Roman Frigg, Erica Thompson and Charlotte Werndl

Philosophy of climate science part II: modelling climate change

Philosophy Compass 10/12 (2015): 965–977, 10.1111/phc3.12297

Philosophy of Climate Science Part II: Modelling Climate Change[1]

Roman Frigg[1], Erica Thompson[1] and Charlotte Werndl[1,2]*
[1]*London School of Economics*
[2]*University of Salzburg*

Abstract

This is the second of three parts of an introduction to the philosophy of climate science. In this second part about modelling climate change, the topics of climate modelling, confirmation of climate models, the limits of climate projections, uncertainty and finally model ensembles will be discussed.

1. *Introduction*

After having discussed the observation of climate change in Part I, we now turn to climate modelling, which is the focus of Part II. We start with a primer on climate modelling (Section 2) and then ask how climate models are confirmed (Section 3). Many models are used to make projections about the future climate, which raises the question of what the limits of such projections are (Section 4), and of how uncertainty in models can be understood (Section 5). A popular method of dealing with uncertainty is the use of model ensembles. We describe this method and discuss the difficulties that attach to it (Section 6) and then end with a brief conclusion (Section 7).

2. *A Primer on Climate Models*

A climate model is a representation of certain aspects of the climate system. One of the simplest climate models is an energy-balance model, which treats the earth as a flat surface with one layer of atmosphere above it. It is based on the simple principle that in equilibrium, the incoming and outgoing radiation must be equal (see Dessler 2011, Chapters 3–6, for a discussion of such models). This model can be developed further by dividing the earth into zones, allowing energy transfer between zones, or describing a vertical profile of the atmospheric characteristics. Despite their simplicity, these models provide a good qualitative understanding of the greenhouse effect.

Modern climate science aims to construct models that integrate as much as possible of the known science (for an introduction to climate modelling, see McGuffie and Henderson-Sellers 2005). Typically, this is done by dividing the earth (both the atmosphere and ocean) into grid cells. At the time of writing, state-of-the-art global climate models had a horizontal grid scale of around 150 km. Climatic processes can then be conceptualised as flows of physical quantities such as heat or vapour from one cell to another. These flows are mathematically described by fluid dynamical equations. These equations form the 'dynamical core' of a global circulation model (GCM). The equations typically are intractable with analytical methods, and powerful computers are used to solve them. For this reason, they are often referred to as 'simulation models'. To solve equations numerically, time is discretised. Current state-of-the-art simulations use time steps of approximately 30 minutes, taking weeks or months in real time on supercomputers to simulate a century of climate evolution.

In order to compute a single hypothetical evolution of the climate system (a 'model run'), we also require *initial conditions* and *boundary conditions*. The former are a mathematical description of the state of the climate system (projected into the climate model's own domain) at the beginning of the period being simulated. The latter are values for any variables which affect the system but which are not directly output by the calculations. These include, for instance, the concentration of greenhouse gases in the atmosphere at a given time, the amount of aerosols and the rate of solar radiation received by the earth. Since these are drivers of climatic change, they are often referred to as *external forcings* or *external conditions*. Different models regard different physical quantities as boundary conditions. For instance, in some models, the atmospheric concentration of carbon dioxide is set to follow a predetermined pathway (and is hence treated as an external boundary condition), while other models directly simulate the carbon cycle (and hence regard it as a dynamical internal variable). Faster simulation time can be achieved by 'switching off' aspects of a complex model, for instance replacing an interactive ocean circulation by prescribed sea surface temperatures.

Processes that occur on a scale smaller than can be resolved by the grid may be included via *parameterisation*, whereby the net effect of the process is separately calculated as a function of the grid variables. For instance, cloud formation is a physical process which cannot be directly simulated because typical clouds are much smaller than the grid. So the net effect of clouds is usually parameterised (as a function of temperature, humidity, etc.) in each grid cell and fed back into the calculation. Sub-grid processes are one of the main sources of uncertainty in climate models.

There are currently around 20 complex climate models, which are under continuous development by national modelling centres like NASA, the UK Met Office and the Beijing Climate Center. In order to be able to compare outputs of these different models, the Coupled Model Intercomparison Project (CMIP) defines a suite of standard experiments to be run for each climate model. One standard experiment is to run each model using the historical forcings experienced during the 20th century, so that the output can be directly compared against real climate system data. Chapter 9 of the most recent report from Working Group I of the IPCC discusses model evaluation in some detail (IPCC 2013).

Models can be put to different uses. For instance, simulation results can be used either to develop physical insight about the system by means of comparison with reality, or to make projections of future climate (see Held 2005 for a discussion of this contrast; see Schmidt and Sherwood 2015 for a practitioner's view of climate modelling). 'Projection' is a technical term in the climate modelling literature and refers to a forecast that is conditional on a certain forcing scenario. The scenario is specified by the changes in total radiative forcing, which depend primarily upon the amount of greenhouse gas emissions and aerosols added to the atmosphere but also on natural processes, solar activity and land use change. Human aspects of these contributory factors depend on other drivers such as socioeconomic, political and technological developments. Note that the use of scenarios in climate research is under development, with a change from emissions scenarios to forcing scenarios between AR4 and AR5, and moves now in progress towards new Shared Socio-Economic Pathways, which have greater and more consistent detail of socioeconomic change (O'Neill et al. 2014).

