On the appropriate and inappropriate uses of probability distributions in climate projections, and some alternatives

Forthcoming in *Climatic Change*

Joel Katzav[1], Erica L. Thompson[2], James Risbey[3], David A. Stainforth[4,5], Seamus Bradley[6], Mathias Frisch[7]

1 School of Historical and Philosophical Inquiry, The University of Queensland, St. Lucia, Queensland 4072, Australia
2 Data Science Institute, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK
3 CSIRO Oceans & Atmosphere, Hobart, Tasmania, Australia
4 Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK
5 Department of Physics, University of Warwick, Coventry, CV4 7AL, UK
6 University of Leeds, Leeds, LS2 9JT, UK
7 Institute of Philosophy, Leibniz University Hannover, 30167 Hannover, Germany

Abstract

When do probability distribution functions (PDFs) about future climate misrepresent uncertainty? How can we recognise when such misrepresentation occurs and thus avoid it in reasoning about or communicating our uncertainty? And when we should not use a PDF, what should we do instead? In this paper we address these three questions. We start by providing a classification of types of uncertainty and using this classification to illustrate when PDFs misrepresent our uncertainty in a way that may adversely affect decisions. We then discuss when it is reasonable and appropriate to use a PDF to reason about or communicate uncertainty about climate. We consider two perspectives on this issue. On one, which we argue is preferable, available theory and evidence in climate science basically excludes using PDFs to represent our uncertainty. On the other, PDFs can legitimately be provided when resting on appropriate expert judgement and recognition of associated risks. Once we have specified the border between appropriate and inappropriate uses of PDFs, we explore alternatives to their use. We briefly describe two formal alternatives, namely imprecise probabilities and possibilistic distribution functions, as well as informal possibilistic alternatives. We suggest that the possibilistic alternatives are preferable.

Keywords: climate projection; uncertainty representations; probability; deep uncertainty; possibility theory

1. Introduction: the common use of PDFs to represent future weather and climate

Information about future climate is gained from past experiences and observations, conceptual/theoretical understanding of relevant physical processes, forward evolution of simulation models, and expert judgement. In the case of weather, predictions are largely developed with forecast models. In each case, the available information is not sufficient to enable one to write down a single unique description of the future state, so that we must somehow represent our uncertainty with a range of outcomes. One common way of doing this is to generate probability distribution functions (PDFs) (see, e.g., IPCC [2013], Lowe et al. [2018] and Lee and Marotske [2021]). Full PDFs are distributions in which each possible outcome is specified and assigned a specific weight, a probability, with the total probability adding up to 1. Partial PDFs are distributions in which a range of possible outcomes are specified, and the range is, at least, assigned a qualitative probability, e.g., the range is taken to be the likely or very likely range. Except where otherwise specified, the (partial or full) PDFs we discuss here are distributions of real-world outcomes (and not of, for example, model runs). This paper is aimed at an interdisciplinary audience of producers and consumers of PDFs in the climate context. We aim to clarify, for this audience, when it is appropriate to use PDFs to represent uncertainty about future climate and how this uncertainty should be represented when using PDFs is not appropriate. We develop a general argument to answer these questions, largely by drawing on existing literature in climate science and philosophy about representing uncertainty and about problems with the use of PDFs in climate science. We use work in philosophy, in particular, to provide a general understanding of uncertainty and of when not to use PDFs.

We draw, in section 2, on the philosophy of probability [Hájek, 2019] to provide a general characterisation of kinds of uncertainty and how these might be misrepresented. We also use this characterisation to illustrate some of the ways in which PDFs might misrepresent our uncertainty about future weather or climate and some of the effects such misrepresentation might have on decision making.

We go on, in section 3, to consider the conditions under which it is appropriate to provide PDFs for future climate. We do so with the help of our characterisation of uncertainty and with the help of worries that scientists and philosophers of science have raised about representing uncertainty about future climate with PDFs. Such worries have been based on the limited opportunities for quantitative evaluation of PDFs, reliance on ensembles of opportunity in their generation and the limitations of supporting theory about the climate system (see, e.g., Hall et al. [2007], Stainforth et al. [2007], Parker [2010], Knutti et al. [2010], Katzav [2014] and Baumberger et al. [2017]). We go beyond existing discussions by considering the conditions in which their worries about the use of PDFs apply, including whether the worries can be mitigated by expert appeals to a variety of types of evidence. Moreover, we provide advice about when using PDFs is appropriate. We follow existing practice (e.g., Stainforth [2007] and Parker [2010]) and contrast the case of climate projections with that of weather prediction to illuminate some of the challenges of the former.

Section 3 presents two perspectives on whether PDFs are appropriate in the climate context. One, which reflects current practice, but we argue is problematic, is to continue to use PDFs but to explicitly recognise the limited representation of uncertainty they provide and their resulting limitations as tools in supporting policy decisions. Alternatively, there is our preferred perspective, namely that PDFs should not be used in the climate context.

Finally, in section 4, we explore two alternatives to PDFs, one provided by imprecise probability theory and one by possibility theory. We note that imprecise probabilities permit improving on PDFs but fail to avoid the key problems with their use. Possibility theory is our preferred option. Section 5 is our conclusion.

2. When probabilities misrepresent uncertainty

2.1 Three kinds of uncertainty

Our focus here will be on three of the main kinds of uncertainty, though these are usually presented in a probabilistic context while we extend them to cover non-probabilistic uncertainty [Halpern 2017; Hájek, 2019]. The first kind concerns how events will unfold over time. This is the kind of uncertainty we seem to be talking about when we say that the chance that a coin will land heads is 50%, or that the chance of rain in South Brisbane tomorrow is about 40%. Here, uncertainty is a measure of the potentiality, propensity, or frequency, of certain kinds of events. We can call such uncertainty **aleatoric uncertainty**. It is also sometimes called **objective uncertainty,** since it seems to refer to objective features of the world (the potentialities, propensities or frequencies).

The second kind of uncertainty comprises how confident we are in propositions. This is the kind of uncertainty we seem to be talking about when we say things such as that we have no doubt that Earth is not flat and that we would bet our lives on it, or that we have very high confidence that global warming has caused the cryosphere to shrink. Here, uncertainty is a measure of the strength of an individuals', or a group's collective, beliefs and thus can be called **subjective uncertainty**. Importantly, we take subjective uncertainties to be actual degrees of belief. Subjective uncertainty is, by contrast, often identified with degrees of belief that cohere, i.e., obey the axioms of probability theory [Hájek, 2019]. We avoid such an identification so as to cover probabilistic and non-probabilistic uncertainty.

The third kind of uncertainty captures the degree of support an individual's evidence or data, or a group's collective evidence or data, provides for an hypothesis. This is the kind of uncertainty we seem to be talking about when we say that general relativity is highly probable given the evidence for it, or that attribution studies provide support for the thesis that global warming is influencing extreme weather events. We will call such uncertainty **evidential uncertainty**. The difference between subjective and evidential uncertainty is that the former is concerned with our beliefs while the latter is concerned with logical relations between propositions or statements we might believe.

