience of users while laying bare the limited utility of the product currently available? The UK's goal to continue to make world leading contributions towards the understanding of climate risks will be hampered if future government projects like the Climate Change Risk Assessment adopt the naïve interpretation of the UKCP09 outputs suggested by the report's headlines and fail to take on board the deep limitations and

²⁷See footnote 25.

²⁸See footnote 25.

²⁹See footnote 25.

implication of the dominant uncertainties identified in the body of the report. How might future studies promote critical information on their own shortcomings in their headline results?

4. Conclusions: Insights and improvement

You can't have a light without a dark to stick it in.
Arlo Guthie

Like weather forecasting and medical diagnosis, climate modelling is riddled with error and inadequacy. That is not pessimistic, even if it is often suggested to be so; it is an opportunity for the sciences, not a reason to abandon science (Beven 2006). Medical diagnosis has the advantage that there are many many people from which we can more quickly learn the behaviours of all but the most rare diseases. This aids our ability to advance the science and make better decisions regarding patient care. While long-standing philosophical arguments as to what (and how many) probabilities 'are' remain (Good [1983] 2009), there is firm empirical evidence that they are useful; many variations³⁰ on the Bayesian paradigm have provided a valuable organising framework for managing probabilities when we think we have them, and have improved the approach taken to a problem even when we do not. In the case of weather, we have only one planet to observe. Still we observe effectively independent weather events happening all over the world every day, and we have the opportunity to learn by observing systematic failures in our weather models day after day. Week after week, after week. Operational centres actively save observations corresponding to past forecast busts to re-examine with each future generation of weather models.

Climate modelling is different (Smith 2002), in that the lifetime of a model is less than the typical forecast time and that the forecast time is often longer than the span of a scientist's career. This makes it much more difficult to learn from our mistakes. Of course, the fact that climate science comes under attack from a politically-motivated anti-science lobby complicates things, a point we will return to below. One scientific challenge to climate science is that it is not at all obvious we will ever be able to model climate at lead times of 80 years on length scales of neighbourhoods in any manner that is quantitatively informative to adaptation decisions. If true, merely knowing this is itself of great value to policy-makers. Knowing the length-scales on which we can expect a 'Big Surprise' deviating from our model-based probabilities in 80 years, or 40 years, or 20 years, requires a new approach to quantifying 'deep uncertainty'. Even only twenty years out, is the probability of a big surprise less than 1 in 200? What is this probability for 2050, given the known limits of our current models? How much more will we know in 2017? Such questions are of immediate value in terms of rational mitigation policy.

While the United Kingdom has pushed the outside of the envelope with the UKCP09 projections, the so-called 'Bayesian' framework adopted by UKCP09 was, even after its launch, unable to clarify how the likelihood that its model-based probabilities will prove mal-adaptive as it looks farther and farther into the future. The probability

³⁰See footnote 6.

distributions for the 2080s are presented as if they were as robust as those for the 2020s. Or our knowledge from observations of the 2000s which, of course, come with probability distributions due to observational uncertainties. How do we move to a broader and deeper understanding of the implications of incomplete knowledge when we are denied the luxury of millions of experiments an hour? Stirling (2010) stresses the importance of keeping it complex: of not simplifying advice under the pressure to make it useful or accessible. This advice may lead us beyond probabilities, and the ability to cope with instances when the denominator in Bayes' Theorem is zero (if indeed it is well-defined at all). But how do we do this?

