Title:

Assessing the quality of state-of-the-art regional climate information: the case of the UK Climate Projections 2018

Authors:

Marina Baldissera Pacchetti
University of Leeds, School of Earth and Environment and ESRC Centre for Climate Change Economics and Policy, UK
m.baldisserapacchetti@leeds.ac.uk
https://orcid.org/0000-0002-5867-6893

Suraje Dessai
University of Leeds, School of Earth and Environment and ESRC Centre for Climate Change Economics and Policy, UK
s.dessai@leeds.ac.uk
https://orcid.org/0000-0002-7879-9364

David A. Stainforth
London School of Economics, Grantham Research Institute on Climate Change and the Environment, and ESRC Centre for Climate Change Economics and Policy, UK
Department of Physics, University of Warwick, UK
d.a.stainforth@lse.ac.uk

https://orcid.org/0000-0001-6476-733X

Seamus Bradley

University of Leeds, School of Philosophy, Religion and History of Science and ESRC Centre for Climate Change Economics and Policy, UK

s.c.bradley@leeds.ac.uk

https://orcid.org/0000-0001-9663-7919

Abstract:

In this paper, we assess the quality of state-of-the-art regional climate information intended to support climate adaptation decision-making. We use the UK Climate Projections 2018 as an example of such information. Their probabilistic, global and regional land projections exemplify some of the key methodologies that are at the forefront of constructing regional climate information for decision support in adapting to a changing climate. We assess the quality of the evidence and the methodology used to support their statements about future regional climate along six quality dimensions: transparency; theory; independence, number and comprehensiveness of evidence; and historical empirical adequacy. The assessment produced two major insights. First, a major issue that taints the quality of UKCP18 is the lack of transparency, which is particularly problematic since the information is directed towards non-expert users who would need to develop technical skills to evaluate the quality and epistemic reliability of this information. Second, the probabilistic projections are of lower quality than the global projections because the former lack both transparency and a theory underpinning the method used to produce quantified uncertainty estimates about future climate. The assessment also shows how different dimensions are satisfied depending on the evidence used, the methodology chosen to analyze the evidence, and the type of statements that are constructed in the different strands of UKCP18. This research highlights the importance of knowledge quality assessment of regional climate information that intends to support climate change adaptation decisions.

Keywords:

Knowledge Quality Assessment, Regional Climate Information, Climate Models, Uncertainty, Adaptation

1. Introduction

Adapting to a changing climate is an increasingly urgent necessity. Anthropogenic greenhouse gas emissions have already caused about 1 ºC of global warming, and even for the most optimistic mitigation scenarios, we are likely committed to 1.5 ºC warming with respect to the pre-industrial period by 2030-2050 (IPCC 2018). Informing the preparations needed to manage the risks, limit the damages and take advantage of the opportunities that arise in light of this changing climate is a grand challenge of climate change science (Moss et al. 2013).

There is an increasing interest in understanding how to address information needs for climate change adaptation decisions. For example, Knutti (2019) argues that despite the improvements in scientific understanding of climate and climate change, we need "more useful knowledge oriented toward solutions" (p. 22). One of the ways in which physical climate science can address this is by providing "more local climate information" (p. 22).

Decadal and multi-decadal regional climate information is increasingly important for making adaptation decisions and varies in temporal and spatial resolution. However, information about future changes in regional climate also comes with high degrees of uncertainty – an important element of the information given the high stakes of climate change adaptation decisions. This information is usually derived from Global Climate Models (GCMs) and Earth System Models (ESMs). State of the art modeling techniques are used to explore uncertainties and model sensitivities with ensemble experiments, dynamical downscaling with regional climate models

(RCMs), statistical downscaling, and the use of high-resolution convection permitting models (CPMs).

However, model-based information is difficult to interpret: the non-stationarity of the system and the time scales of forward looking model simulations imply that these simulations cannot be verified or confirmed (Stainforth et al. 2007b), the nature and scope of ensemble experiments is not clearly defined (Pirtle et al. 2010; Parker 2011; Masson and Knutti 2011; Jebeile and Crucifix 2020), excessive focus on uncertainty quantification risks being misleading (Parker and Risbey 2015), and it is not always clear that there is an escape from "model land", i.e., statements from models about models, rather than statements from models about the world (Thompson and Smith 2019).

So a legitimate question that can be asked is whether information about future climate derived from ESMs and other types of evidence does meet the quality standards that are needed to make decisions about how to adapt to a changing climate. Just because the information is provided, doesn't mean it is adequate for the purpose of informing climate change adaptation decisions. For example, Fiedler et al. (2021) argue that rules need to be developed to evaluate the reliability of climate information for decision support in the private sector.

To assess the quality of regional climate information for decision making, we apply a slight modification of the quality assessment framework of Baldissera Pacchetti et al. (2021). In that paper quality is specified along five dimensions for statements or estimates about future climate: transparency, theory, diversity and completeness, and adequacy for purpose. We slightly modify these dimensions in two ways. First, we break down diversity and completeness into: number, independence and comprehensiveness to more clearly capture the way the typology of evidence and its analysis bear on statements about future climate. Second, we change “adequacy for purpose” to “historical empirical adequacy” to more clearly specify this dimension and differentiate it from more general notions of adequacy for purpose (e.g., Parker 2020). These dimensions are designed to assess the epistemic reliability of statements about future climate, which requires that the information and related probabilities suitably represent the likelihood of different realizations of future climate, and that there is an explanation of why this is the case.

The aims of this paper are twofold. First, to assess the quality of state-of-the-art information about future regional climate intended to inform adaptation decisions using the UK Climate Projections 2018 (UKCP18) as a case study. We consider what is needed to achieve higher quality to inform future efforts in constructing regional climate information. Second, this study serves as an empirical test for the quality framework itself.

We start by describing the modified framework in Section 2. Here, we describe “quality” in the context of providing information for decision support. We specify the target of the framework in terms of the elements of information about future regional climate which need to be taken into consideration for a meaningful assessment. In Section 3, we motivate the choice of UKCP18 as an exemplar of state-of-the-art regional climate information and assess the quality of three products of the land projections according to the framework of Baldissera Pacchetti et al. (2021). In Section 4 we discuss the findings of the assessment. We conclude with future research directions in Section 5.

2. The framework

The framework introduced by Baldissera Pacchetti et al. (2021) specifies what is meant by quality in the context of informing climate change adaptation decisions. In particular, this framework focuses on the epistemic requirements of a concept of quality in this context. These epistemic requirements can provide guidance on what it means for information to be *credible* enough to be decision-relevant. Credibility refers to the scientific adequacy of the technical details and arguments used as evidence for the information (Cash et al. 2003).

For information to be of high quality, it needs to be epistemically reliable, i.e., the information about future climate and related probabilities need to suitably represent the likelihood of different realizations of future climate, *and there needs to be an explanation of why this is the case*. This understanding of reliability becomes important when statistical-empirical evaluations of reliability are not available to scientists, as is the case for long term climate predictions and projections (see, e.g., Winsberg 2006, Stainforth et al. 2007a, Stainforth et al. 2007b, Baldissera Pacchetti 2020). Epistemic reliability is also important when connecting model-based statements about models to model-based statements about the real world (see Thompson and Smith 2019).

The target of the framework is information in the form of "statements or estimates about future regional climate",[1] on decadal and longer time scales, that are produced by scientific research (Baldissera Pacchetti et al. 2021, p. 477). Beyond the statements themselves there are two further

[1] We will use estimate or statement as appropriate to the context, but our discussion is relevant for both.

aspects that need to be taken into consideration: the evidence underpinning the statements and the methodology used to analyze this evidence.