The forcing scenario is specified by the amount of greenhouse gas emissions and aerosols added to the atmosphere, and these in turn depend on number of other factors such as changes in land use as well as future socioeconomic and technological developments. The essential task of climate policy is to compare these projections and decide what actions to take. Uncertainty about the climate impacts related to a particular forcing scenario derives from the different representations of climate within different models.

3. *Confirmation of Climate Models*

In Section 3 of Part I, we have seen that there are various ways data are obtained in climate science. *Confirmation* concerns the question whether, and to what degree, a certain model is supported by these data. To give an example, climate scientists often look at whether a model manages to simulate data of past temperature changes sufficiently well and, if so, claim that the model is confirmed (Harvey and Kaufmann 2002; IPCC 2013, Chapter 9; Knutti et al. 2003).

In the case of climate models, it is important to distinguish between variables we expect to have predictive skill and those we do not; confirmation concerns only the former. That is, model confirmation concerns certain variables of a model but does not extend to others (cf. Parker 2009). For example, one might claim that a certain climate model adequately predicts the global temperature increase by 2100 (relative to certain initial conditions and a certain emission scenario). Yet, one might also claim that the predictions of global mean precipitation by 2100 by the same model cannot be trusted.

It is widely held that many *aspects of climate models are confirmed* (Lloyd 2010). For instance, data confirm that models including natural and anthropogenic factors adequately retrodict the past mean temperature changes within a certain margin of error (cf. Section 5 in Part I). To reach such a judgement, a number of obstacles have to be overcome. We now review the difficulties involved in confirming climate models.

Suppose climate scientists find that a model retrodicts certain variables well and is thus confirmed (e.g. that the model's retrodictions of global mean temperature changes can be trusted because it simulates the past temperature changes well). Further, suppose that another model does not predict certain variables well and is thus disconfirmed (e.g. that the model's retrodictions of the global mean temperature changes cannot be trusted because it does not simulate past temperature changes well). Then one would like to know the reasons why the first model is successful and the second is not. Lenhard and Winsberg (2010) argue that this is hopeless: For complex climate models, a *strong version of confirmation holism* makes it impossible to tell where the failures and successes of climate models lie. In particular, they claim that it is impossible to assess the merits and problems of submodels, and parts of models more generally. They also argue that this strong form of confirmation holism is here to stay. Note that the issue here is not just the widely accepted claim of confirmation holism, i.e. that a single hypothesis cannot be tested in isolation because tests depend on auxiliary hypotheses. Lenhard and Winsberg defend the more radical claim that one will never be able to say where the successes and failures of climate models come from.

Lenhard and Winsberg present case studies (e.g. about the Coupled Model Intercomparison Project) which nicely illustrate some of the difficulties involved. Still, there remains a question whether complex models in climate science always suffer from this strong confirmation holism. Complex models have different modules, which can be run individually and as well as together. At least in certain cases, by comparing these individual and combined runs, it seems plausible that one can obtain an understanding of the effect of modules and then attribute successes and failures to certain submodels. Indeed, there are cases of this in the literature. For instance, Qu and Hall (2014) examine snow-albedo feedback of 25 climate simulations of the Coupled Model Intercomparison Project version 5. They find that while for some models the snow-albedo feedback is in line with the observations, for others it is not. A detailed analysis of different modules leads them to the conclusion that the wide variation in treatments of effects due to the presence of vegetation on snow-covered surfaces is a major factor responsible for the variations in model performance.

Data are often used in the construction of climate models. In particular, the values of many parameters of models are unknown because the physical processes are poorly understood and because parameters often stem from an attempt to simplify theory when it is available. Therefore, data are used to estimate these parameters (a process called *calibration*). For example, the net effect of cloud cover is estimated from observations rather than determined by a bottom-up theoretical calculation (e.g. Golaz, Horowitz and Levy 2013).

When data have been used for calibration, the question arises *whether the same data can be used to confirm the model*. To come back to our example, if data about temperature changes are used to calibrate the parameter of a climate model, can the same data be used again to confirm it? If data are used for confirmation which have not already been used before for calibration, they are *use-novel*. If data are used for both calibration and confirmation, this amounts to *double-counting*. This raises the question whether double-counting is acceptable. Many philosophers and climate scientists argue for a negative answer to this question (Lloyd 2010; Shackley et al. 1998; Worrall 2010). They think that double-counting is illegitimate and that data have to be use-novel to be confirmatory.

Against this view, Steele and Werndl (2013) argue that according to *Bayesian* and relative-likelihood confirmation use-novelty is beside the point and double-counting is legitimate. In a Bayesian analysis, data will always be used for calibration, and in many cases – in particular, when the model has a very good fit with the data and this fit is highly nontrivial – the same data will also confirm the model. For instance, in standard linear regression, all data are used to estimate the slope of the curve. Here, the same data can also confirm the linear model if the fit between the data and the estimated linear curve is good.

Another statistical theory that has interesting implications for use-novelty and double-counting is *model selection theory* (e.g. Michaelsen 1987). Model selection theory is relevant for climate science because the methods it provides are applicable whenever one wants to compare the predictive abilities of a number of competing models (as e.g. in standard linear regression contexts which often arise for detection and attribution studies). Steele and Werndl (2014) argue that the picture model selection theory presents us with about use-novelty is more nuanced than the commonly endorsed positions. More specifically, first, there are methods such as cross-validation where the data are required to be use-novel. For cross-validation, the data are split up into two groups: the first group is used for calibration and the second for confirmation. Second, there are the methods such as the Akaike Information Criterion for which the data need not be use-novel. Still, there is a penalty term in the expression for the degree of confirmation penalising the use of data that have already been used for calibration.