It is possible to categorise uncertainties differently, or to argue that, strictly speaking, there is only one kind of uncertainty and that the others are reducible to it or are somehow based on confusion [Hájek, 2019]. For example, one might argue that one's evidential uncertainty really is just the subjective degree of belief one would have if one's beliefs cohered. We, however, propose to bracket the question whether there is one kind of uncertainty. Instead, we use our understanding of the different kinds of uncertainty as guides to when it is appropriate to represent uncertainty using probabilities.

2.2 How precise probabilities might misrepresent uncertainty

Consider aleatoric uncertainty. It might be that there is no limit frequency, that is, no stable ratio to which the ratio between the number of events of a certain kind and specified intervals of time converges as the ratio is looked at across longer and longer periods of time [Fine, 1988]. For example, the number of purple scarves worn in winter may fluctuate in an unstable way over winters. In the climate context, we know that average temperatures on timescales of tens of thousands of years do not converge on a mean but instead fluctuate [Lovejoy, 2015]. Where the timescale of prediction is similar to or longer than the timescale of fluctuations in the limit frequency, it would be a mistake to represent the probability of the event using a precise probability.

Consider subjective uncertainty. In some circumstances, we do not have a precise degree of confidence in a prediction. Accordingly, to describe our confidence as having a precise probability would be to misrepresent it and thus to misrepresent our subjective uncertainty. For instance, I do not have any particular degree of confidence that a certain climate model provides a "good" representation of global mean precipitation change over the next eighty years: even though I might clearly be able to identify one model as "better" than another, they may both, as far as I can tell, be quite poor or quite good in absolute terms. Climate scientists sometimes exhibit such non-probabilistic subjective uncertainty [Millner et al., 2013].

Consider evidential uncertainty. In some cases, it makes sense to combine different, precise probabilistic projections into a single one, e.g., using Bayesian model averaging. In other cases, however, evidence is fundamentally ambiguous, pointing to incompatible precise probabilities for an hypothesis. In such cases, the evidence does not give the hypothesis a precise probability. For instance, we might have two competing models of the El Niño/La Niña–Southern Oscillation, each based on plausible but differing theoretical mechanisms. The two models offer substantially different, precise probabilistic forecasts about the formation of La Niña conditions towards the end of the year. In this case, it only makes sense to say that there is evidence pointing both ways.

In addition, many uncertain situations contain elements of more than one of these types of uncertainty, and a probability function which effectively represents one kind of uncertainty may not be adequate for representing another.

2.3 How PDFs might misrepresent uncertainty about weather and climate, and some potential risks of such misrepresentation

The above illustrative examples are of cases in which precise probabilities misrepresent uncertainty (in one sense or another of 'uncertainty'). Our focus now is on what are perhaps less obvious cases, cases in which PDFs misrepresent uncertainty. We explore how such cases can occur as well as why they might be problematic. We emphasise that here too the examples we give are hypothetical rather than real-world examples. We engage with a real-world example later, when (in section 3.3) we discuss the use of expert opinion by the IPCC to transform PDFs produced by climate model ensembles.

The underlying issue brought out in our examples, and subsequent discussion, is that when we misrepresent uncertainty in one way or another, we misrepresent what available evidence justifies. For example, if we present our evidential uncertainty as being within a narrower or more precise range than it is, we ignore evidence indicating possibilities that are outside of the range. Even when probabilities appear to misrepresent only a scientist's subjective uncertainty, this often also involves a misrepresentation of available evidence and what it supports, since scientists' subjective uncertainty is dependent on their familiarity with, and interpretation of, the evidence. Inaccurate or incomplete appraisals of the evidence, we take it, are inherently unreliable but are also potentially problematic in that they might misdirect research and lead to further false beliefs about the world. Such appraisals also threaten to be a poor basis for policy or, more broadly, for practical decisions.

2.3.1 Model frequencies misinterpreted as real-world probabilities

Consider a weather forecast for next week generated from a set of five simulations with a weather model. The "chance of rain" is derived directly from the number of model runs which show rain at any particular location and point in time. None of the runs show rain at my location in 10 days time at 4pm, so the "chance of rain" is given as 0%. However, comparisons between models and out-of-sample outcomes show that model-based weather predictions at these lead times are not fully reliable (see, e.g., Risbey et al. [2021]).

In this case, taking the frequencies represented in the model to be those of the weather system is, because of model unreliability, to misrepresent aleatoric uncertainty. It also involves misrepresenting our subjective and evidential uncertainty, because the unreliability is known. The identification of 0% probability is particularly unfortunate because it implies certainty. For instance, an outdoor event organiser might choose not to make contingency plans for rain based on false certainty that it will not occur.

2.3.2 Biased ensembles of opportunity

An ensemble of opportunity is an ensemble of models that happen to be available at a time [Tebaldi and Knutti, 2007]. Some ensembles of opportunity bring together available state-of-the-art climate models, that is, climate models that are individually designed to be as good as possible along a variety of dimensions. Examples of such ensembles of opportunity are those of the Coupled Model Intercomparison Project (CMIP) [Taylor et al, 2012; Katzav and Parker, 2015; Lee and Marotzke, 2021]. These ensembles are used to project full or partial PDFs for key climatic quantities over the coming century, including temperature change and precipitation. Unlike the case presented in 2.3.1, the focus is on spatially and temporally aggregated information such as ten-year means, medians or 90th percentiles of daily values. Collections of small numbers of model runs from each individual model, which are assumed to be probabilistically robust representations of model output, are combined to produce distributions of these quantities. Even if this assumption is accurate, however, there is evidence that the results from multiple models are PDFs which have their probability mass in the wrong place (as 'best guesses' with biases, including many shared biases, the models produce results that tend to cluster in the wrong place) and do not span the space of possibilities (as 'best guesses', the models do not adequately explore extremes, even of aggregated data about temperature and other quantities) [Taylor et al., 2012; Borodina et al., 2017; Lee and Marotzke, 2021].

In this case, PDFs do not represent the evidential uncertainty about the provided projections (since the projections are biased and overemphasise central ranges) and thus do not represent what our subjective uncertainty ought to be or even is. The potential results include, among other things, being overconfident that scenarios in the extreme values (tails) of the PDF distribution will not occur and being overconfident that the consensus is correct. Overconfidence here may lead to over-optimised adaptation strategies.

2.3.3 Failure to propagate assessed uncertainties through chains of models

The IPCC's fifth assessment [IPCC, 2013] used the outputs of CMIP to derive temperature-change projections by turning the observed model temperature-changes into a probability distribution of model runs. The authors then judge that the 90% model range is only "likely" (66%+) in the real world. Climate impact modellers, however, typically do not use inputs from outside the range of the models, even though the IPCC implies a nontrivial possibility of such an outcome (for example, Mendlik and Gobiet [2016] describe how to select a subset of simulations representative of the range of a larger ensemble, for use in impact modelling). Doing so might misrepresent IPCC expert judgement, and by implication the subjective and evidential uncertainties that underpin this judgement, either in an explicitly probabilistic way or by implying that a full range is covered (such as by selecting a "high", "medium" and "low" case from the available model runs). As a result, secondary projections may systematically underestimate uncertainty [Thompson et al, 2016] and inadvertently neglect a long tail of potential outcomes that may be of interest to decision makers.[1]

2.3.4 Misrepresenting the breadth of expert opinion

Consider a situation where an expert elicitation procedure is used to determine a probability distribution for future sea level. However, all experts are selected from one institution. Or perhaps all experts are ocean dynamicists, none are ice sheet specialists, and they do not take much account of the contributions or uncertainties related to ice sheet flows, stability and melt. A real-life case in which this occurred may simply involve a poorly designed study. In principle, however, experts may not be fully aware of diversity of opinion or of all relevant available information, or they may be worried about including non-mainstream views in studies. This is a case of misrepresenting the subjective and evidential uncertainty of the community and, as in the case of the ensemble of opportunity, might lead to over-optimised adaptation strategies.