Baldissera Pacchetti et al. (2021) identify five dimensions along which quality can be assessed: transparency, theory, diversity, completeness and adequacy for purpose. Transparency requires that both the evidence and methodology be accessible enough for the other quality dimensions to be assessed, even by non-experts. Theory refers to the strength of the theoretical foundations for the statement about future climate; it covers both physical processes and methodological approaches to the data. This dimension is particularly important for giving epistemic reliability and is recognized to some extent in recent process-based model evaluations (Daron et al. 2019; Jack et al. 2021). Diversity and completeness track different but related aspects of how evidence is sourced and combined. For clarity, these two dimensions have been further divided into three sub-dimensions: Independence, Number and Comprehensiveness (see Table 1). Independence tracks the extent to which different types of evidence can be considered independent. Types of evidence can, for example, be ESM or GCM models that share model genealogy and any derivative thereof (e.g., emulators), theoretical process-based understanding, expert judgment, observations, paleoclimatic data (see also Fig. 1 in Baldissera Pacchetti et al. 2021). Independence can be assessed by evaluating the provenance of the evidence such as model genealogy and overlapping modeling assumptions, training and background of scientists chosen for expert elicitation, geographical location of research activity, etc. Number tracks how many of these different types are taken into account. Comprehensiveness tracks whether each type of evidence is exhaustively assessed, i.e. whether model space is sufficiently explored, whether enough of the relevant experts are consulted, or whether all plausible physical theories are taken into consideration. These three sub-dimensions contribute to an exhaustive uncertainty assessment – an important component of policy-relevant statements about future climate.

Adequacy for purpose, in general, is invoked to highlight that model evaluation should take account of the purpose for which a model is being used (Parker 2020). In the present case, the purpose of statements about future climate is to inform decision making, and to achieve this requires epistemic reliability. To more clearly specify what can be assessed as adequate in the context of this purpose, we call this dimension *historical empirical adequacy.* This dimension refers to the empirical adequacy of the model evaluation for the stated purpose of the output (e.g., has model output been compared with historical observations for each variable of interest at the relevant spatial and temporal scales, etc.).

Score	Qualifier	Transparency	Theory	Diversity and Completeness			Historical Empirical Adequacy
				Independence	**Number**	**Comprehensiveness**	
0	Not satisfied	No access	No theoretical support that warrants X. Or Can't assess.	Only one type of evidence is taken into consideration to justify X. Or Can't assess	No (scientific) evidence is taken into consideration. Or Can't assess	No exploration of uncertainty within individual lines of evidence. Or Can't assess	No empirical tests (e.g. hincasts) for X. Or Can't assess
1	Minimally satisfied	Evidence and Methodology are mentioned but not well explained and not appropriately traceable.	Weak theoretical support that warrants X. (theoretical underpinning is weak, and doesn't justify the precision of X)	There is considerable overlap among the evidence.	Few of the available lines of evidence are taken into account.	Minimal exploration of uncertainty within individual lines of evidence.	Empirical tests are performed but only of few components relevant to X.
2	Somewhat satisfied	Evidence and methodology are somewhat accessible and traceable, but there are gaps.	Medium theoretical support that warrants X.	The evidence overlaps somewhat.	Multiple, but not most available lines of evidence are taken	Partial exploration of uncertainty within individual lines of evidence.	Empirical tests are performed but not for all

					into account.		components[2] relevant to X.
3	Generally satisfied	Evidence and methodology are well-explained, and all evidence is traceable.	Strong theoretical support that warrants X.	There is little overlap among sources of evidence.	Most available lines of evidence are taken into account.	Sufficient exploration of uncertainty within individual lines of evidence.	Extensive empirical tests are performed for all components relevant to X.
4	Satisfied	Evidence and methodology are well-explained, and all evidence is immediately available.	Theory unequivocally justifies X.	Completely independent types of evidence are taken into account.	All possible lines of evidence are taken into account.	Comprehensi ve exploration of uncertainty within individual lines of evidence.	All possible empirical tests for all components relevant to X.

Table 1 Qualitative descriptors for each quality dimension across a quantitative scale (0-4).

Table 1 provides qualitative descriptors for each quality dimension across a quantitative scale, and how various dimensions can be satisfied. These dimensions are not to be considered “necessary and sufficient conditions” for quality, and there is no absolute scale along which they can be assessed. The last row represents an in practice unattainable ideal, that can nevertheless provide guidance on how to achieve high quality information. In practice, the degree to which each dimension should or can be satisfied is influenced by the kind of statement under consideration and also the relation of the dimensions to one another (Baldissera Pacchetti et al. 2021, p. 488).

The order in which the above dimensions are presented is not prescriptive but highlights the relation between the dimensions. Transparency is assessed first because it provides an explanation for why other dimensions may not be satisfied if there is no access to relevant

[2] Components: model output for variable(s) of interest at the relevant spatial and temporal scale.

evidence and methodology for the estimate or statement under assessment. Theory follows transparency because the theoretical support for an estimate or statement can guide the extent to which diversity and completeness need to be satisfied: the stronger and more established the theoretical support, the less important diversity and completeness are for epistemic reliability. Finally, historical empirical adequacy is a minimal empirical requirement for epistemic reliability.

3. The assessment

The UKCP18 projections exemplify key characteristics of state-of-the-art information about future regional climate. Here we assess to what extent different strands of the UKCP18 land projections (Murphy et al. 2018) satisfy the quality dimensions of the framework. The probabilistic projections combine multi-model-ensembles (MME) and perturbed-physics-ensembles (PPE) to provide a probabilistic estimate of the uncertainties tied to future changes in regional climate. The global projections provide model-derived trajectories for future climate which aim to sample a broad range of possible future responses to anthropogenic forcing (Murphy et al. 2018, p. 38). The regional projections include dynamical downscaling using a PPE of regional climate models (RCM).

We apply the quality assessment framework to these three strands of UKCP18 and assess how they satisfy the dimensions of the quality framework. When appropriate, we show whether quality varies depending on the variable of interest within a particular strand or across strands. For example, the theory dimension highlights that quality is better satisfied for estimates about variables that depend on thermodynamic principles (such as global average temperature) than fluid dynamical theory (such as regional precipitation) (see, e.g., Risbey and O'Kane 2011) independently of the strand under assessment. Table 2 provides a summary of the products of the UKCP18 land projections.

	Probabilistic projections	**Global projections**	**Regional Projections**
Description	Probabilistic changes in future climate based on an assessment of model uncertainties	A set of 28 climate futures with detailed data on how it may evolve in the 21st century: • 15 x Hadley Centre Model variants HadGEM-GC3.05 • 13 x other climate models (CMIP5-13)	A set of 12 high-resolution climate futures over Europe downscaled from the global projections (PPE-15) using Hadley Centre model HadREM-GARA11M
Period	1961-2100	1900-2100	1981-2080 for 12km 1981-200, 2021-2040, 2061-2080 for 2.2km
Temporal resolution	Monthly Seasonal Annual	Daily Monthly Seasonal Annual	Subdaily for 2.2km Daily Monthly Seasonal Annual
Spatial resolution	25km	60km	12km 2.2km
Geographical extent	UK & regions	UK & regions Global	UK & regions Europe for 12km
Emission scenarios	RCP2.6 RCP4.5 RCP6.0 RCP8.5 SRES A1B	RCP8.5	RCP8.5
Why should you use it?	Explores emissions scenario uncertainty Explores uncertainty in key processes in climate models Helps characterize future extremes in risk assessment	Long time series Spatially and temporally coherent Direct access to "raw" climate model data	Enhanced spatial detail Spatially and temporally coherent Improved extremes

		Met Office Hadley Centre global climate model HadGEM3-GC3.05	Direct access to "raw" climate model data

Table 2. Summary of UKCP18 Land Projections (adapted from Fung et al. 2018, pp.3-4)

3.1 Probabilistic Projections

The probabilistic projections provide probabilistic estimates for potential future climate over the UK, based on an assessment of model uncertainties (Murphy et al. 2018).