Finally, let us mention three other problems about calibration and confirmation. First, a common practice is to rule out those models that do not show a close enough fit with the data and use those that are left (weighted equally) for predictions. Here again, the question arises why these predictions should be trusted and whether some of the models that are left are not better than others. Second, climate scientists question to what extent past data are relevant for assessing how well models predict the future. The worry is that processes important on longer time scales, such as changes in the ice coverage, will take the system out-of-sample with respect to existing observations, such that empirically derived relationships and parameterisations could change significantly. Because of this, they argue that *it is really hard to tell how relevant past data are or that past data are not relevant at all for assessing the predictive accuracy of the models in the future*. Consequently, past data cannot be counted on for calibration or confirmation of models that concern the future (Oreskes et al. 1994; Parker 2009; Stainforth et al. 2007a, 2007b; Steele and Werndl 2013). Third, a fundamental worry about confirmation is that there is *radical uncertainty* about other possible climate models that could account for the data. In the Bayesian framework, this implies that it is impossible to tell whether models are confirmed because the performance of the model at hand compared to all other possible models cannot be judged (Steele and Werndl 2013). For these reasons and others, some climate scientists argue that complex models are just possibilities (Stainforth et al. 2007a, 2007b). We return to this position in Section 6 below.

4. The Limits of Climate Projections

Some climate research is undertaken in an 'experimental' spirit: 'play' around with models with the aim to gain insight about the physics of the earth system and its processes. This does not require the ability to produce reliable projections. However, much research these days is undertaken (and funded) with the aim of generating projections about the actual future climate upon which policies are made and real-life decisions are taken. In these cases, it is necessary to quantify and understand how good those projections are likely to be.

It is doubtful that this question can be answered along traditional lines. One such line would be to refer to the confirmation of a model against historical data (Chapter 9 of IPCC (2013) discusses model evaluation in detail) and argue that the ability of a model to successfully reproduce historical data should give us confidence that it will perform well in the future too. It is unclear at best whether this is a viable answer. The problem is that climate projections for high-forcing scenarios take the system well outside any previously experienced state, and at least *prima facie* there is no reason to assume that success in low-forcing contexts is a guide to success in high-forcing contexts. Although we hope that the laws of physics will not alter in a warmer world, many of the relationships and parameters described numerically in models are derived semi-empirically, either by observation of real-world properties or by tuning the parameter in order to reproduce expected larger-scale behaviours. For example, a model calibrated on data from a world with the Arctic Sea covered in ice might no longer perform well when the sea ice is completely melted and the relevant dynamical processes are quite different. For this reason, calibration to past data has at most limited relevance for a model's predictive success (Oreskes et al. 1994; Stainforth et al. 2007a, 2007b; Steele and Werndl 2013). For a more optimistic take on tuning see Frisch (2015).

This brings into focus the fact that there is no general answer to the question of the trustworthiness of model outputs. Our aim here is to draw attention to the problem and make some progress in framing it. Assume the task is to predict the occurrence of an event sometime in the future under a given emission scenario. In general, the trustworthiness of such a prediction depends on at least five factors:

- The lead-time, detailing how far in the future the event is to occur. For instance, if in the year 2014 the increase in global mean temperature by 2050 is predicted, the lead-time is 36 years.
- The time scale, specifying the event's temporal extension. An event can occur at a specific instance of time (e.g. the maximum temperature in July 2090) or can be an average over a longer time period or a number of events (e.g. the average maximum summer temperature between 2080 and 2099).
- The spatial extension, specifying the event's geographical size. Events can be relatively precisely located (e.g. the number of storms hitting Florida between 2080 and 2099) or occur on larger scales (e.g. the average maximum wind speed over North America between 2080 and 2099).
- The specificity, indicating how precisely the event is circumscribed. Events can be only vaguely specified (e.g. there is an upward trend in the global mean temperature) or be given precisely (e.g. global mean temperature will have increased by 3.7 degrees by 2010).
- The physical understanding of the causal relationships involved. Some events are brought about by one well-understood physical mechanism, while others are the result of a number of poorly understood factors.

There is widespread consensus that predictions are better for longer time averages, larger spatial averages, low specificity and better physical understanding, and, all other things being equal, shorter lead times (nearer prediction horizons are easier to predict than longer ones). Global mean temperature trends are considered trustworthy, and it is generally accepted that the observed upward trend will continue (Oreskes 2007), although the basis of this confidence is usually a physical understanding of the greenhouse effect with which the models are consistent, rather than a direct reliance on the output of models themselves. The latest IPCC report (IPCC 2013, Summary for Policymakers, section D.1) professes that modelled surface temperature patterns and trends are trustworthy on the global and continental scale.

By contrast, controversy besets the other end of the spectrum. Some believe that current models are capable of providing detailed information about the future local climate. The *United Kingdom Climate Impacts Program's* UKCP09 project, for instance (Sexton et al. 2012; Sexton and Murphy 2012), aims to make high-resolution probabilistic projections of the climate up to 2100 based on HadCM3, a global climate model developed at the UK Met Office Hadley Centre. Probabilities are given for events on a 25 km grid for finely defined specific events such as changes in the temperature of the warmest day in summer, the precipitation of the wettest day in winter or the change in summer mean cloud amount, with projections given for 30-year segments out to 2100. It is projected, for instance, that under a medium emission scenario, there is a 0.5 'probability level central estimate' for the reduction in summer mean precipitation in central London to be between 20% and 30% (Jenkins et al. 2009, p. 36).