3. Recognising when a PDF is (not) appropriate

A PDF is a formal way to distribute probability mass. The most fundamental attributes of such a distribution are the state space over which it is defined and its shape. Depending on the application for which it is to be used, a user may be interested in any function of the distribution, such as the mean/median value, the spread, the proportion beyond a certain threshold, or the tails.

[1] The 'H++' scenarios of future, regional United Kingdom sea levels are examples of explorations of extremes [Lowe et al., 2009].

There are some standard arguments for tending to stick with the application of probability theory in expressing uncertainty. These include the existence of well-developed theory that guides their updating in light of evidence and well-developed theory that guides decisions in light of the probabilities of future events [Halpern, 2017]. In addition, practical considerations may, depending on the context, favour the use of probabilities in communicating uncertainty [Hinkel et al., 2019].

That said, the formal non-probabilistic approaches discussed in the next section also have well-developed theories of updating and corresponding decision theories [Halpern, 2017]. Moreover, convenience and historical use are not, from a scientific perspective, good reasons for continuing to use an inappropriate methodology. We have described above some potential consequences for real-world decision making which stem from inappropriate use of PDFs. In this section, we more systematically consider when it is appropriate to provide PDFs that describe climate and offer some more concrete guidelines for identifying these circumstances.

When is a PDF appropriate? A simple answer is: when it represents what our subjective probability *ought* to be given available evidence, including evidence concerning our uncertainty. In such a case, subjective and evidential uncertainty will align. Ideally, we would also want these probabilities to match aleatoric probabilities (when these exist in the real world).

How can we detect when our PDF is appropriate? A first test for subjective uncertainty is simple: we ask whether we really believe what the PDF says. If the PDF is subject to caveats which are such that we do not believe it, it does not fully represent our uncertainty [Parker 2010]. In addition, however, we want the subjective uncertainty we arrive at to match the subjective degree of uncertainty we ought to have. Even if we already have a level of subjective uncertainty, we want to make sure that it is the one we ought to have. We know that our actual confidence in hypotheses often does not fully reflect available evidence, either because of biases in our reasoning [Benjamin, 2019; O'Hagan, 2019] or because, as is common in complex, interdisciplinary fields of research such as climate science, individuals only have a partial understanding of the relevant evidence.

Making sure that our subjective degree of belief is what it ought to be involves ensuring that our beliefs cohere, as far as we can tell, with each other. In particular, accepting a PDF should not involve misrepresenting our understanding of the evidential uncertainty, that is, accepting a PDF should not involve misrepresenting how much evidence is seen to support the PDF, including any limitations in our knowledge of the extent of this support.

In the remainder of this section, we apply this general requirement in working out, in more detail, the circumstances in which our PDFs might capture our uncertainty about climate and weather. In subsections 3.1 and 3.2, we note that two sufficient conditions for providing a PDF, namely repeated quantitative evaluations and theory-based evaluations, are not fulfilled in the case of climate projections, though the first is fulfilled in the case of weather predictions. We then provide, in 3.3, one perspective (P1) according to which having either a quantitative evaluation of a PDF or a theory-based one is also necessary for providing the PDF. Although we prefer perspective P1, we recognise that the use of PDFs is ubiquitous in climate science and unlikely to be deprecated soon. Thus, we also describe a second perspective (P2), that in special circumstances expert assessment can compensate for limited data and theory sufficiently to use PDFs. We bring out the challenges of this perspective by providing some necessary conditions for the adequacy of expert generated PDFs.

3.1 Repeated, quantitative evaluation of past probabilistic forecasts

In the case of a repeated forecast, such as a weather forecast, we can generate probabilistic predictions and use our data (actual observations over the forecast period) to determine how accurate our probabilistic predictions are [Risbey and O'Kane, 2011]. Ideally, in such a process, the probabilistic predictions match observed frequencies well, so that using the predictions to guide expectations means our degree of confidence will match evidential and aleatoric probabilities. Formal measures can assess the value of the information content of the forecast, and a user might choose to set some threshold of error-tolerance relating to the costs and losses associated with incorrect forecasts. Where the probabilistic predictions are found to be unacceptably inaccurate, they can be revised, and the performance of the new forecast quantitatively compared with the old one. For this reason, when high-quality data are available in sufficient quantity, quantitative evaluation against relevant out-of-sample data can be adopted as the gold standard for a defensible PDF.

Where, as in seasonal climate forecasts, relevant out-of-sample testing is possible (data quality is high) but only a small amount of data is available (data quantity is low), similar trials can be undertaken using formal measures of reliability, but statistical confidence in the assessment will be lower. Additional forecast-outcome data may be generated using past data/conditions ("hindcasts") and these can provide a good quantitative measure of reliability, though with the caveat that they are not truly out-of-sample even where rigorous cross-validation approaches are employed [Risbey et al., 2021].

In either of the above cases, if empirical reliability assessment suggests that the observed data are not consonant with the forecast distribution (within the threshold of user tolerance), then it should not be provided as a PDF. Further, even where repeated testing is possible, we need to be confident that the system is stationary in that it does not change significantly on timescales comparable with the timescales of the forecasts. Where this is not so, we have theoretical or empirical reason to suspect that our PDFs will not span the range of possibilities or will not have an appropriately distributed probability mass.

For most climate forecasts, we have little out-of-sample data on the relevant timescales. While, as noted above, seasonal climate forecasts are tested against a small amount of such data, it remains unclear what skill such forecasts have [Weisheimer and Palmer, 2014; Risbey et al., 2021]. On decadal and longer timescales, out-of-sample testing is even more limited, though it is used [Schmidt and Sherwood, 2015, Hausfather et al, 2020]. Moreover, in climate forecasts on all timescales of interest, we are unsure about the similarity of the testing conditions and the conditions obtaining in the future [Baumberger et al., 2018; Lee and Marotzke, 2021]. This, in combination with theoretical understanding of the system's nonlinearities, gives reason to doubt the reliability of bias correction methods [Risbey et al., 2021]. Therefore, we need to fall back on theoretical evaluation of model output or on more qualitative arguments about model quality and adequacy for purpose.

3.2 Theory-based evaluation of PDF credibility

In the absence of repeated trials against new evidence, our ability to explore the range of possible behaviours within a complex system such as climate is limited. More specifically, the opportunities for determining the envelope of possible evolutions of the system as well as the (aleatoric) relative likelihood of these evolutions will be limited.