Transparency

The probabilities can be interpreted as an outcome of the methodology used. The authors of UKCP18 say that "the available models are sufficiently skillful that the conditional probabilistic projections…provide useful advice about known uncertainties in future changes" (Murphy et al. 2018, p. 10) but recognize that "systematic errors represent an important but unavoidable caveat" (Murphy et al. 2018, p. 10). Furthermore, they warn the user that the probabilities do not reflect the confidence the scientists have in the strength of the evidence (see, e.g., Murphy et al. 2018, p. 9). This implies that the probabilities do not provide a measure of what can be concluded from the evidence.

These statements do not clarify how to interpret the usefulness of the information provided. If the uncertainty ranges do not represent the possible ranges of future climate but rather are conditional on the particular methodology and the evidence used, what are the consequences for the statements about future climate? A non-expert user would probably not be able to use this information to assess the consequences for the epistemic reliability of the probabilistic estimates and therefore for the suitability of the information for their particular purpose.

The decision-relevance of the information and the expertise required by a user to assess the epistemic reliability of the uncertainty estimates are not clarified by the additional available documents. For example, consider the following:

"We have designed the probabilistic projections to provide the primary tool for assessments of the ranges of uncertainties in UKCP18. However, they may not capture all possible future outcomes." (Fung et al. 2018, p. 3)

"The future probabilistic projections in UKCP18 are an update of those produced for UKCP09. You should interpret the probabilities as being an indication of how much the evidence from models and observations taken together in our methodology support a particular future climate outcome. […] The relative probabilities indicate how strongly the evidence from models and observations, taken together in our methodology, support alternative future climate outcomes." (Ibid.)

These statements show that the evaluation of the merits of a complex methodology is left to the user to decipher. It is unclear how a user who is not an expert in uncertainty assessments could assess the extent to which these estimates are suitable for their purposes. So, while the availability of multiple reports and guidance notes would suggest that the probabilistic projections satisfy the transparency dimension, the opacity of the method to derive the projections and the lack of explanation of how this affects the statements about future climate indicates that the probabilistic projections only *minimally satisfy* this dimension (score: 1). In order to score higher along this dimension, it should be clearly stated what it means for the uncertainty ranges to be conditional on the evidence and methodology, and what the consequences of this conditionality are. For example, it could be specified how much wider the uncertainty range could be, and what kind of information the probabilistic estimates can provide – do they represent the degree of belief UKCP18 scientists have regarding future regional climate?

Theory

Theoretical understanding is an important component of climate information for adaptation, and models do not directly encapsulate all theoretical knowledge (Baldissera Pacchetti et al. 2021). In order to show how epistemically reliable the results are, model output should be assessed based on the scientists' theoretical understanding of climatic processes and the theoretical justification for how the model output is processed. The theory dimension of the framework does not only address the process understanding of the underlying mechanisms responsible for observed and future climate, but also the use of methodology. Here we focus on methodology.

Murphy et al. (2018) use the Bayesian framework of Goldstein and Rougier (2004) to develop probabilities. The probabilistic projections are mainly constructed by developing three PPEs. Two of these are updated with observational constraints and combined with an MME obtained from CMIP5 "to achieve a combined sampling of parametric and structural uncertainties in physical and carbon cycle responses" (Murphy et al. 2018, p. 13). The model output is then further downscaled with an RCM PPE to produce the projections at the 25km resolution. There are several issues with this methodology.

While Murphy et al. (2018) state that the probabilities do not reflect their confidence in the evidence, the probabilities are presented as some kind of knowledge claim about future climate. The main issue here is that probabilities cannot be interpreted as a measure of likely futures – not even subjective probabilities as intended by the original methodology introduced by Goldstein and Rougier (2004)–unless the subjective nature of this approach is made explicit and discussed in more detail. These probabilities are a quantified measure resulting from the methodology and the modelling choices, but it is unclear whether they are a measure of uncertainty about future climate. We further substantiate this claim below.

Murphy et al. (2018) do not usefully discuss how UKCP18 addresses the issues raised in Frigg et al. (2015), who argue that the use of the discrepancy term to generate decision-relevant probabilities is problematic. The use of the discrepancy term rests on the informativeness assumption, i.e., the assumption that the distance between the model and the truth is small (Frigg et al. 2015, p. 3993).

Murphy et al. (2018) assume that the MME from CMIP5 can be an adequate proxy to estimate this distance, but CMIP5 output cannot be considered a representative sample of the real world and thus a good basis for assessing structural model uncertainty. This assumption is flawed because of shared assumptions and shared biases of models (see Masson and Knutti 2011; Knutti et al. 2013; and the discussion in Baldissera Pacchetti et al. 2021, p. 481).

While these criticisms are acknowledged in UKCP18, it is not explained how UKCP18 overcomes the consequences for generating decision-relevant knowledge so the concerns over the informativeness of the discrepancy term identified by Frigg et al. in UKCP09 persist in UKCP18. Probabilistic estimates would be better justified if supplemented with physical interpretation of the model output. As we and others have argued elsewhere (Stainforth et al. 2007a; Frigg et al. 2015; Thompson et al. 2016; Baldissera Pacchetti et al. 2021) extrapolatory inferences can be unreliable for complex, nonlinear systems like the climate system, and certain methodological assumptions used to produce probabilistic estimates about future regional climate do not warrant the claims of decision-relevance for the information obtained from these projections. Further, these estimates cannot be considered to represent subjective credences of a group of experts, since the authors of the technical report themselves state that “the probabilistic format should not be misinterpreted as an indication of high confidence in the weight of evidence behind specific outcomes” (Murphy et al. 2018, p. 9). The probabilistic projections therefore *do not satisfy* (score 0) the theory dimension. To improve theory with respect to the methodology, the subjective nature of these probabilities should be fully embraced, the justification for the informativeness assumption and its limitations should be described, and alternative methodologies to aggregate model output should be taken into consideration (e.g. Stainforth et al. 2007b).

Diversity and Completeness

Diversity and completeness assess some key characteristics of the evidence and how the evidence is analyzed. These dimensions are subdivided into Independence, Number and Comprehensiveness, which respectively assess the shared assumptions and origin, the number of different types of evidence and the extent to which individual types of evidence are explored.

The main lines of evidence used are an MME, three PPEs (the output of which is augmented with a statistical emulator), and observational data. To assess the diversity of this evidence, we discuss the extent to which these sources of evidence are different from one another, and, relatedly, whether they share substantive assumptions. In addition, expert knowledge is used to estimate the ranges of the parameters of the PPEs (Murphy et al. 2018, p. 13). However, the process for extracting the knowledge and the uncertainty implications for the probabilistic projections are unclear. The UKCP18 science reports (Murphy et al. 2018; Lowe et al. 2018) do not reveal any other sources of evidence for the probabilistic projections. The lack of a thorough description of the use of expert judgment to select the parameter ranges is problematic because the methodology used to process the PPEs was designed as an approach for quantifying expert knowledge (Goldstein and Rougier 2004). It is unclear however whether Murphy et al. (2018) intend their methodology to represent expert judgement (or expert knowledge). Besides, it has been argued that probabilistic expert elicitation can be ambiguous and can underestimate the uncertainty associated with the knowledge claims of groups of scientists (Millner et al. 2013). The consequences of such issues are impossible to assess because the expert judgement aspect of the approach is not described and indeed is undermined by various caveats (see above and

Murphy et al. 2018, p. 9). We cannot therefore assess the role expert knowledge plays as a source of evidence, so the discussion below focuses on model-based and observational evidence.