The trustworthiness – and policy-relevance – of such projections has been disputed. A model has structural model error if the model's dynamics differs from the dynamics in the target system. Frigg et al. (2014a) point out that any structural model error in nonlinear models may compromise the ability to generate decision-relevant predictions. Furthermore, there is little reason to expect that post-processing of model outputs can correct for the consequences of such errors (Frigg et al. 2014b, 2015), except by the introduction of subjective probabilities, as is used by IPCC (2013, Chapter 12). This casts doubt on our ability, today, to make trustworthy, high-resolution probabilistic projections out to the end of this century. The research question is to determine the timescales on which such projections are likely to be reliable, and beyond those timescales to estimate the effect of model inadequacy. Where is the boundary between trustworthy and non-trustworthy projections? That is, where in between global temperature trends and precise projections on a 25 km grid does trustworthiness come to an end? This is a crucial – and eminently policy-relevant – question in the epistemology of climate science, and one that is hitherto unsolved.[2]

5. *Coming to Grips With Uncertainty*

Uncertainty features prominently in discussions about climate models and yet it is a concept that is poorly understood and that raises many difficult questions. In most general terms, uncertainty is lack of knowledge. The first challenge is to circumscribe more precisely what is meant by 'uncertainty'. A typical reaction to this question is to draw a distinction between uncertainty and risk. On a common view, we are faced with risk if we are in a situation in which we are uncertain about the occurrence of specific events but are able to attach precise probabilities to them. If, by contrast, relevant probabilities are either unavailable, or only partially available, we face uncertainty.[3] For example, we make a decision under risk if we bet on getting 'black' when drawing balls from an urn which contains a known proportion of well-mixed black and red balls; we operate under uncertainty if we do not know the proportion or do not even know whether there are balls with colours other than black and red in the urn.

Describing uncertainty as a situation without precise probabilities provides a first point of orientation, but it leaves many difficult questions unanswered. First, what is the reason for there not being precise probabilities? Do we not have sufficient information to commit to a particular probability measure, or are we unsure about the event space itself? If the latter, are there known unknowns, or do we reckon with unknown unknowns? Second, can uncertainty be quantified? There not being probabilities does not *ipso facto* exclude the existence of other quantitative measures for uncertainty. Third, how can uncertainty be communicated to decision-makers? Fourth, what is a rational way of making decisions under uncertainty?

Neither is this list exhaustive nor are the questions independent from one another. We now review the contributions made to the first, second and third questions in the climate science context. The second and the fourth question have attracted some attention in decision theory, and more will be said about them from that point of view in Part III.

One reaction to the first problem is to devise a classification of different kinds of uncertainties. A number of proposals have been made, but the discussion is still in a 'pre-paradigmatic' phase. Smith and Stern (2011) identify four relevant varieties of uncertainty: imprecision, ambiguity, intractability and indeterminacy. Spiegelhalter and Riesch (2011) consider a five-level structure with three within-model levels – event, parameter and model uncertainty – and two extra-model levels concerning acknowledged and unknown inadequacies in the modelling process. Wilby and Dessai (2010) discuss the issue with reference to what they call the cascade of uncertainty, studying how uncertainties magnify as one goes from assumptions about future global emissions of greenhouse gases to the implications of these for local adaption. Petersen (2012, Chapters 3 and 6) introduces a so-called uncertainty matrix listing the sources of uncertainty in the vertical and the sorts of uncertainty in the horizontal direction. Lahsen (2005) looks at the issue from a science studies point of view and discusses the distribution of uncertainty as a function of the distance from the site of knowledge production. And these are but a few of the many proposals currently available.

In response to the second problem, a number of methods to quantify uncertainty have been devised. Among these, ensemble methods occupy centre stage, and we turn to them in the next section. A number of authors emphasise the limitations of model-based methods (such as ensemble methods) and submit that any realistic assessment of uncertainties will also have to rely on other factors, most notably expert judgement. Petersen (2012, Chapter 4) outlines the approach of the Netherlands Environmental Assessment Agency (PBL), which sees expert judgement and problem framings as essential components of uncertainty assessment. Aspinall (2010) suggest using methods of structured expert elicitation. Specifically, he proposes the use of the so-called Cooke method, which is based on individual expert elicitation and followed by weighted judgement aggregation to assess the level of future climate uncertainty. So quantifying uncertainty is no less a formidable task than classifying it.

As regards the third problem, the most prominent framework for communicating uncertainty is the IPCC's. The latest version of the framework, which is used throughout AR5, is explicated in the 'Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties' and further explicated in Mastrandrea et al. (2011). The framework qualifies the 'degree of certainty' concerning a particular result on a qualitative scale from very low to very high. The degree of certainty is a function of two features: the type of evidence and the degree of agreement within the scientific community. The Guidance Note also advises communicating uncertainty at an appropriate level of precision and gives examples of a number of different levels. A discussion of this framework can be found in Budescu et al. (2014) and Adler and Hadorn (2014).

It should also be noted that the role of ethical and social values in relation to uncertainties in climate science is controversially debated. Winsberg (2012) appeals to complex simulation modelling to argue that results produced by climate scientists are influenced by ethical and social values. More specifically, his argument is that assignments of probabilities to hypotheses about future climate change are influenced by ethical and social values because of the way these values come into play in the building and evaluating of climate models. Parker (2014) objects that often it is not social or ethical values but pragmatic factors that play a role in evaluating and building climate models. She also objects that focussing on precise probabilistic uncertainty estimates is misguided and argues that uncertainty about future climate change is more appropriately depicted with coarser estimates, which are less influenced by values. She concludes that while Winsberg's argument is not mistaken, it exaggerates the influence of ethical and social values. Finally, Parker argues that a more traditional challenge to the value-free ideal of science might fit to the climate case: it can be argued that estimates of uncertainty are themselves always somewhat uncertain; thus, offering a particular estimate of uncertainty involves value judgements (cf. Douglas 2009; for further literature on this topic, see also Inteman 2015).