Sometimes, extremely well confirmed theory might help−that is, extremely well confirmed general principles which are potentially true; unlike models, where their construction involves explicit idealisation. Extremely well confirmed theory might strongly guide modelling and enable understanding its limitations. Perhaps this is sometimes the case when theories of gravitation are used to predict aspects of the evolution of a solar system. In such cases, we also make idealisations, but theory tells us to what extent these will impact our predictions and thus how confident we can be in the predictions. With such constraints and understanding, PDFs or even precise predictions might be justified.

In climate modelling, however, theory provides limited guidance in model construction and in assessing model limitations [Gleckler et al., 2008; Parker, 2010; Katzav, 2014]. For example, though there is a well known set of equations governing flow on a sphere, there are no exact solutions, and the flow is subject to small scale processes that are poorly represented, yet impact even the largest scales (such as cloud and aerosol processes) [Lee and Marotzke, 2021]. So too, we have no highly confirmed, general and implementable quantitative theory of how patterns of internal variability develop and impact overall climate variability [Katzav, 2014]. Thus, although we have large-scale theories of climate that guide model development, including very high confidence in the greenhouse effect and large-scale geographical features such as polar amplification, we have no detailed theory of climate per se. This means, in particular, that theory provides us with limited guidance in constructing climate models and climate model ensembles. So too, theory provides limited guidance in interpreting model and ensemble output, including what model biases imply for output accuracy. Thus, theory tends to leave open the extent to which output spans the range of possibilities or whether these possibilities are weighted in a way that reflects reality.

Indirect empirical support resulting from the examination of retrospective forecasts over periods in the past can assist here. As we noted in the previous section, however, if these compare well with past data, then we must next ask whether the climate of any future forecast state is sufficiently similar to the past that we can take confidence from the assessment of the past. So too, we must ask whether alternative modelling efforts, which give rival projections, might also have adequately made the retrospective forecasts [Katzav, 2013]. With limited guidance from theory, it will typically not be possible to answer these questions in a principled way.

Another attempt to compensate for the limited theoretical guidance in model construction and evaluation uses model ensembles. A “Model Land” PDF can, to begin with, be constructed from a model by running it multiple times with slightly different inputs and then treating the resulting frequency of outcomes as a sample from a probability distribution which can be inferred. Initial condition ensembles are used in weather and near-term climate forecasting, for example, to capture the measurement uncertainty over initial conditions and propagate this through into a PDF output [Risbey et al., 2021]. Sampling of initial conditions is in practice usually extremely limited on longer climate scales, with, for example, individual CMIP models typically contributing fewer than 10 members to CMIP ensembles [Milinski et al, 2020]. But at least in the theoretical case of a perfect model, this procedure results in a perfect description of the *initial condition uncertainty*.

However, models are idealisations so that, in addition to initial condition uncertainty, there is also *parameter uncertainty* about the appropriate values of parameters in model equations, and *structural uncertainty* about the representation of the physical process by means of the model equations. One approach to estimating a lower bound of this uncertainty is to use perturbed parameter ensembles (variations of a single model by changing the parameters, essentially a sensitivity analysis), and a second approach is to use multiple-model ensembles (statistically analysing together a set of different models for the same output) [Baumberger, 2018]. In the absence of appropriate guidance from theory, however, it is unclear what range of model structures needs to be explored in order to estimate structural uncertainty in a principled way and ensembles of opportunity are used to provide substitutes for such estimates [Katzav and Parker, 2015]. Ensembles, therefore, do not fundamentally alter our inability to estimate uncertainty in the absence of guidance from adequate theory. We still do not have a principled way for judging the extent to which the models are exploring the full range of possibilities or are appropriately distributing the probability mass across possibilities.

3.3 Subjective evaluation of PDF credibility

In the absence of sufficient guidance either from theory or repeated testing, we must rely on expert judgement. In the following, we describe two alternative perspectives about the justification for the use of PDFs in this situation. According to perspective P1, which we outline first, PDFs should not be offered and alternative means of representing uncertainty should be used.

Expert judgement is based on one or more theory, model and data-based studies (including, e.g., on multiple-model ensemble studies and evidence from the palaeo-record of analogue cases). By hypothesis, our concern here is with circumstances in which theory is limited and sufficient data to directly evaluate reliability is unavailable. Further, our concern is with circumstances in which the models by which we draw conclusions from data have limitations that impact conclusions in unquantified, but substantial, ways. In such a situation, experts are to a substantial, but not fully known, extent in the dark about the space of possible hypotheses that might explain data, and thus about the space of possible projections that are compatible with the data. As a result, experts will not be able to assess the space of possible projections, never mind how likely they are. Any single PDF that experts produce will misrepresent (leave out) subjective uncertainty; experts will understand that it involves arbitrary bounding and weighting of projected possibilities. So too, a selected PDF would misrepresent evidential uncertainty, since the evidence permits a variety of ranges of possibilities and, within each range, of weightings of possibilities.

For example, experts might have to decide whether the latest CMIP ensemble's output captures the range of possible future evolution of precipitation in a given region in coming decades. They will understand that the ensemble includes substantial relevant biases relating to representation of clouds and convection, to representation of the spatial patterns and seasonal cycles of key precipitation-governing processes in the ocean and atmosphere and to representation of internal variability. So too, the ensemble does not include all relevant forcing factors and feedback mechanisms that could result in different climate forcing and/or different responses of rainfall to climate forcing [Risbey and O'Kane, 2011; Shepherd, 2014; Lee and Marotzke, 2021]. But what the limited ability to simulate these phenomena implies for the range and distribution of projections is unknown to a substantial degree, since neither very highly confirmed theory nor empirical evidence sufficient to determine this are available. Unfortunately, the models are the primary way of getting a handle on the evolution of these complex, highly-interdependent phenomena, so that our ability for model-independent assessment is limited. Thus, any provided PDF that results from transforming the CMIP ensemble's results will come with the caveat that it is unclear how well it fits the evidence and, as a result, will misrepresent subjective and evidential uncertainty.

To be sure, experts do bring additional sources of knowledge to the construction and evaluation of PDFs and thus can help us develop a better understanding of our evidential uncertainty. For example, in adjusting climate model parameters so that model output better fits data, climate modelers learn about the ranges within which parameters can be varied given the data. Modelers also learn about the extent to which accommodating the data requires specific modeling assumptions and thus about the extent to which the assumptions are robust. Modelers can make use of this knowledge during elicitation exercises [Schmidt et al., 2017]. Climate scientists have, further, knowledge relevant to judging the relative independence of the different studies upon which they draw in preparing a PDF. And where the various lines of evidence are judged to be somewhat independent, there may be increased confidence in a PDF (see, e.g., Sherwood et al. [2020] and Lee and Marotzke [2021]). However, experience with tuning reflects some further exploration of parameter space while knowledge of robustness of model assumptions reflects some further exploration of structural uncertainty. These sources of knowledge, as we have noted, do not by themselves appear sufficiently to compensate for the limited availability of extremely well confirmed theory. Nor does the appeal to a variety of evidence fundamentally change the situation. The absence of background theory that delimits the space of hypotheses that might explain the data means we cannot determine whether it might be equally, or better, explained by other hypotheses than those being worked with [Katzav, 2013 and 2014]. So, we are not in a position to estimate how confident we should be in the shared hypotheses explaining diverse lines of evidence and by implication, how confident we should be in the resulting PDFs.