Independence is *somewhat satisfied* (score 2) with respect to model-based and observational evidence. We consider the MME and PPEs to be one type of evidence. In principle, these ensembles explore different sources of uncertainty: the MME explores structural uncertainty, whereas the PPE explores parameter uncertainty. Nevertheless, there is considerable overlap in the model structure and, consequently, shared biases in model output (Masson and Knutti 2011; Knutti et al. 2013). However, we can consider observations to be a different type of evidence. Take the HadCRUT3 dataset (Brohan et al. 2006) used for temperature as an example. This dataset is evaluated with reanalysis data but the overlap in model-based assumptions is not considerable (Parker 2016). Number is *minimally satisfied* (score 1) as few types of evidence are taken into account. Comprehensiveness is *somewhat satisfied* (score 2) with respect to model-based and observational evidence: structural model uncertainties are explored with a large MME by today’s standards and the uncertainties regarding the choice of parameters within one of the models is also on the large side by today’s standards although climate*prediction*.net demonstrated that a wider range of behavior can be found with much bigger ensembles (Stainforth et al. 2005).

Since the probabilistic projections aim to provide an estimate of uncertainty, there is one more way in which comprehensiveness should be assessed. Singh and AchutaRao (2020) show that observational uncertainty can affect estimates of future change, as the assessment of model performance varies depending on the observational dataset used. This uncertainty may be minimal for datasets of variables that have an extensive record in space and time and bias may be easily removed for variables that are well understood–such as temperature. However, this uncertainty may become severe for other variables of interest and can change depending on the metric used (Kennedy-Asser et al. 2021), and this difficulty should be explicitly acknowledged to provide epistemically reliable information.

In order to improve quality along these dimensions, expert elicitation should be thoroughly documented, a wider range of models coming from different modeling centers should be taken into account, and parametric uncertainty should be systematically explored across different models. Reanalysis data could also be taken from different centers as European and global reanalysis datasets are produced by several international research centers. This diversity could help control for some of the idiosyncrasies in modeling assumptions and data processing methodologies that are tied to each research centre.

Historical Empirical Adequacy

Historical empirical adequacy assesses whether statements about future regional climate intended for climate change adaptation decisions have been subjected to adequate empirical tests. Empirical adequacy for the variables for which probabilistic estimates are provided is not itself an indicator that the probabilistic estimates will be epistemically reliable, but if they are not empirically adequate it is a strong indicator that they won't be epistemically reliable. In this sense, empirical adequacy for the purpose of evaluating model behavior for variables of interest is a minimal requirement for quality. The importance of empirical adequacy for evaluating models has been stressed recently by Nissan et al. (2020). The following analysis is based only on the information that can be accessed.

The output of the probabilistic projections is assessed and updated mostly by studying anomalies in key variables. For example, Murphy et al. (2018, Fig. 2.4a, p. 20 and Fig. 2.5, p.25) assess temperature changes with respect to a chosen baseline period. This evaluation of empirical adequacy of a model or a group of models does not satisfy historical empirical adequacy. While anomalies may be useful for supporting a strong inference about the need for mitigation, it does not adequately support epistemically reliable estimates about future climate for adaptation. We provide a motivation for this claim below.

Empirical adequacy with respect to an anomaly is only a measure relative to a chosen baseline, makes strong implicit assumptions about the linearity of the climate system, and can be achieved without a good representation of some of the details of the earth system. Take the time series data of GMST for the 1900-2000 period from CMIP5 alongside a time series of observations shown in Frigg et al. (2015, p. 3994). While the warming signal appears consistent among model output, there is considerable difference across models for the absolute value of GMST. As Frigg et al. (2015, p. 3994) note, these differences–albeit only of a few degrees Celsius–are an indication that different models represent the earth system differently: the location of sea-ice, vegetation, etc., varies across models, and so do associated feedbacks. While this may be of less significance for evaluating the historical empirical adequacy of a *global signal* of climate change and related uncertainties, estimating *how much temperature will change locally* needs to rely on an adequate representation of the relevant earth system components, and associated processes and feedbacks–which is not captured by the empirical adequacy of anomalies.

This issue is particularly relevant when information is downscaled: heterogeneities across models in the representation of physical features of the earth system and associated processes and feedbacks may not matter when model output is averaged globally, but they will be of crucial importance when evaluating model performance at regional scales (Ekström et al. 2015). Because of the importance of evaluating historical empirical adequacy for the purpose of informing decision making in terms of absolute values of the relevant variables, historical empirical adequacy is *not satisfied* for the probabilistic projections (score 0). To improve along this dimension, model performance should be evaluated (and shown to be evaluated) for absolute values of the variables provided.

3.2 Global Projections

The focus of the global projections is on estimates and statements about future climate derived directly from individual CMIP5 and HadGEM-GC3.05 simulations rather than processed ensemble output. This also shifts the focus of the quality assessment. These projections aim to show "how the 21st century climate may evolve under the highest emission scenario RCP8.5" (Lowe et al. 2018, p. 1). The purpose of these projections is to provide "a multi-variable dataset for impacts analysis … and [to support the] development of storylines relating to future climate variability and extremes on a broad range of timescales" (Murphy et al. 2018, p. 35). Further details about the global projections can be found in Table 2.

Transparency

The global projections provide information on most of the sources of evidence and describe their methodology, but there are components of the evidence and how the evidence analyzed that are not accessible or traceable. Again, the user is left to assess certain key features of the quality of the projections with little support from the UKCP18 documents or user interface.[3]

There are various instances where this occurs. For example, as we discuss below, the user is left to assess which models perform best and what this implies for the epistemic reliability of the information. Moreover, the UKCP18 user interface does not aid in the evaluation of the performance of models against observations. Take the time series data for precipitation from the global projections (Fig. 1). When producing these images through the user interface, one can highlight up to 5 members of the ensemble, but one cannot distinguish between PPE and CMIP5 members. Furthermore, one cannot compare the model output with observations through the user interface. Unless the user has the skills to download the relevant data and process it themselves, they cannot easily assess the historical empirical adequacy dimension.

[3] The user interface can be found here: https://ukclimateprojections-ui.metoffice.gov.uk/ui/home