One could never usefully refute the claims of long-range climate predictions empirically. We can, however, explore the strength of evidence (*all* the available evidence) for their likely relevance. The quotes given in the previous section show that the seeds of doubt are sown deep within the scientific text of the report, yet they are all but suppressed if not obfuscated in the headline sections. Every user³¹ of the UKCP09 products can hardly be expected to know all the meteorological phenomena relevant to an application of interest, nor is it reasonable to expect individuals to examine 'diagnostics from the RCM simulations and the weather generators'. Known shortcomings of each generation of models should allow more informative information on the expected robustness and relevance of particular UKCP09 probability distributions as a function of lead time. Information on the relevant dominant uncertainty is more useful when it is identified clearly *as* the Relevant Dominant Uncertainty; it is less useful when buried amongst information of other uncertainties that are well quantified, have small impacts, or are an inescapable fact of all scientific simulation. Given the understandable tendency of modellers to defend their models as the best of the current generation of models, and the difficulties of a team constructing a report cataloguing its shortcomings, perhaps an accompanying second scientific report, a 'minority opinion', is called for: an independent study clarifying the limits of robustness, could be included in all such highly policy-relevant scientific reports. Ideally scientific risk assessments will translate both the implications and the limitations of the science for decision-makers in each sector. Extrapolation problems like climate change can benefit from new insights into: how to better apply the current science, how to advertise its weaknesses and more clearly establish its limitations; all for the immediate improvement of decision support and the improvement of future studies. This might also aid the challenge of training not only the next generation of expert modellers, but also the next generation scientists who can look at the physical system as a whole and successfully use the science to identify the likely candidates for future RDU. Finally, note that government sponsorship of these minority reports would not only aid immediate decision making, but could also have the second order effect of tightening the claims and clarity of future primary reports³².

³¹ Arguably, every user of the probabilities for 2080 is 'exposed to' the impacts of the global model generating the 'wrong weather' worldwide, and accumulating impacts of the failure of local downscaling to simulate 'blocking' realistically, for instance, over the intervening six decades.

³² One might argue that the contents of a minority opinion should be part and parcel of the study itself. While we would agree with this goal, the facts on the ground indicate that, for whatever reason, clear information of the potential irrelevance of the finding of the report is often not highlighted in practice. A major aim in suggesting a minority opinion is to ultimately raise the profile of this information in the primary study.

An alternative approach to using high-resolution climate predictions that has been developed in the Netherlands holds significant promise and seems likely to be more informative than the 2009 UK approach. We here refer to the approach taken by the Dutch government and its Royal Netherlands Meteorological Institute (KNMI) towards climate scenarios for the Netherlands (KNMI 2012). The Dutch approach explicitly departs from the view that a full probabilistic approach to climate predictions at the regional scale is currently feasible. Recognising the inherent predictability limits of climate and the different epistemic values held by different groups of experts, the Dutch climate experts argue for a different approach. For instance with respect to changes in extreme weather events, narratives describing existing knowledge of the physics of extreme weather events accompanied by simulations of extreme weather events in well-calibrated numerical weather prediction models in present-day climate and a potential future climate setting are provided to give a realistic and physically consistent picture of both the types of events that need preparatory actions and the impacts of the adaptation decisions taken (Hazeleger et al. in preparation).

Climate policy on mitigation and decision-making on adaptation provide a rich field of evidence on the use and abuse of science and scientific language. We have a deep ignorance of what detailed weather the future will hold, even as we have a strong scientific basis for the belief that anthropogenic gases will warm the surface of the planet significantly. It seems rational to hold the probability that this is the case far in excess of the '1 in 200' threshold which the financial sector is required to consider by law (regulation). Yet there is also an anti-science lobby which uses very scientific sounding words and graphs to bash well-meaning science and state-of-the-art modelling. If the response to this onslaught is to 'circle the wagons' and lower the profile of discussion of scientific error in the current science, one places the very foundation of science-based policy at risk.

Failing to highlight the shortcomings of the current science will not only lead to poor decision-making, but is likely to generate a new generation of insightful academic sceptics, rightly sceptical of oversell, of any over-interpretation of statistical evidence, and of any unjustified faith in the relevance of model-based probabilities. Statisticians and physical scientists outside climate science (even those who specialise in processes central to weather and thus climate modelling) might become scientifically sceptical, sometimes wrongly, of the basic climate science in the face of unchecked oversell of model simulations. This mistrust will lead to a low assessment by these actors of the reliability₃ (public reliability) of findings from such simulations, even where the reliability₁ and reliability₂ are relatively high (e.g. with respect to the attribution of climate change to significant increases in atmospheric CO₂).

It is widely argued that we need decision-makers to think about climate more probabilistically, as we will never be able to tell them exactly what will happen in the far future. That argument does not justify the provision of probability distributions thought not to be robust. Indeed, numerate users already use probability distributions as such. As additional numerate users appear, what would be the impact of having to explain in 2017 that the probabilities of 2009 were still being presented as robust and mature in 2011 even when they were known not to be? How, in that case, would one encourage use of the 2017 'probabilities' (or those of 2027) to that same numerate audience? Learning to better deal with deep uncertainty (ambiguity) and known model inadequacy can advance

significantly and foster the more effective use of model-based probabilities in the real world.