An alternative perspective (P2), which reflects a significant strand of current practice, including in IPCC reports (e.g., IPCC [2013] and Lee and Marotzke [2021]), is to offer PDFs as the best probabilistic representations of evidential uncertainty currently achievable while acknowledging that they are unreliable to some unquantified extent and that guidance is not being provided about what to suppose if the PDF is misleading. In this way, probabilistic representations are kept while their limitations, including potential risks involved in their use, are recognised. Even if we are willing to accept these limitations, however, implementing P2 is challenging.

If a PDF is to be the best probabilistic representation of uncertainty achievable, it cannot merely be a plausible PDF or a consensus PDF (in the sense that it represents what the community is most confident about). A plausible PDF need not take into account information about uncertainty included in alternative plausible PDFs. A consensus PDF does not take into account the uncertainty represented in second-best alternatives. The best achievable PDF is rather one which somehow takes into account as much of the relevant, established uncertainty while minimizing the loss of information about uncertainty that results from insisting on a probabilistic representation. Practically, such a PDF can be developed by an ensemble of experts that reflect the diversity of the field of knowledge relevant to PDF variables. Such an ensemble of experts will have to make decisions to exclude information about uncertainty from PDFs where experts disagree. Arguably, which information to exclude will depend on what users want the information for and thus on the values of users [Parker and Winsberg, 2018]. There are various procedures available for eliciting, comparing and combining the PDFs of experts (see, e.g., O'Hagan [2019]); we focus on qualitative considerations for assessing whether a PDF captures uncertainty as well as possible. Some elicitation procedures do encourage this kind of discussion [O'Hagan, 2019]. Further, we focus on the use of expert judgement in combination with model ensembles, to bridge as far as feasible the gap between a distribution of model outcomes and reality. We spell out steps in which the appropriateness of PDFs generated in this way might be assessed in accord with P2. Such expert-based assessments should come with the already noted caveats that they are uncertain to an unquantified degree and thus are potentially a risky basis for action. We do not here tackle the difficult question of who counts as a domain expert but note that the answer is of first-order importance to results.

If model-derived distributions are to be transformed, using expert judgement, into real-world forecast PDFs, there are two key qualitative questions for assessing the quality of the model-derived PDFs. The first concerns the construction of the ensemble of models, which may take one of four forms:

i. Ad-hoc "ensemble of opportunity": just the models that happen to be available.
ii. Structured ensemble resulting from systematic variation of a subset of parameters of a single model.
iii. Structured ensemble resulting from systematic variation of all parameters of a single model (keeping in mind that the number of parameters in state-of-the-art climate models makes this unfeasible in their case).
iv. Unstructured ensemble of models resulting from deliberate attempt to maximise diversity of physical representations/approaches or model responses.

We have already noted that PDFs derived directly from ensembles of climate models, thus including cases (i)-(iv), cannot adequately represent our uncertainty. That said, our current task is to do the best to represent uncertainty with a PDF, so the question is how we can do better than merely accepting what the models tell us.

Case (i) clearly results in model frequencies that do not represent any kind of systematic sampling. Cases (ii) and (iii) are better in this regard. They are partial and full sensitivity analyses in model space. (ii) and (iii) come, however, in varying degrees of adequacy. Sensitivity analyses should explicitly look for nonlinearity or non-robustness to parameter variation, and highlight it if found, because this provides some indication of the extent to which the PDF is a realistic representation of current uncertainty rather than an artefact of relatively unconstrained modeling choices.

Case (iv) is likely to be more informative than (i)-(iii) due to the deliberate effort to increase uncertainty ranges. Experimental designs should therefore prioritise the maximisation of diversity within ensembles.

The second question, regarding the input of expert judgement, concerns the way in which the ensemble output is related to the real variable [Thompson and Smith, 2019]. This expert judgement could take one of the following forms:

a. We make an expert judgement that the model (ensemble) is perfect: the real world is statistically indistinguishable from the ensemble distribution.
b. We make an expert judgement that the model (ensemble) is perfect, minus some empirically-determinable "discrepancy" term or "bias-correction" procedure.

c. We make an expert judgement that the ensemble range probably contains the real-world outcome and the relative model frequency is a qualitative rather than quantitative guide to the more likely outcomes.
d. We make an expert judgement that the ensemble range contains the real-world outcome, with some probability.
e. We make an expert judgement that a synthesis of (possibly bias-corrected) ranges from two or more ensembles contains the real-world outcome, with some probability.
f. We make an expert judgement that the ensemble cannot be interpreted as a probabilistic guide to the real-world outcome.

Examples of most of the above strategies can be found in climate literature. For instance, UKCP18 [Lowe et al., 2018] take approach (b) for projections of UK climate variables to 2100 and the IPCC's Working Group 1 [IPCC, 2013; Lee and Marotzke, 2021] take approaches (d) and (e) for projections of global mean temperature. Under P2, note, the judgements under (a)-(f) should reflect the diversity of expert opinion in the domain. The exercises just referred to did not aim explicitly to do so. So too, recall that, under P2, judgements (a)-(e) are to be acknowledged to be risky to an unquantified degree. IPCC reports do often take a step towards such acknowledgement by qualifying their confidence in PDFs, e.g., giving them medium or high confidence, as a function of the quality of supporting evidence. This gives the misleading impression that we can quantify how confident we should be in the PDF, contrary to our arguments earlier in this section.

There is (e.g., Bamber and Aspinall [2013]) a second, less widely used, approach to subjective evaluation of PDF credibility in the climate context, namely using expert elicitation directly: is a PDF which has been produced by experts using all available evidence credible? We do not here spell out the steps of using this approach under P2, but note that a directly elicited PDF will require a clear defence of its characteristics as capturing as far as possible diversity of opinion among domain experts about the best probabilistic representation of evidential uncertainty. We further note that PDFs based on this approach must also come with the caveat that they are uncertain to an unquantified extent.

4. What to do when a PDF won't do

4.1 Formal treatments of imprecise probabilities

Imprecise probability theory provides us with alternatives to representing uncertainty with a (partial or full) PDF. Here, sets of PDFs can be used (see Bradley [2019] for an overview). Using sets of PDFs allows us to represent, without loss of information, a range of expert opinions about which PDF best captures future uncertainty about climate as well as a range of model generated distributions. Moreover, there are strategies for mitigating the unwieldy situation of having to deal with large numbers of PDFs.[2]

For example, competing studies of climate sensitivity sometimes reflect fundamental disagreement resulting from reliance on, e.g., different model structures and expert judgements. To avoid masking the uncertainty associated with such disagreement, Hall et al. [2007] use a set of PDFs from then available studies to represent uncertainty about climate

[2] Partial PDFs already depart from probability theory. The point being made in this section is that imprecise probability theory provides allows further, useful departures.

sensitivity. The imprecise probability distribution that results is summarised by an outer envelope on the cumulative probability distributions of available PDFs. For a given value of climate sensitivity, the cumulative probability distribution of a PDF gives the probability, according to that PDF, that climate sensitivity will be equal to, or lower than, the value.