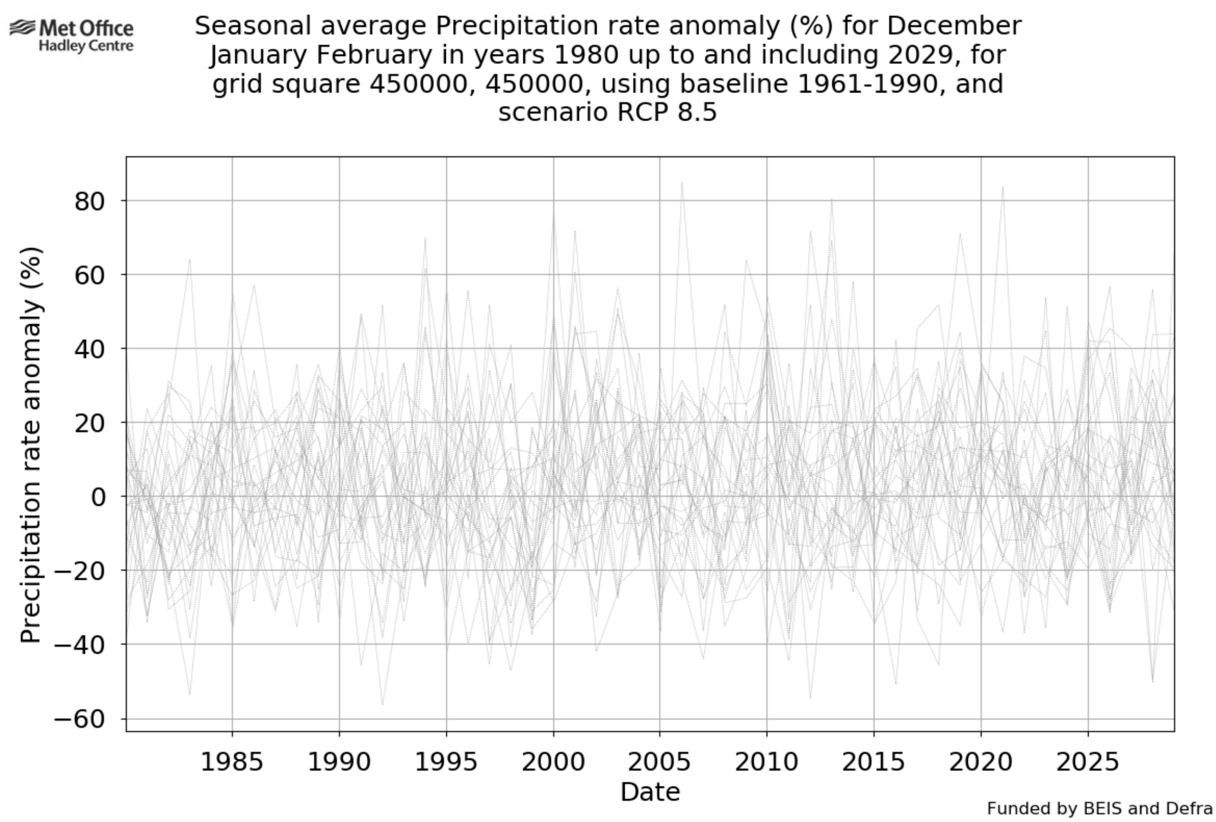


Figure 1. Global Seasonal Projections (30-year average) of precipitation rate anomaly for a 60km grid point over Leeds in the period 1980-2029. The projections are derived from 15 variants of HadGEM-GC3.05 and 13 members of CMIP5. Obtained from https://ukclimateprojections-ui.metoffice.gov.uk/ui/home in January 2021.

Furthermore, while most of the data sources are cited, it is not always clear what kind of data sets are used at various stages of the projection development process. For example, Murphy et al. (2018) cite the paper from which they borrow the methodology for model evaluations using 5-day simulations as the source of their data, but that paper only vaguely references the data set used (Williams et al. 2013, p. 3259). Another example of lack of transparency in the model development process is the use of expert elicitation in the construction of the PPE. Murphy et al. (2018) do not specify who the experts are and how they were chosen.

These considerations indicate that the global projections *somewhat satisfy* the transparency dimension (score: 2). The raw data can be downloaded from the interface, but the user would need to have high numerical literacy and programming skills to fully trace the model output. To improve transparency, the origin of the output of the global projections and the data sources used for the model verification should be fully traceable through the user interface and, ideally, thoroughly described in the supporting documents.

Theory

The description of the theoretical underpinning of how global atmospheric circulation patterns can affect UK weather is discussed at various stages in relation to the global projections (Murphy et al. 2018). For example, theoretical understanding of key processes is taken into consideration when choosing which parameters to perturb in the PPE and when choosing what synoptic system metrics to use to assess the performance of the simulations. However, the use of theoretical understanding is not explored in much depth in the science report.

The overview report of the scientific output (Lowe et al. 2018, p. 35) provides some further insight into how this theoretical understanding can be used. For instance, theory about large scale circulation patterns and their effect on local weather is combined with model output to provide statements about possible future climate over the UK. While this use of theoretical insight contributes to satisfying the theory dimension of the quality framework, the overview report exemplifies the use of theory only for pressure; there is no discussion of how it affects temperature or other variables. These considerations suggest that the global projections do *somewhat satisfy* the theory quality dimension (score 2). To improve quality along this dimension, there should be better integration between theoretical evaluation of the physical processes represented by models, and how it bears on the epistemic reliability of model output for individual variables.

Diversity and Completeness

There are several different sources of evidence for the global projections: MME, PPE, expert elicitation in building the PPEs, reanalysis data and observations (Murphy et al. 2018). As we have discussed in the evaluation of the probabilistic projections, MME and PPE count as one type of evidence.

Model output is derived from both a PPE and an MME. The MME output is similar to the one used for the probabilistic projections, but the PPE is constructed and forced differently (see Murphy et al. 2018, Section 3). Model output here is assessed as a source of evidence as it is used at various stages of the filtering stages to satisfy the principles of “plausibility and diversity” that drive the projection development process (Murphy et al. 2018, p. 37).

Expert elicitation follows Sexton et al. (2019), which is itself partly based on the Sheffield Elicitation Framework (SHELF) method of Oakley and O’Hagan (2010). Expert elicitation is used to set up the parameter space for the PPE. The parameters and the respective ranges are elicited from experts following the protocol suggested by SHELF but not using the software. The experts were advised "to base their ranges on their own sensitivity analyses, theoretical understanding, or empirical evidence excluding any knowledge they had of the effects of the parameters in climate simulations." (Sexton et al. 2019, p. 995). The experts also provided guidance on selecting the shape of the distribution.

Observations are used at various stages of the production process. First, they are used to filter the PPE to extract the most plausible and diverse set of models. Reanalysis datasets from the ECMWF are used to assess the short term (5-day) hindcasts (see Williams et al. 2013, p. 3259) and the Met Office HadISST2 data (Titchner and Rayner 2014) for the longer term (5-year) simulations (see Murphy et al. 2018, pp. 41-45). Observations are also used to assess how PPE performs in simulating large scale circulation, like AMOC.

So, the global projections draw from three different types of evidence, and *generally satisfy* the "number" component of diversity and completeness (score: 3). We note that the score of this component depends on the variable in question. For example, if we assess global projections about mean temperature, the level of theoretical understanding of thermodynamic response to GHG concentrations warrants a lower number of types of evidence to achieve the same score as model derived statements about regional precipitation patterns.

We can now evaluate the independence and comprehensiveness of the evidence. Independence cannot be assessed for expert elicitation and model-based evidence, because the origin of the experts is not disclosed (score 0), but it is generally satisfied for model-based evidence and observations (score 3). For any variable, the PPE represents a more comprehensive evaluation than the MME, because the "plausibility and diversity" principles are applied only for developing the PPE and not the MME. Nevertheless, both ensembles contribute to the overall projections, and overall comprehensiveness is therefore *somewhat satisfied* (score 2). To improve along both diversity and completeness, then, the source of the experts should be revealed – and the experts should be sought from international research centers. Moreover, "plausibility and diversity" principles could also be applied for the evaluation and selection of components of the MME.

Historical Empirical Adequacy

Different datasets are used to assess the historical performance of models at different timescales (e.g., the 5-day and 5-year evaluations described in Murphy et al. 2018, p. 41). The discussion in Murphy et al. (2018) does not provide information about the empirical adequacy of the output of individual models, but the agreement between model output and observations is discussed with examples in Lowe et al. (2018).

Fig. 2 shows the output two random models from the global projections (Model A and Model B) and the NCIC observations for temperature anomaly, wind speed anomaly and precipitation rate. There are several problems with this evaluation of empirical adequacy. First, the issues tied to using anomalies to assess the empirical adequacy of models discussed above are also relevant here. Second, the comparison of observations and model output for wind speed anomaly and precipitation do not support a high score on this dimension. The models illustrated do not appear to capture enough of the variability for wind speed anomaly although whether this is an artifact of model selection or a more general issue is unclear. The precipitation rate output shows a lot of variation between different models but there is no guidance on how to interpret this variation? Understanding these issues is important because the features of atmospheric systems that influence variables such wind speed and precipitation are not as well understood as those that influence temperature (see Risbey and O'Kane 2011) so the theory quality dimension cannot take the slack for limited empirical adequacy.