6. Model Ensembles

Many investigations into the climate system are now made with an ensemble of models rather than with a single model.[4] Current estimates of climate sensitivity and increase in global mean temperature under various emission scenarios, for instance, include information derived from ensembles containing multiple climate models. The reason to use ensembles are the acknowledged uncertainties in individual models, which concern both the model structure and the values of parameters in the model. It is a common assumption that ensembles help mitigate the effects of these uncertainties either by producing and identifying 'robust' predictions, or by providing estimates of the uncertainty about future climate change.

Before discussing the epistemic function of ensembles, a distinction needs to be made between two types of ensembles. As we have seen above, a climate model has a number of parameters in it. Some represent physical magnitudes such as the viscosity of water, while others are 'effective summaries' of sub-grid processes that are not explicitly resolved (such as cloud coverage). A *perturbed parameter ensemble* (PPE, sometimes alternatively a 'perturbed physics ensemble') explores how sensitively the outputs of one model depend on the parameters by running the same underlying model a number of times, each time with different parameter values. In this way, the ensemble explores the impact of parametric uncertainty on predictions (i.e. a sensitivity analysis with respect to the chosen parameters). By contrast, a *multi model ensemble* (MME) consists of several different models – i.e. models that differ in mathematical structure and physical content rather than only in parameter values. Such an ensemble is used to investigate how predictions of relevant climate variables are impacted by uncertainty about the model structure.

A result is *robust* if all or most models in the ensemble show the same result. If, for instance, all models in an ensemble show more than a 4° increase in global mean temperature by the end of the century when run under a certain emission scenario, this result is robust. Lack of robustness is often regarded as a good reason to doubt a result; there is also a pervasive intuition in the scientific community that robustness warrants increased confidence in predicted outcome.

Does robustness justify increased confidence? Parker (2011) discusses this problem in detail and reaches a sober conclusion: 'When today's climate models agree that an interesting hypothesis about future climate change is true, it cannot be inferred [...] that the hypothesis is likely to be true or that scientists' confidence in the hypothesis should be significantly increased or that a claim to have evidence for the hypothesis is now more secure' (Parker 2011, 579). One of the main problems is that if today's models share the same technological constraints posed by today's computer architecture and understanding of the climate system, then they inevitably share some common errors. Indeed, such common errors have been widely acknowledged (see, for instance, Knutti et al. 2010), and studies have demonstrated and discussed the lack of model independence (Bishop and Abramowitz 2013; Jun et al. 2008, 2008). But, if models are not independent, then agreement between them carries little, if any, epistemic weight.

It is important to be clear on the nature of this problem. The difficulty does not lie in the existence *per se* of these common errors, which is a normal situation in many modelling contexts. The difficulty lies with the statistical methods used in the analysis of these ensemble experiments. These methods assume independence in the sense that the errors are assumed to be randomly distributed about the correct answer. If the shared errors could be quantified and removed with reference to data, as is common practice in weather forecasting where large amounts of data are available for calibration purposes, then it would not be a problem at all. In the case of climate forecasting, where calibration data are of limited availability (due to long timescales) and limited applicability (due to changes in the earth system over time taking us into extrapolatory conditions), this is almost impossible to do. But, the use of these methods without reference to shared inadequacies of models is likely to result in over-confident prediction. Agreement between models which have been constructed by genuinely different approaches would indeed give epistemic support, but in practice, most approaches to this kind of modelling are built upon the same physical understanding, the same observations and the same computational methods.

When ensembles do not yield robust predictions, then the spread of results within the ensemble is often used to quantitatively estimate the uncertainty of the outcome. There are two main approaches to this. The first approach aims to translate the histogram of model results directly into a probability distribution. These methods assign higher probabilities to outcomes on which more ensemble members agree: in effect, the guiding principle is that the probability of an outcome is proportional to the fraction of models in the ensemble which produce that result. The thinking behind this method is first to treat models as exchangeable sources of information (in the sense that there is no reason to trust one ensemble member any more than any other) and then invoke a frequentist approach to probabilities. As we have seen above, the assumption that models are independent is problematic for a number of reasons. But, there is a further problem: current MMEs are 'ensembles of opportunity', grouping together existing models. Even the best currently available ensembles such CMIP5 are not designed to systematically explore all possibilities (nor were they developed independently), and it is therefore possible that there are vast classes of models that produce entirely different results. There is thus no reason to assume that the frequency of ensemble results should double as a guide to probability. The IPCC acknowledges this limitation (see discussion in Chapter 12 of IPCC 2013) and thus downgrades the assessed likelihood of ensemble-derived confidence intervals. For instance, the 5–95% range of model results for global mean temperature change in 2100, under forcing scenario RCP8.5, is 2.6 to 4.8 degrees (IPCC 2013, Table SPM.2). This result is not presented as 'very likely' (>90% probability), which would correspond to a direct use of model frequencies as probabilities; instead, it is deemed only 'likely' (>66% probability). In this way, the ensemble information is used but supplemented with expert judgement about the chance that models are misinformative. In effect, 24% of the probability mass has been reassigned in an undetermined manner, which we might interpret as a probability of approximately up to one-in-four that something occurs which the models are incapable of simulating. Note that the IPCC's assignment of numerical probabilities to certain terminology is intentionally imprecise, which in this context reflects the difficulty in interpretation of language, the need for fairly wide categories and the impossibility of accurately quantifying subjective judgements about the capability of models. The important point is that this reassignment is non-negligible and points to a significant likelihood of climate model inadequacy resulting in incorrect projections for 2100.