Such approaches, however, still leave untouched the question whether available PDFs span the range of possible distributions which are compatible with our knowledge. Consider, for example, the ensembles of opportunity provided by state-of-the-art climate models. We have observed that they do not fully explore the space of possible evolutions of the climate system that are compatible with our knowledge and thus that we should not interpret their output as a real-world PDF, or even as capturing evidential uncertainty. However, for the same reason, these models are not, in their current form, suitable for generating sets of PDFs that might represent our uncertainty about quantities of interest, or even for producing lower and upper probability bounds. Insisting on imprecise probabilities here leads us to an option that is similar to P2 and retains the core problems with PDFs, though with more limited worries regarding the adequacy of exploration of extremes.

The lower computational cost of simpler climate models means that they are better suited to estimating a broad range of possibilities [Katzav and Parker, 2015]. We cannot, however, straightforwardly use the results of studies with such models to create appropriate PDFs or collections of PDFs for the reasons described in 3 above. For such models, the problems of determining the range of appropriate structures to be used in trying to represent uncertainty are exacerbated by the particularly limited grounding of models in physical theory [Katzav and Parker, 2015; Baumberger et al., 2018]. One could again choose something like option P2 and provide a range of low probabilities, or collections of such, for extremes produced by these studies. But doing this will reflect somewhat arbitrary decisions. Further, a better framing of our uncertainty in such cases seems to be that certain extremes may turn out to be possibilities that should be taken seriously in decision making. This framing is neutral about the probability of the extremes and thus, unlike an assignment of ranges of low probabilities, does not ignore the possibility that the extremes will turn out to be as serious as any other possibilities already acknowledged to be serious.

Thus, for example, simple models have suggested that Antarctica might contribute more than a meter of sea-level rise by the year 2100 (see, e.g., DeConto and Pollard [2016]) but there has been some discussion about how seriously to take these possibilities (see, e.g., Clerc et al. [2019] and Pattyn and Morlighem [2020]). It seems reasonable just to acknowledge that it is unclear whether these possibilities are serious rather than to assign them a range of low probabilities and not explicitly to acknowledge that they might turn out to be as serious as any others.

4.2 Formal possibilistic approaches

Possibilistic representations of uncertainty represent uncertainty with possibility distribution functions rather than PDFs. Possibilities come in degrees, like probabilities. But possibilities are not additive, unlike probabilities. If for example two events are fully possible, their disjunction is so too, no more and no less than the individual events. Quantitative possibility distributions assign to each state in a set of states a number from the real interval [0,1], where 0 stands for impossibility and 1 for full possibility. Further, at least one state is assigned the value 1. We can define the possibility measure for any set of states *A* on which a possibility distribution $\pi(x)$ is defined as $\Pi(A) = sup_{(x\in A)}\, \pi(x)$. $\Pi(A)$ gives us the degree to which *A* is possible (see Dubois and Prade [2015] and Halpern [2017] for more on possibility theory).

Presenting a range of model outputs without committing to how uncertain the output is, as was done, for example, in the 2002 United Kingdom climate projections [Hulme et al., 2002], differs from presenting a possibility distribution. When presenting a possibility as a full possibility, one is committed to taking it seriously in decision making [Betz, 2016]. When a possibility is not yet a full possibility, one is indicating that it is unclear whether it is to be taken seriously in this way. Roughly, when the understanding on which we base a claim that something is a possibility is more realistic, the closer the possibility is to being full.

An advantage of using quantitative possibilistic representations of uncertainty over imprecise probabilities is that possibilistic representations more felicitously represent cases in which, as with the possibility of extreme sea-level rise, it is not clear how serious the possibilities are.

For example, in estimating uncertainty about sea level rise by the end of the century, Le Cozannet et al. [2017] take the AR5 IPCC assessment that it is at least likely (66%) that, under RCP8.5, sea level rise will be between 0.52 to 0.98 meters by 2100 and transform it into the possibilistic assessment that it is fully possible that sea-level rise will be within this range. They assign less than full possibility to values lower than 0.52 and higher than 0.98 meters and take the full range of projections to be given by lower and upper bounds of available estimates, including the IPCC scenarios and other estimates of more extreme levels of sea-level rise. Notice that, in this way, Le Cozannet et al. assume that state-of-the art climate models, which were key to deriving the IPCC AR5's conclusion, contribute to estimating our uncertainty but avoid making the mistake of taking agreement or disagreement between such models to contribute to estimating the likelihood of the range of projections or the full range of possibilities.[3]

Qualitative possibilities are defined using partially or totally ordered ordinal scales [Dubois and Prade, 2015]. Representing uncertainty about different sources of evidence using an ordinal scale allows being neutral about how to compare the uncertainties (e.g., the outputs from different models or different studies can be represented as possibilities without deciding whether the outputs are equally good or not, or by preferring one but to an unquantified degree). Qualitative possibilities have not, as far as we know, been used in the climate context.

4.3 What to do when what is possible is unknown

The formal possibilistic assessments discussed above also have their limitations. Thus, for example, a key aspect of the evolution of our uncertainty about climate is that the space of (partial and full) possibilities itself evolves. While extreme levels of sea-level rise might not have been considered as partial possibilities in the past, they are so now. It is thus important to consider and develop formal and informal approaches to handling such situations in the context of climate projections. This includes informal approaches which guide us through articulating decision-relevant possibilities, while noting where such possibilities might not be

[3] Quantitative possibility measures can be interpreted as upper probabilities, which are tools of imprecise probability theory. This allows interpreting possibilistic representations using the tools of imprecise probability, though plausibly with a loss of information about uncertainty [Dubois and Prade 1993 and 2015].

known, and in reasoning about these possibilities.[4] Further, where the space of possibilities is only partially known, it is often important to invest in attempts to determine the bounds on the range of possibilities that are, for better or for worse, to be taken seriously (specifying non-discountable envelopes). This information could be particularly valuable for risk-averse decision makers [Hinkel et al., 2019].

The "storyline" approach complements the possibilistic exploration of extremes. Within the storyline approach, theoretical and expert knowledge is first used to build pictures of highly uncertain futures; for instance a scenario for greenhouse gas emissions might be combined with climate or earth system sensitivities which are outside the range simulated by today's state-of-the-art models. These pictures can then be filled in with details from high resolution, weather-model simulations. Informal tools for articulating serious possibilities can, in principle, be used to guide interpreting resulting scenarios in possibilistic terms (for more on storylines, see Risbey et al. [2002] and Shepherd et al. [2018]).

5. Conclusions

We have seen that PDFs can misrepresent uncertainty and that this might have negative consequences for decision-making. Further, while the reliability of a PDF in the case of weather can empirically be evaluated by repeated, quantitative testing against out-of-sample data, this is generally not possible in the climate context. Extremely well confirmed theory could in principle compensate here, by indicating what uncertainty to associate with climate projections, but theory of climate is not sufficiently developed to do so. In such circumstances, PDFs about future climate will be unreliable to an unquantified extent. It, accordingly, seems reasonable to go for what we called perspective P1, which is that PDFs should be used only when evaluated quantitatively or with extremely well confirmed theory, and to avoid using PDFs in representations of our uncertainty about future climate. We also, however, provided a second alternative, P2. On this alternative, PDFs are to be offered when these reflect a best attempt at capturing domain expert uncertainty and while acknowledging that they are uncertain to an unquantified extent.