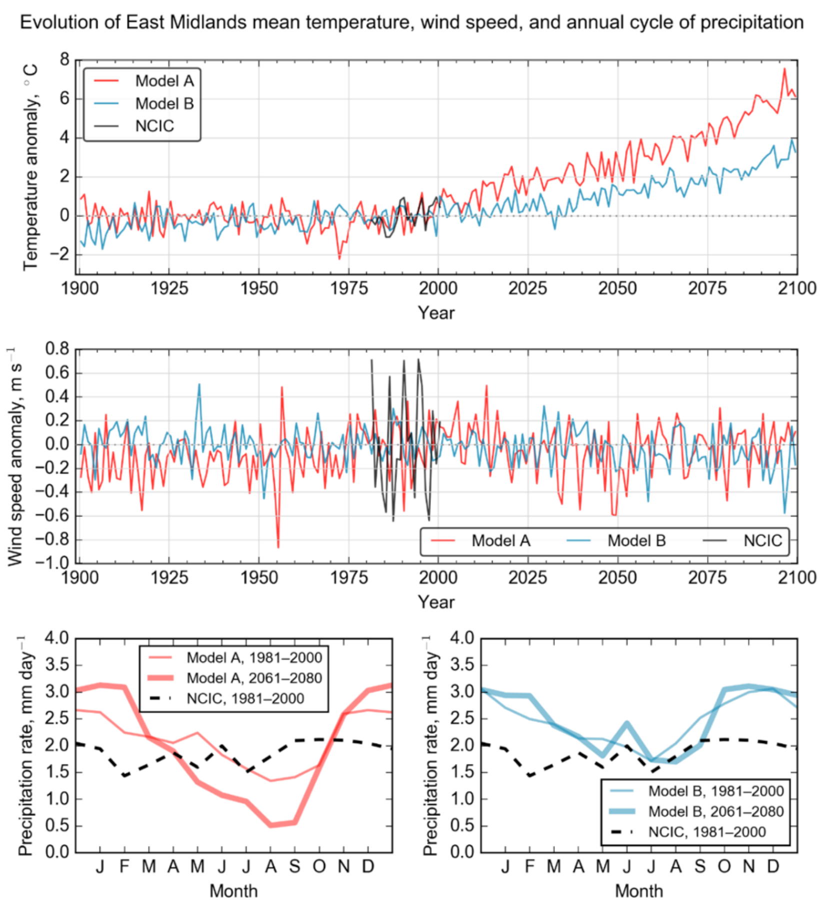


Figure 2. Agreement between model output and NCIC observations for the global projections over the East Midlands. The model resolution is 60km. The top two panels show model output and observations on annual timescales and the bottom panel shows model output and

observations at monthly time scales. Thin and thick curves show averages over different time periods for the same model (Lowe et al. 2018, p.33).

There are further issues with how observations are used to assess model output. The global projections pass two filtering stages where hindcasts are assessed for 5-day and 5-year periods. The selection of these periods is not described in much detail. For example, 5-day hindcasts are only performed for data within the 2008/09 period (Williams et al. 2013, p. 3259), and the science report of Murphy et al. (2018) does not specify the years for which the 5-year simulations have been performed. Furthermore, the adequacy of all the output of the global projections cannot be assessed for many of the variables of interest. Moreover, Fig. 2 suggests that empirical adequacy is not satisfied for variables such as wind speed anomaly and precipitation by some or all of the models. The historical empirical adequacy dimension is therefore *not satisfied* (score 0). To improve this score, the performance of individual models with respect to absolute values of the variables of interest should be more explicitly discussed for each model of the ensemble.

3.3 Regional Projections

The regional projections serve the same purpose as the global ones and follow a similar methodology (Murphy et al. 2018). There is therefore considerable overlap in the assessment and recommendations for improvement of these projections with the above global projections. There are, however, two main differences between these projections. First, the regional projections only use models from the Hadley Centre (no CMIP5 data). Second, the regional projections are developed using a one-way nesting approach to dynamically downscale the projections over the UK by forcing a PPE of regional models with a PPE of global models.

Transparency

The regional projections *somewhat satisfy* the transparency dimension (score 2) for similar reasons as the global projections. As we will discuss below, some of the dimensions are difficult to assess either because the sources of evidence are not easily accessible or because accessing them would require a user to have the skills to analyze the data themselves. For example, the analysis given by Murphy et al. (2018, pp. 95-107) only shows model performance with respect to temperature and precipitation, while many other variables (such as wind speed, cloud cover, relative humidity) are available through the user interface (Fung 2018). Higher transparency could be achieved by following the same recommendations that were given for the global projections above.

Theory

While the regional projections share many methodological assumptions with the global projections, the evaluation of the regional projections includes some additional theoretical considerations. For example, model performance in reproducing European climatology is part of the assessment process. As with the global projections, model performance in reproducing past climatology and major synoptic systems doesn't guarantee that they can predict future changes. Theoretical support is needed to relate past model performance to key processes and how these processes might respond to higher GHGs concentrations. There are many difficulties in making such an assessment. For instance, the extent to which large scale systems such as "atmospheric blocks" will affect temperature extremes over Europe and nearby regions such as the UK is still a matter of debate (Voosen 2020).

These considerations are important for the global projections but are magnified in the case of downscaled information. Possible biases introduced by downscaling are assessed for temperature and precipitation (Murphy et al. 2018, pp. 95-107). However, Giorgi (2020, p. 435) notes that the dynamical components of climate models are not well understood, and downscaling adds complexity to the evaluation of the model. Hence, as for the case of the probabilistic projections, reliance on only one modelling strategy may hide significant biases the consequences of which are not explicitly addressed. The theory dimension is therefore only minimally satisfied by the regional projections (score: 1). To improve the theory dimension, more explicit justification for the choice of downscaling method (see, e.g., Rummukainen 2010, 2016; Ekström et al. 2015) and possible consequences for model output should be included in the documents.

Diversity and Completeness

Observations, model output and expert elicitation are the three main types of evidence used here. So, like the global projections, the regional projections *generally satisfy* number (score: 3) and *somewhat satisfy* comprehensiveness (score: 2).

However, the regional projections only minimally satisfy independence (score: 1). First, the models that are used for the regional projections are all from the Hadley Centre. Watterson et al. (2014, pp. 607-698) show that CMIP models have an advantage simulating temperature, precipitation and pressure levels over their home territory. But skill in reproducing past data does not directly imply good representation of the underlying physical processes–and global scale phenomena and/or teleconnections may influence future changes in the UK climate. So, the exclusion of CMIP5 models may undermine the principles of "plausibility and diversity" that guide the production of the global projections. Second, as discussed above, the downscaling step adds complexity, introducing further assumptions into the modeling process. To improve independence and comprehensiveness, more models that were not developed by the Hadley Centre should be taken into consideration. The provenance of the experts involved in the elicitation process should be diverse, too.