Acknowledgements

We would like to thank Mike Hulme for his suggestion in June 2010 to add the notion of reliability₃ to the notions of reliability₁ and reliability₂ that were first introduced in Petersen ([2006] 2012), and Brian Hoskins, Wendy Parker, Dave Stainforth and Nick Stern for continuing conversations on the matters of this chapter, and beyond.

References

- Beven, K. (2006), 'On undermining the science?', *Hydrological Processes* 20: 3141–3146.
- Funtowicz, S.O. and Ravetz, J.R. (1993), 'Science for the post-normal age', *Futures* 25: 739–755.
- Good, I.J. ([1983] 2009), *Good Thinking: The Foundations of Probability and Its Applications*, Mineola, NY: Dover Publications.
- Hazeleger, W., van den Hurk, B.J.J.M., Min, E., van Oldenborgh, G.J., Wang, X., Petersen, A.C., Stainforth, D.A., Vasileiadou, E. and Smith, L.A. (in preparation), 'Tales of future weather'.
- Held, I. (2005), 'The gap between simulation and understanding in climate modeling', *Bulletin of the American Meteorological Society* 86: 1609–1614.
- Hon, G. (1989), 'Towards a typology of experimental errors: An epistemological view', *Studies in History and Philosophy of Science* 20: 469–504.
- Hulme, M. and Mahony, M. (2010), 'Climate change: What do we know about the IPCC?', *Progress in Physical Geography* 34: 705–718.
- IAC(2010), *Climate Change Assessments: Review of the Processes and Procedures of the IPCC*, Amsterdam, The Netherlands: InterAcademy Council. Available at: <http://reviewipcc.interacademycouncil.net>.
- IPCC (2005), *Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties*, Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC). Available at: <http://www.ipcc.ch/pdf/assessment-report/ar4/wg1/ar4-uncertaintyguidancenote.pdf>.
- Jenkins, G.J., Murphy, J.M., Sexton, D.M.H., Lowe, J.A., Jones, P. and Kilsby, C.G. (2009), *UK Climate Projections: Briefing Report*, Exeter, UK: Met Office Hadley Centre. Available at: <http://ukclimateprojections.defra.gov.uk/media.jsp?mediaid=87868&filetype=pdf> (version 2, December 2010).
- Kandlikar, M., Risbey, J.S. and Dessai, S. (2005), 'Representing and communicating deep uncertainty in climate change assessments', *Comptes Rendus Geosciences* 337: 443–455.
- KNMI (2012), *Advisory Board Report: Towards the KNMI '13 Scenarios – Climate Change in the Netherlands*, De Bilt, The Netherlands: Royal Netherlands Meteorological Institute (KNMI), KNMI Publication 230. Available at: http://www.knmi.nl/climatescenarios/documents/AdvisoryBoard_report_towards_KNMI_13.pdf.
- Lamb, H.H. (1982), *Climate History and the Modern World*, Oxford: Oxford University Press.
- Mach, E. (1856) Error and Uncertainty
- Manabe S. and Wetherald, R.T. (1975), 'The effects of doubling the CO₂ concentration on the climate of a General Circulation Model', *Journal of the Atmospheric Sciences* 32: 3–15.

- Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., Plattner, G.-K., Yohe, G.W. and Zwiers, F.W. (2010), *Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties*, Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC). Available at: <http://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf>.
- Millner, A., R. Calel, D. Stainforth, and G. MacKerron, Do probabilistic expert elicitation capture scientists' uncertainty about climate change? *Climatic Change*, 2013. 116(2): p. 427-436.
- Moss, R. and Schneider, S.(2000), 'Uncertainties in the IPCC TAR: Recommendations to Lead Authors for more consistent assessment and reporting', in R. Pachauri, T. Taniguchi and K. Tanaka (eds.), *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*, Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC). Available at: http://climateknowledge.org/figures/Rood_Climate_Change_AOSS480_Documents/Moss_Schneider_Consistent_Reporting_Uncertainty_IPCC_2000.pdf.
- Murphy, J.M., Booth, B.B.B., Collins, M., Harris, G.R., Sexton, D.M.H. and Webb, M.J. (2007), 'A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles', *Philosophical Transactions of the Royal Society A* 365: 1993–2028.
- Palmer, T. N., Doblas-Reyes, F. J., Weisheimer, A. & Rodwell, M. J. (2008) Toward seamless prediction: calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society*, 459–470.
- Parker, W.S. (2009), 'Confirmation and adequacy-for-purpose in climate modeling', *Proceedings of the Aristotelian Society* 83 (Suppl.):233–249.
- Petersen, A.C.(2000), 'Philosophy of climate science', *Bulletin of the American Meteorological Society* 81: 265–271.
- Petersen, A.C. ([2006] 2012), *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*, Second Edition, Boca Raton, FL: CRC Press.
- Petersen, A.C., Cath, A., Hage, M., Kunseler, E. and van der Sluijs, J.P. (2011), 'Post-normal science in practice at the Netherlands Environmental Assessment Agency', *Science, Technology, & Human Values* 36: 362–388.
- Randall, D.A. and Wielicki, B.A.(1997), 'Measurements, models, and hypotheses in the atmospheric sciences', *Bulletin of the American Meteorological Society* 78: 399–406.
- Risbey, J.S. and Kandlikar, M. (2007), 'Expressions of likelihood and confidence in the IPCC uncertainty assessment process', *Climatic Change* 85: 19–31.
- Scaife, A. A., Buontempo, C., Ringer, M., Sanderson, M., Gordon, C. K. & Mitchell, J. (2008) Comment on Toward seamless prediction: calibration of climate change projections using seasonal forecasts.
- Sivia, D.S. (2000), *Data Analysis: A Bayesian Tutorial*, Oxford: Clarendon Press.
- Smith, L.A. (2002), 'What might we learn from climate forecasts?', *Proceedings of the National Academy of Sciences of the United States of America* 99 (Suppl. 1): 2487–2492.
- Smith, L.A. (2006) Predictability past predictability present. Chapter 9 of Predictability of weather and climate (eds T. Palmer & R. Hagedorn). Cambridge, UK: Cambridge University Press
- Smith, L.A. (2013) 'Insight and Action: Probabilistic Prediction in Practice' *International J of Forecasting* (to appear).
- Smith L.A. and Stern, N. (2011), 'Scientific uncertainty and policy making', *Philosophical Transactions of the Royal Society A* 369: 1–24.
- Smith L.A. and Stainforth, D. (in preparation), 'Putting the weather back into climate: On the definition of climate'.

- Stainforth, D.A., T.E. Downing, R. Washington, A. Lopez, and M. New (2007a) , Issues in the interpretation of climate model ensembles to inform decisions. Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 2007. 365(1857): p. 2163-2177.
- Stainforth, D.A., Allen, M.R., Tredger, E.R. and Smith, L.A. (2007b) [Confidence, uncertainty and decision-support relevance in climate predictions](#), *Phil. Trans. R. Soc. A*, **365**, 2145-2161.
- Stirling, A, (2010) ‘Keep it complex’, *Nature* 468: 1029–1031.
- Strengers, B.J., Meyer, L.A., Petersen, A.C., Hajer, M.A., van Vuuren, D.P. and Janssen, P.H.M. (2013), *Opening up Scientific Assessments for Policy: The Importance of Transparency in Expert Judgements*, PBL Working Paper 14. The Hague: PBL Netherlands Environmental Assessment Agency.
- Swart, R., Bernstein, L., Ha-Duong, M. and Petersen, A.C. (2009), ‘Agreeing to disagree: Uncertainty management in assessing climate change, impacts and responses by the IPCC’, *Climatic Change* 92: 1–29
- van der Sluijs, J.P.(1997), *Anchoring amid Uncertainty: On the Management of Uncertainties in Risk Assessment of Anthropogenic Climate Change*, PhD dissertation, Utrecht: Utrecht University.
- Weinberg, A.M. (1972), ‘Science and trans-science’, *Minerva* 10: 209–222.