A more modest approach regards ensemble outputs as a guide to possibility rather than probability. On this view, the spread of an ensemble presents the range of outcomes that cannot be ruled out. The bounds of this set of results – sometimes referred to as a 'non-discountable envelope' – provide a lower bound of the uncertainty (Stainforth et al. 2007b). In this spirit, Katzav (2014) argues that a focus on prediction is misguided and that models ought to be used to show that certain scenarios are real possibilities.

While undoubtedly less committal than the probability approach, non-discountable envelopes also raise serious questions. The first is the relation between non-discountability and possibility. Non-discountable results are ones that cannot be ruled out. How is this judgement reached? Do results which, given current knowledge, cannot be ruled out indicate possibilities? If not, what is their relevance for estimating lower bounds? Betz (2015), for instance, argues that climate model results cannot be straightforwardly interpreted as serious possibilities because their assumptions are known to be false. Furthermore, it is important to keep in mind that the envelope just represents some possibilities. Hence, it (as the IPCC method described above) does not indicate the *complete* range of possibilities, making certain types of formalised decision-making procedures impossible.

7. *Conclusion*

In this paper, we have reviewed issues and questions that arise in connection with climate models. We have done this from a philosophy of science perspective, and with special focus on epistemological issues. Needless to say, this is not the only perspective. Much can be said about climate science from the points of view of science studies, sociology of science and political theory. For want of space, we have not been able to review contributions from these perspectives and refer the interested reader to Oreskes and Conway (2012) for an analysis of climate change from such an angle.

Acknowledgement

Many thanks to Reto Knutti, Wendy Parker, Lenny Smith, Dave Stainforth and Katie Steele for valuable discussions and suggestions. Funding support for the research was provided by the Arts and Humanities Research Council (via the 'Managing Severe Uncertainty' project) and by the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council. Frigg further acknowledges financial support from the grant FFI2012-37354 of the Spanish Ministry of Science and Innovation (MICINN).

Short Biographies

Roman Frigg was born in Basel, Switzerland. He holds a PhD in Philosophy from the University of London and master's degrees in both theoretical physics and philosophy from the University of Basel. He is Professor of Philosophy in the Department of Philosophy, Logic and Scientific Method, Director of the Centre for Philosophy of Natural and Social Science and Co-Director of the Centre for the Analysis of Time Series at the London School of Economics and Political Science. He is a permanent visiting professor in the Munich Centre for Mathematical Philosophy of the Ludwig-Maximilians-University Munich. He held visiting appointments in the Rotman Institute of Philosophy of the University of Western Ontario, the Descartes Centre for the History and Philosophy of the Sciences and the Humanities of the University of Utrecht, the Sydney Centre for the Foundations of Science of the University of Sydney and the Department of Logic, History and Philosophy of Science of the University of Barcelona. He is associate editor of the *British Journal for the Philosophy of Science* and member of the steering committee of the European Philosophy of Science Association, and he serves on a number of editorial and advisory boards. Professor Frigg is a member of the European Society for the Philosophy of Science, the Philosophy of Science Association and the British Society for the Philosophy of Science.

Erica Thompson was born in Scotland. She has a PhD in physics from Imperial College, London (2013), and an MMath from the University of Cambridge (2007). She is presently a Research Officer in the Centre for the Analysis of Time Series within the Department of Statistics at the London School of Economics and Political Science. She has previously worked for the Grantham Institute for Climate Change at Imperial College and the UK Energy Research Centre. Her current research interests include quantification of scientific uncertainty, the use and utility of climate information for real-world decision-making and the interpretation and communication of climate model output. Dr Thompson is or has previously been a member of the Royal Meteorological Society, the American Geophysical Union, Scientists for Global Responsibility and the British Beekeepers' Association.

Charlotte Werndl was born in Salzburg. She completed a PhD in Philosophy at the University of Cambridge in 2010 and master's degrees in both mathematics and philosophy at the University of Salzburg in 2006. She is Professor of Logic and Philosophy of Science at the Department of philosophy at the University of Salzburg, Austria, a Visiting Professor at the Centre for Philosophy of Natural and Social Science at the London School of Economics and an affiliate of the Grantham Research Institute on Climate Change and the Environment at the London School of Economics. Previously, she was an Associate Professor at the Department of Philosophy, Logic and Scientific Method at the London School of Economics and before that a research fellow at the University of Oxford. She is an editor of the *Review of Symbolic Logic* and an associate editor of the *European Journal for the Philosophy of Science* and serves on a number of editorial and advisory boards. She has published papers on climate change, statistical mechanics, mathematical knowledge, chaos theory, predictability, confirmation, evidence, determinism, indeterminism, observational equivalence and underdetermination. Her current work focuses on the philosophy of climate science, evidence and the philosophy of statistics and the foundation of statistical mechanics. Professor Werndl is a member of the Aristotelian Society, the British Society for the Philosophy of Science, the European Society for the Philosophy of Science and the Philosophy of Science Association.