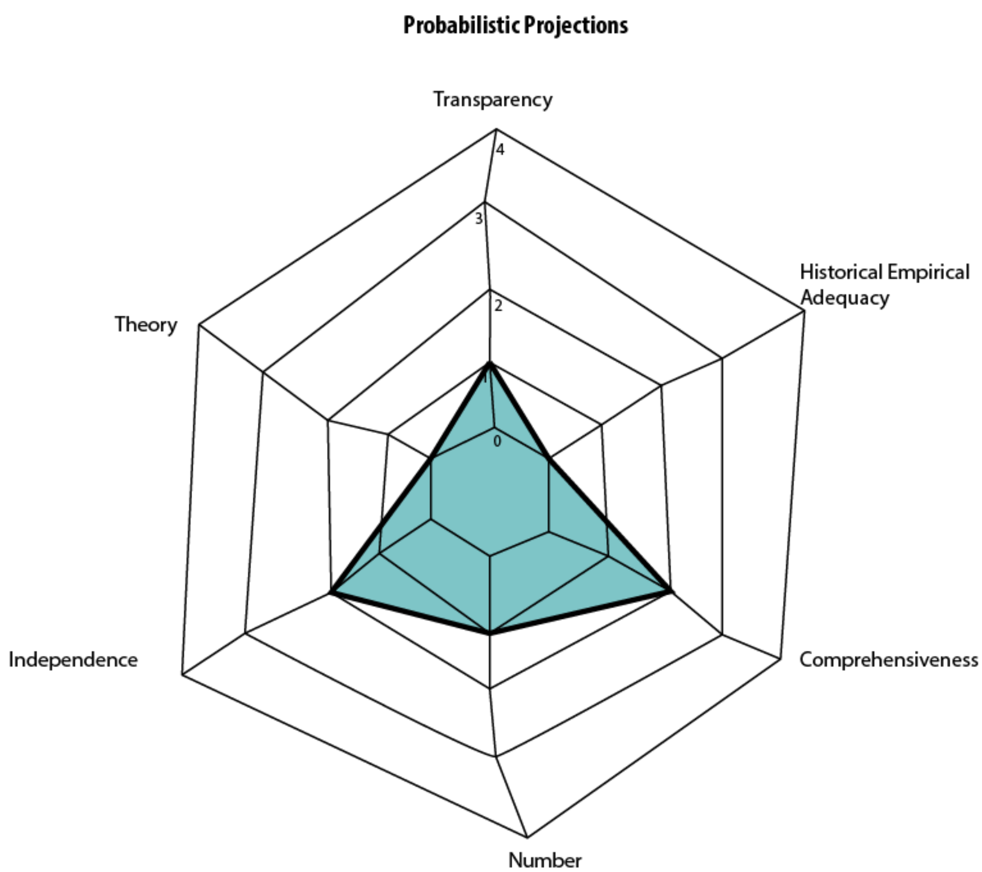
Historical empirical adequacy

The empirical adequacy of the regional projections is assessed by evaluating the performance of the regional models in reproducing European climatology, surface temperature, precipitation and AMOC strength using the NCIC dataset and the standard configuration of the GCM used for the global projections. Murphy et al. (2018) claim that model performance is also assessed for other variables, but it is not discussed in detail in the report and so cannot be assessed.

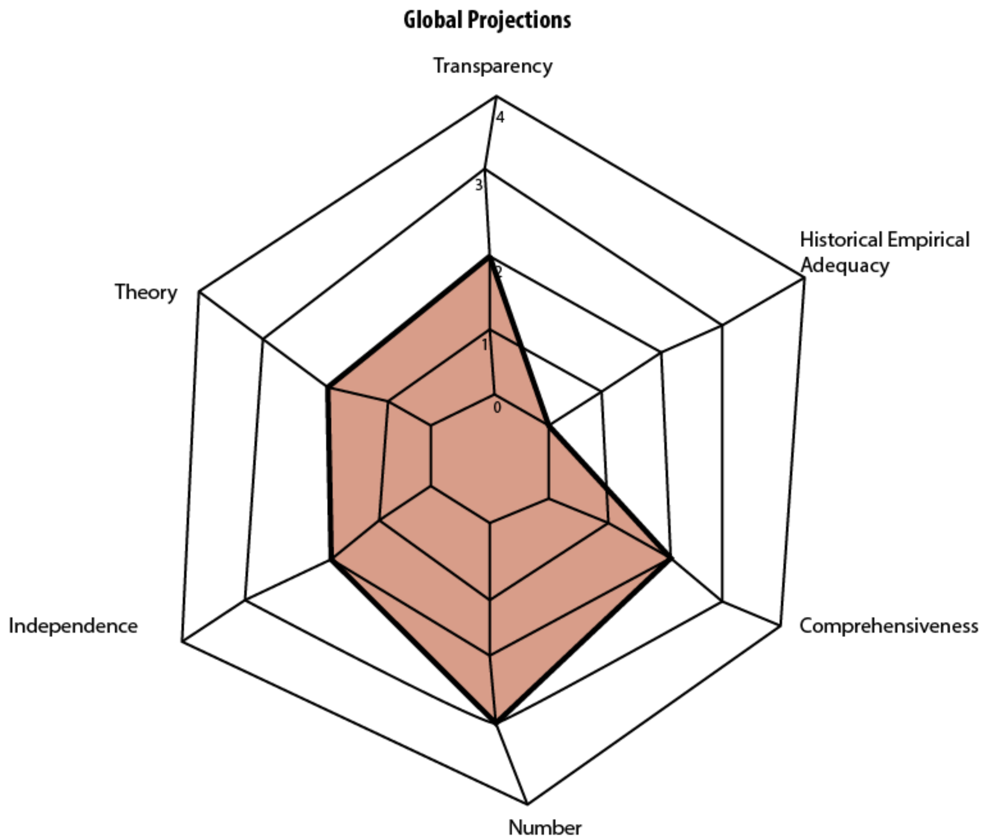
The empirical adequacy of the regional projections is described more thoroughly than for the global projections, and as discussed above, there is an extensive discussion of how data and model output are compared to observations to eliminate models with possible biases. The acknowledgement of biases in model performance for absolute values of temperature and precipitation at different spatial resolutions (see, e.g., Fig 4.5a in Murphy et al. 2018) suggest that the regional projections *generally satisfy* empirical adequacy (score: 3) for some of the variables of interest. However, there are some important caveats. First, the empirical adequacy cannot be assessed for all variables available in the regional projections. Second, a higher historical empirical adequacy does not imply a higher overall quality of the information. Furthermore, even if the regional projections have a higher historical empirical adequacy score than the global projections, they cannot have overall higher quality than the global projections due to the additional assumptions introduced by the downscaling step. Historical empirical adequacy could be improved by explicitly discussing model performance for each variable provided.

Overall assessment

The overall quality of a product cannot be assessed as the sum of the individual evaluations along the different dimensions (Baldissera Pacchetti et al. 2021). Interdependencies of the assessed products, of the quality dimensions and their relation to statements about different variables makes overall quality comparisons difficult. Nevertheless, the dimensions highlight the major strengths and weaknesses of the projections and how these are related to features of the projection construction process. Fig. 3 provides a visualization of the scores of the quality assessment for the different projections. This figure shows that the probabilistic projections have the lowest quality, and that its main shortcomings derive from lack of transparency, theoretical support and lack of adequate empirical tests. The global projections have higher quality but also lack historical empirical adequacy.