The IPCC takes steps towards P2 in, for example, transforming model-output for global temperatures over the rest of the century into likely or very likely ranges of projections using expert judgement, a step to which we have referred above. The IPCC approach, however, does not explicitly acknowledge that resulting PDFs are uncertain to an unquantified extent. The IPCC approach also focuses on consensus PDFs while our proposal is that PDFs that better reflect the breadth of domain expert opinion be provided. More generally, we have argued that P2 is challenging to implement.

Alternative P1 does not leave us without actionable information or without more adequate means of representing uncertainty. Formal treatments of imprecise probabilities include methods for presenting multiple PDFs simultaneously, but although they clarify the disagreement between models and/or experts rather than seeking to condense it into a single projection, they are less than ideal for representing the full range of uncertainty, including cases where it is unclear what the full possibilities are.

Formal possibilistic approaches are available for wider use in climate science and take another step towards quantifying "deep uncertainty" by representing the range of partial

---

[4] See Betz [2016] and the references therein for a general discussion of informal approaches. See Heifetz et al. [2006] for an example of a formal system for representing an evolving possibility space in the context of economics.

and fully possible outcomes. Informal, non-probabilistic approaches to assessing uncertainty are also available. We note that the possibilities in the envelope of possibilities provided in the very likely range of IPCC temperature ranges [Lee and Marotzke, 2021] are all serious and thus should not be ignored in decision making. This claim, further, is not subject to the worries raised about the appropriate location of probability mass or about where to locate PDF extremes.

Possibilistic approaches would benefit from discussion of when to take extreme possibilities seriously and from development in the context of exploring extremes with, for example, the storyline approach. More broadly, use of non-probabilistic approaches to represent uncertainty would require more familiarity with these in the climate science community and a culture that makes explicit disagreement that is masked by consensus PDFs. Similarly, further work is needed to consider how non-probabilistic approaches are impacted when considering different spatial scales and different climatic variables.

Accurate representations of genuine levels of uncertainty about future climate outcomes are very important for decisions about mitigation and adaptation. We have argued that probability distributions of future climate change do not accurately represent genuine levels of uncertainty, that they can indeed be misleading. This suggests that other approaches such as those described above should be explored and implemented and that, where probabilistic representations are used, caution should be used and warnings provided.

References

Bamber, J. and Aspinall, W. (2013) “An expert judgement assessment of future sea level rise from the ice sheets.” *Nature Climate Change*, 3: 424–427 (2013). https://doi.org/10.1038/nclimate1778

Benjamin, D. J. (2019) “Errors in probabilistic reasoning and judgment bias.” In B. Douglas Bernheim, S. DellaVigna and D. Laibson (eds.) *Handbook of Behavioral Economics - Foundations and Applications 2*, pp. 69-186. Elsevier: London and Amsterdam.

Baumberger, C., Knutti, R., Hadorn, G. H. (2017) “Building confidence in climate model projections: an analysis of inferences from fit,” *WIREs Climate Change*, 8:e454. https://doi.org/10.1002/wcc.454

Betz, G.. (2016) Accounting for possibilities in decision making. In: Hansson S., Hirsch Hadorn G. (eds) *The Argumentative Turn in Policy Analysis. Logic, Argumentation & Reasoning* (Interdisciplinary Perspectives from the Humanities and Social Sciences), vol. 10: 135-170. Springer, Cham.

Borodina, A., Fischer, E. M., and Knutti, R. (2017) “Potential to constrain projections of hot temperature extremes,” *Journal of Climate*, *30*(24), 9949-9964. https://doi.org/10.1175/JCLI-D-16-0848.1

Bradley, S. (2019) "Imprecise probabilities." In C. Beisbart and N. J. Saam (eds.) *Computer Simulation Validation*, 525-540: Springer.

Clerc, F., Minchew, B. M., and Behn, M. D. (2019) “Marine Ice Cliff Instability Mitigated by Slow Removal of Ice Shelves”, *Geophysical Research Letters*, 46(21), 12108-12116. https://doi.org/10.1029/2019GL084183

DeConto, R. M. and Pollard, D. (2016) “Contribution of Antarctica to past and future sea-level rise”, *Nature*, 531: 591-597. https://doi.org/10.1038/nature17145

Dubois, D., Prade, H., and Sandri, S. (1993) “On possibility/probability transformations.” In R. Lowen and M. Roubens (eds.) *Fuzzy Logic: State of the Art*, 103-112. Kluwer Academic: Dordrecht.

Dubois, D., and Prade, H. (2015) “Possibility Theory and Its Applications: Where Do We Stand?” In Kacprzyk J., Pedrycz W. (eds) *Springer Handbook of Computational Intelligence*, 31-59. Springer Handbooks. Springer, Berlin, Heidelberg.

Fine, T. L. (1988) “Lower probability models for uncertainty and nondeterministic processes,” *Journal of Statistical Planning and Inference*, 20(3): 389-411. https://doi.org/10.1016/0378-3758(88)90099-7

Gleckler P. J., Taylor K. E., Doutriaux C. (2008) “Performance metrics for climate models,” *Journal of Geophysical Research*, 113, D06104. https://doi:10.1029/2007JD008972.

Hájek, A., "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2019/entries/probability-interpret/>

Hall, J., Fu, G., Lawry, J. (2007) “Imprecise probabilities of climate change: aggregation of fuzzy scenarios and model uncertainties,” *Climatic Change*, 81: 265-281. https://doi.org/10.1007/s10584-006-9175-6

Halpern, J. Y. (2017) *Reasoning about uncertainty*, 2nd ed. The MIT Press, Cambridge, MA. https://doi.org/10.7551/mitpress/10951.001.0001

Hausfather, Z., Drake, H. F., Abbott, T., & Schmidt, G. A. (2020) “Evaluating the performance of past climate model projections,” *Geophysical Research Letters*, *47*(1), e2019GL085378. https://doi.org/10.1029/2019GL085378

Heifetz, A., Meier, M. and Schipper, B. C. (2006) "Interactive unawareness," *Journal of economic theory*, 130.1: 78-94. https://doi.org/10.1016/j.jet.2005.02.007

Hinkel, J., Church, J. A., Gregory, J. M., Lambert, E., Le Cozannet, G., Lowe, J., et al. (2019) “Meeting user needs for sea level rise information: A decision analysis perspective,” *Earth's Future*, 7: 320–337. https://doi.org/10.1029/2018EF001071

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R., Hill, S. (2002) *Climate change scenarios for the United Kingdom: The UKCIP02 Scientific Report*, Tyndall Centre for Climate Change Research and School of Environmental Sciences, University of East Anglia, Norwich.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Katzav, J. (2013) “Severe testing of climate change hypotheses,” *Studies in History and Philosophy of Modern Physics*, 44(4), 433-441. https://doi.org/10.1016/j.shpsb.2013.09.003

Katzav, J. (2014) “The epistemology of climate models and some of its implications for climate science and the philosophy of science,” S*tudies in History and Philosophy of Modern Physics*, 46(B), 228-238. https://doi.org/10.1016/j.shpsb.2014.03.001

Katzav, J. and Parker, W. S. (2015) “The future of climate modeling,” *Climatic Change*, 132: 475-487. https://doi.org/10.1007/s10584-015-1435-x

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. A. (2010), “Challenges in Combining Projections from Multiple Climate Models”, *Journal of Climate*, 23(10): 2739–2758. https://doi:10.1175/2009JCLI3361.1

Le Cozannet, G., Manceau, J. C., Rohmer, J. (2017) “Bounding probabilistic sea-level projections within the framework of the possibility theory,” *Environmental Research Letters*, 12, 014012. https://doi.org//10.1088/1748-9326/aa902a

Lee, Y.-J. and Marotzke, J. (eds.) 2021: Chapter 4: Future global climate: scenario-based projections and near-term information. In IPCC 2021: AR6 Climate Change 2021: The Physical Science Basis.. Cambridge University Press.