Notes

* Correspondence: Department of Philosophy, University of Salzburg, Austria, and Department of Logic and Philosophy Logic and Scientific Method, London School of Economics, UK. Email: charlotte.werndl@sbg.ac.at

[1] Names are listed alphabetically. To contact the authors, write to r.p.frigg@lse.ac.uk; e.l.thompson@lse.ac.uk; c.s.werndl@lse.ac.uk.

[2] For a discussion of the impact of UKCP09 on planning policies, see Heaphy (2015).

[3] This distinction has become standard in the philosophical literature on decision theory (see, for instance, Sturgeon 2008). It is often attributed to Knight (1921), but it is hard to pin it down in the original text, and Knight scholars have doubted that the attribution is correct (Cooke 2014). Note, however, that this terminology is not standardised across scientific disciplines, and in many circumstances, exactly the opposite definitions are used.

[4] Parker (2013) provides an excellent discussion of ensemble methods and the problems that attach to them.

Work Cited

Adler, C. E. and G. H. Hadorn. 'The IPCC and Treatment of Uncertainties: Topics and Sources of Dissensus.' *Wiley Interdisciplinary Reviews: Climate Change* 5.5 (2014): 663–676.

Aspinall, W. 'A Route to More Tractable Expert Advice.' *Nature* 463 (2010): 294–5.

Betz, G. 'Are Climate Models Credible Worlds? Prospects and Limitations of Possibilistic Climate Prediction.' *European Journal for Philosophy of Science* 5.2 (2015): 191–215.

Bishop, C. H. and G. Abramowitz. 'Climate Model Dependence and the Replicate Earth Paradigm.' *Climate Dynamics* 41 (2013): 885–900.

Budescu, D. V., H. Por, S. B. Broomell, and M. Smithson. 'The Interpretation of IPCC Probabilistic Statements Around the World.' *Nature Climate Change* 4 (2014): 508–512.

Cooke, R. M. Deep and Shallow Uncertainty in Messaging Climate Change. *Safety, Reliability and Risk Analysis: Beyond the Horizon*. Eds. R. Steenbergen, P. van Gelder, S. Miraglia, and A. Vrouwenvelder. Boca Raton, Florida: CRC Press, 2014. 13–25.

Dessler, A. *Introduction to Modern Climate Change*. Cambridge: Cambridge University Press, 2011.

Frigg, R. P., S. Bradley, H. Du, and L. A. Smith. 'Laplace's Demon and the Adventures of His Apprentices.' *Philosophy of Science* 81.1 (2014a): 31–59.

Frigg, R. P., D. A. Stainforth, and L. A. Smith. 'The Myopia of Imperfect Climate Models: The Case of UKCP09.' *Philosophy of Science* 80.5 (2014b): 886–897.

Frigg, R., D. A. Stainforth, and L. A. Smith. 'An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of UKCP09.' *Forthcoming in Synthese* (2015). DOI:10.1007/s11229-015-0739-8.

Frisch, M. 'Predictivism and Old Evidence: A Critical Look at Climate Model Tuning.' *European Journal for Philosophy of Science* 5.2 (2015): 171–190.

Golaz, J.-C., L. W. Horowitz, and H. Levy. 'Cloud Tuning in a Coupled Climate Model: Impact on 20th Century Warming.' *Geophysical Research Letters* 40.10 (2013): 2246–2251.

Harvey, D. and R. K. Kaufmann. 'Simultaneously Constraining Climate Sensitivity and Aerosol Radiative Forcing.' *Journal of Climate* 15.20 (2002): 2837–2861.

Held, I. M. 'The Gap Between Simulation and Understanding in Climate Modeling.' *Bulletin of the American Meteorological Society* 80 (2005): 1609–1614.

Heaphy, L. J. 'The Role of Climate Models in Adaptation Decision-Making: The Case of the UK Climate Projections 2009.' *European Journal for the Philosophy of Science* 5.2 (2015): 233–257.

Inteman, K. 'Distinguishing Between Legitimate and Illegitimate Values in Climate Modeling.' *European Journal for the Philosophy of Science* 5.2 (2015): 217–232.

IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Eds. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley. Cambridge and New York: Cambridge University Press, 2013.

Jenkins, G., J. Murphy, D. Sexton, J. Lowe, and P. Jones. *UK Climate Projections Briefing Report*. DEFRA: , 2009.

Jun, M., R. Knutti, and D. W. Nychka. 'Local Eigenvalue Analysis of CMIP3 Climate Model Errors.' *Tellus A* 60.5 (2008): 992–1000.

——. 'Spatial Analysis to Quantify Numerical Model Bias and Dependence: How Many Climate Models Are There?' *Journal of the American Statistical Association* 103.483 (2008): 934–947.

Katzav, J. 'The Epistemology of Climate Models and Some of Its Implications for Climate Science and the Philosophy of Science.' *Studies in History and Philosophy of Modern Physics* 46 (2014): 228–238.

Knight, F. H. *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner & Marx, 1921.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl. 'Challenges in Combining Projections from Multiple Climate Models.' *Journal of Climate* 23.10 (2010): 2739–2758.

Knutti, R., T. F. Stocker, F. Joos, and G.-K. Plattner. 'Probabilistic Climate Change Projections Using Neural Networks.' *Climate Dynamics* 21 (2003): 257–272.

Lahsen, M. 'Seductive Simulations? Uncertainty Distribution Around Climate Models.' *Social Studies of Science* 35.6 (2005): 895–922.