Lovejoy, S. (2015) “A voyage through scales, a missing quadrillion and why the climate is *not* what you expect.” *Climate Dynamics*, 44, 3187–3210. https://doi.org/10.1007/s00382-014-2324-0

Lowe, J. A., Howard, T., Pardaens, A., Tinker, J., Holt, J., Wakelin, S., Milne, G., Leake, J., Wolf, J., Horsburgh, K., Reeder, T., Jenkins, G., Ridley, J., Dye, S., & Bradley, S. (2009). UK Climate Projections science report: Marine and coastal projections. Met Office Hadley Centre. http://ukclimateprojections.defra.gov.uk

Lowe, J. A., Bernie, D., Bett, P., Bricheno, L., Brown, S., Calvert, D., Clark, R., Eagle, K., Edwards, T., Fosser, G. and Fung, F. (2018) *UKCP18 science overview report*. Exeter, UK: Met Office Hadley Centre.

Mendlik, T., & Gobiet, A. (2016) “Selecting climate simulations for impact studies based on multivariate patterns of climate change,” *Climatic change*, *135*(3-4), 381-393. https://doi.org/10.1007/s10584-015-1582-0

Milinski, S., Maher, N., and Olonscheck, D. (2020) “How large does a large ensemble need to be?” *Earth System Dynamics*, 11, 885–901. https://doi.org/10.5194/esd-11-885-2020

Millner, A., R. Calel, D. Stainforth & G. MacKerron (2013) “Do probabilistic expert elicitations capture scientists’ uncertainty about climate change?” *Climatic Change*, 116(2), 427-436. https://doi.org/10.1007/s10584-012-0620-4

O’Hagan, A. (2019) “Expert knowledge elicitation: subjective but scientific,” *The American Statistician*, 73(sup1), 69-81. https://doi.org/10.1080/00031305.2018.1518265

Parker, W. S. (2010) “Whose probabilities? Predicting climate change with ensembles of models,” *Philosophy of Science*, 77(5), 985-997. https://doi.org/10.1086/656815

Parker, W. S. and Winsberg, E. (2018) “Values and evidence: how models make a difference,” *European Journal of Philosophy of Science*, 8, 125–142. https://doi.org/10.1007/s13194-017-0180-6

Pattyn, F. and Morlighem, M. (2020) “The uncertain future of the Antarctic ice sheet”, *Science*, 367: 1331-1335. https://doi.org/10.1126/science.aaz5487

Risbey, J., Lamb, P., Miller, R., Morgan, M., and Roe, G. (2002) “Exploring the structure of regional climate scenarios by combining synoptic and dynamic guidance and GCM output”, *Journal of Climate,* 15 (9): 1036-1050. https://doi.org/10.1175/1520-0442(2002)015<1036:ETSORC>2.0.CO;2

Risbey, J. and O'Kane, T. (2011) “Sources of knowledge and ignorance in climate research.” *Climatic Change*, 108(4): 755-773. https://doi.org/10.1007/s10584-011-0186-6

Risbey, J, Squire, D., Black, A., DelSole, T., Lepore, C., Matear, R., Monselesan, D., Moore, T., Richardson, D., Schepen, A., Tippett, M., and Tozer, C. (2021). “Standard assessments of climate forecast skill can be misleading,” *Nature Communications*, 12 (4346), 1-14. https://doi.org/10.1038/s41467-021-23771-z

Schmidt, G. A., & Sherwood, S. (2015) “A practical philosophy of complex climate modelling,” *European Journal for Philosophy of Science*, 5(2): 149-169. https://doi.org/10.1007/s13194-014-0102-9

Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., Molod, A., Neale, R. B., and Saha, S. (2017) “Practice and philosophy of climate model tuning across six US modeling centers,” *Geoscientific Model Development*, 10, 3207–3223, https://doi.org/10.5194/gmd-10-3207-2017

Shepherd, T. (2014). “Atmospheric circulation as a source of uncertainty in climate change projections,” *Nature Geoscience,* 7(10), 703-708. https://doi.org/10.1038/ngeo2253

Shepherd, T., Boyd, E., Calel, R., Chapman, S., Dessai, S., Dima-West, I., Fowler, H., James, R., Maraun, D., Martius, O., Senior, C., Sobel, A., Stainforth, D., Tett, S., Trenberth, K., van den Hurk, B., Watkins, N., Wilby, R., and Zenghelis, D. (2018) “Storylines: an alternative approach to representing uncertainty in physical aspects of climate change," *Climatic Change*, 151: 555-571. https://doi.org/10.1007/s10584-018-2317-9

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020) "An assessment of Earth's climate sensitivity using multiple lines of evidence," *Reviews of Geophysics*, 58, e2019RG000678. https://doi.org/10.1029/2019RG000678

Stainforth, D. A., Allen, M. R., Tredger, E. R., and Smith, L. A. (2007) "Confidence, Uncertainty and Decision-Support Relevance in Climate Predictions," *Philosophical Transactions of the Royal Society A*, 365(1857): 2145–2161. https://doi.org/10.1098/rsta.2007.2074

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012) "An overview of CMIP5 and the experiment design," *Bulletin of the American Meteorological Society*, 93(4): 485-498. https://doi.org/10.1175/BAMS-D-11-00094.1

Tebaldi, C., & Knutti, R. (2007) "The use of the multi-model ensemble in probabilistic climate projections," *Philosophical Transactions of the Royal Society A: mathematical, physical and engineering sciences*, *365*(1857), 2053-2075. https://doi.org/10.1098/rsta.2007.2076

Thompson, E., Frigg, R., & Helgeson, C. (2016) "Expert judgment for climate change adaptation," *Philosophy of Science*, *83*(5), 1110-1121. https://doi.org/10.1086/687942

Thompson, E. L. and Smith, L.A. (2019). Escape from Model-land. *Economics*, 13(2019-40), 1-15. http://dx.doi.org/10.5018/economics-ejournal.ja.2019-40

Weisheimer A. and Palmer T. N. (2014) "On the reliability of seasonal climate forecasts," *Journal of the Royal Society Interface*, 11: 20131162. http://dx.doi.org/10.1098/rsif.2013.1162