Lenhard, J. and E. Winsberg. 'Holism, Entrenchment, and the Future of Climate Model Pluralism.' *Studies in History and Philosophy of Modern Physics* 41 (2010): 253–262.

Lloyd, E. A. 'Confirmation and Robustness of Climate Models.' *Philosophy of Science* 77 (2010): 971–984.

Mastrandrea, M. D., K. J. Mach, G.-K. Plattner, O. Edenhofer, T. F. Stocker, C. B. Field, K. L. Ebi, and P. R. Matschoss. 'The IPCC AR5 Guidance Note on Consistent Treatment of Uncertainties: A Common Approach Across the Working Groups.' *Climatic Change* 108 (2011): 675–691.

McGuffie, K. and A. Henderson-Sellers. *A Climate Modelling Primer*. New Jersey: Wiley, 2005.

Michaelsen, J. 'Cross-Validation in Statistical Climate Forecast Models.' *Journal of Climate and Applied Meteorology* 26 (1987): 1589–1600.

O'Neill, B. C., E. Kriegler, K. Riahi, K. L. Ebi, S. Hallegatte, T. R. Carter, R. Mathur, and D. P. van Vuuren. 'A new Scenario Framework for Climate Change Research: The Concept of Shared Socioeconomic Reference Pathways.' *Climatic Change* 122.3 (2014): 387–400.

Oreskes, N. 'The Scientific Consensus on Climate Change: How Do We Know We're Not Wrong?.' *Climate Change: What It Means for Us, Our Children, and Our Grandchildren*. Eds. J. F. C. DiMento and P. Doughman. Boston: MIT Press, 2007. 65–99.

Oreskes, N. and E. M. Conway. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York: Bloomsbury Press, 2012.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 'Verification, Validation, and Confirmation of Numerical Models in the Earth Science.' *Science New Series* 263.5147 (1994): 641–646.

Parker, W. S. 'Confirmation and Adequacy for Purpose in Climate Modelling.' *Aristotelian Society Supplementary Volume* 83.1 (2009): 233–249.

—— 'When Climate Models Agree: The Significance of Robust Model Predictions.' *Philosophy of Science* 78.4 (2011): 579–600.

—— 'Ensemble Modeling, Uncertainty and Robust Predictions.' *Wiley Interdisciplinary Reviews: Climate Change* 4.3 (2013): 213–223.

—— 'Values and Uncertainties in Climate Prediction, Revisited.' *Studies in History and Philosophy of Science Part A* 46 (2014): 24–30.

Petersen, A. C. *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*. Boca Raton, Florida: CRC Press, 2012.

Qu, X. and A. Hall. 'On the Persistent Spread in Snow-Albedo Feedback.' *Climate Dynamics* 42 (2014): 69–81.

Schmidt, G. and S. Sherwood. 'A Practical Philosophy of Complex Climate Modelling.' *European Journal for Philosophy of Science* 5.2 (2015): 149–169.

Sexton, D. M. H., J. M. Murphy, M. Collins, and M. J. Webb. 'Multivariate Probabilistic Projections Using Imperfect Climate Models. Part I: Outline of Methodology.' *Climate Dynamics* 38 (2012): 2513–42.

Sexton, D. M. H. and J. M. Murphy. 'Multivariate Probabilistic Projections Using Imperfect Climate Models. Part II: Robustness of Methodological Choices and Consequences for Climate Sensitivity.' *Climate Dynamics* 38 (2012): 2543–58.

Shackley, S., P. Young, S. Parkinson, and B. Wynne. 'Uncertainty, Complexity and Concepts of Good Science in Climate Change Modelling: Are GCMs the Best Tools?' *Climatic Change* 38 (1998): 159–205.

Smith, L. A. and N. Stern. 'Uncertainty in Science and Its Role in Climate Policy.' *Phil. Trans. R. Soc. A* 369.1956 (2011): 4818–4841.

Spiegelhalter, D. J. and H. Riesch. 'Don't Know, can't Know: Embracing Deeper Uncertainties When Analysing Risks.' *Phil. Trans. R. Soc. A* 369 (2011): 4730–4750.

Stainforth, D. A., M. R. Allen, E. R. Tredger, and L. A. Smith. 'Confidence, Uncertainty and Decision-support Relevance in Climate Predictions.' *Philosophical Transactions of the Royal Society A* 365 (2007a): 2145–2161.

Stainforth, D. A., T. E. Downing, R. Washington, A. Lopez, and M. New. 'Issues in the Interpretation of Climate Model Ensembles to Inform Decisions.' *Philosophical Transactions of the Royal Society A* 365 (2007b): 2163–2177.

Steele, K. and C. Werndl. 'Climate Models, Confirmation and Calibration.' *The British Journal for the Philosophy of Science* 64 (2013): 609–635.

——. The Need for a More Nuanced Picture on Use-Novelty and Double-Counting. *Philosophy of Science*, forthcoming (2014).

Sturgeon, S. 'Reason and the Grain of Belief.' *Noûs* 42 (2008): 139–165.

Wilby, R. L. and S. Dessai. 'Robust adaptation to climate change.' *Weather* 65.7 (2010): 180–185.

Winsberg, E. 'Values and Uncertainties in the Predictions of Global Climate Models.' *Kennedy Institute of Ethics Journal* 22.2 (2012): 111–127.

Worrall, J. Error, Tests, and Theory Confirmation. *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*. Eds. D. G. Mayo and A. Spanos. *Cambridge: Cambridge University Press*, 2010. 125–154.