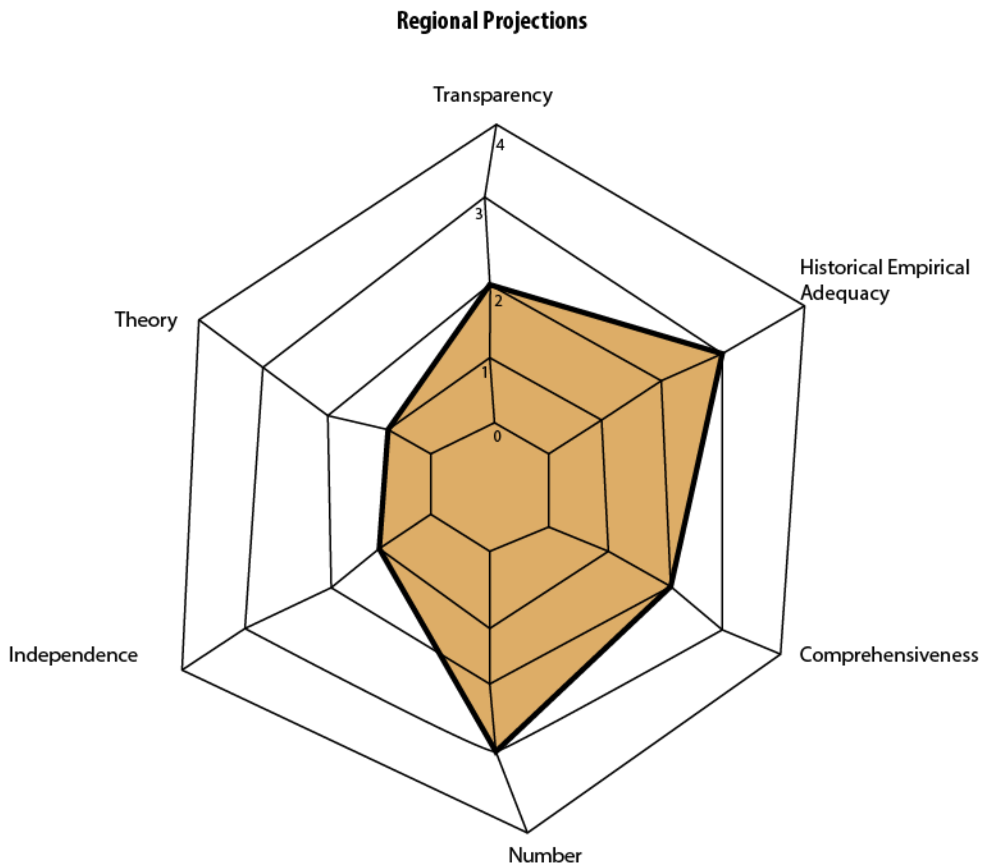


(a)


Global Projections
Transparency
Historical Empirical Adequacy
Comprehensiveness
Number
Independence
Theory
0
1
2
3
4

708

709 (b)



(c)

Figure 3 Visualization of the scores of the assessment of the probabilistic (a), global (b) and regional (c) projections. Note that scores for quality dimensions cannot be simply aggregated and there are interdependencies among different projections, so a larger shaded area does not directly imply a higher overall quality.

The higher quality of the global projections derives from two key differences. First, the global projections are not concerned with probabilistic estimates of future climate but rather with individual model simulations and potential future trajectories. This means that the evidential standards for achieving epistemic reliability are different. Second, the theoretical component– both in terms of underlying physical theory and the justification of the methodology is better justified in the global projections. The importance of synoptic weather systems and their role for understanding changes in regional weather are acknowledged, and the "plausibility" principle draws explicit attention to the physically meaningful representation of the processes that drive regional changes. Nevertheless, the above analysis shows that one cannot adequately assess the extent to which these projections satisfy key dimensions such as historical empirical adequacy of the global projections.

The regional projections have slightly lower quality than the global projections. There is little independence between sources of evidence, and the additional downscaling step, while thoroughly explained, requires additional theoretical justification for the regional projections to be adequately assessed as epistemically reliable. Moreover, the focus on the use of mostly nationally produced models raises questions about the context in which these models are granted epistemic authority (see, e.g., Mahony and Hulme 2016).

4. Towards higher quality regional climate information

We have assessed the quality of UKCP18 as an exemplar of state-of-the art regional climate information that can inform climate adaptation decision-making, and provided some suggestions for improvement. In this section, we consider some of the broader implications of this evaluation.

4.1 Transparency

A significant issue that lowers the overall quality of these products is *transparency*. Not all the data on which quality could be assessed is presented in the science report documents. Where historical empirical adequacy or the limitations of a particular methodological choice are not explicitly assessed, the task is left to the user. These considerations suggest that the concerns raised by Porter and Dessai (2017), who found that the scientists involved in UKCP09 assume that the recipients of the information they produce have similar skills to their own, are somewhat inherited by UKCP18.

This lack of transparency compromises the extent to which a user can evaluate the quality of the information produced by UKCP18. Some recently published research evaluates components of quality such as historical empirical adequacy for some climate-impact relevant metrics such as heat stress (e.g. Kennedy-Asser et al. 2021), but it is primarily aimed at an academic audience. Some documents produced by the UK Met Office such as the UKCP Enhancements[4] produce fact sheets that are aimed at improving transparency and provide more insight into how other dimensions, such as theory, could be satisfied. However, there is little integration between the documents, which itself poses a further barrier transparency; something which becomes even more important when climate information is integrated into climate services (Otto et al. 2016).

[4] To be found here: https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/index (accessed 22 February 2021)

For comparison, take the "traceable accounts" approach of the Fourth U.S. National Climate Assessment (USGCRP 2018, Chapter 2), which provides a thorough description of the information construction process. In a similar vein, the European Union's Earth Observation Programme, Copernicus, is implementing an Evaluation and Quality Control (EQC) system for all of the products available through its climate data store (CDS).[5]

4.2 Uncertainty assessment and quantification

The above quality analysis reveals that the probabilistic projections have the lowest quality. The lower quality of these projections partly lies in the probabilistic nature of the representation of uncertainty estimates, and the lack of an explanation of what these probabilities represent: the estimates provided by the probabilistic projections don't reflect confidence in the strength of the evidence.

One interpretation of the approach to uncertainty quantification followed by UKCP18 is that the authors assume that likelihoods and confidence can usefully be treated separately, and that confidence estimates can be provided at a later stage. This approach is similar to the one described by Mastrandrea et al. (2011) and used, e.g., in the Special Report on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX, IPCC 2012). But this approach has been criticized on the grounds that the distinction between confidence and likelihood is not clear (e.g. see Kandlikar et al. 2005 and Helgeson et al. 2018 for an overview), and all likelihood statements are conditional on confidence levels.

[5] https://climate.copernicus.eu/quality-assurance-climate-data-store

To clarify this point, consider the trade-offs that exist between confidence, precision and evidence described by Helgeson et al. (2018): confidence (in the epistemic reliability of a particular statement) "can be raised…by widening the probability interval … [and] less informative [i.e. precise] probability intervals may enjoy greater confidence because they are supported by additional … lines of evidence from which sharper probabilistic conclusion cannot be drawn" (p. 520). This complex interaction between the evidence and its relationship to statements about future climate emphasizes the importance of clarifying exactly how the different lines of evidence can be integrated into information production.

In particular, these considerations indicate that claiming that the probabilities are "conditional on the evidence" is an insufficient justification for providing probabilistic information aimed at decision-support. If non-quantifiable evidence lowers the confidence in the probability estimates one should one consider alternative ways of representing uncertainties about future regional climate (see, e.g. Risbey and Kandlikar 2007). If, however, a probabilistic framework of higher quality is desired, then metrics such as *theory, diversity* and *completeness* should be satisfied to a greater extent. For example, there should be a better theoretical justification of the derivation of the probability distribution functions and the kind of knowledge claim they represent, an attempt to quantify structural dependencies between the PPEs and MMEs, and an explanation of how the discrepancy term relates to real world observations rather than the MME output.

Alternatively, different ways of exploring uncertainty and knowledge claims about future climate are being developed. For example, Dessai et al. (2018) develop narratives about deeply uncertain future regional climate by drawing from expert elicitation, observation and reanalysis data. Bhave et al. (2018) exemplify this approach by using expert elicitation to develop climate narratives that are combined with socioeconomic narratives. These are then converted into quantitative output that is used to drive a hydrological model. In this approach, expert knowledge is prioritized and used to replace projections to explore plausible futures and their impact on regional scales.

Another related approach prioritizes theoretical understanding of the effects of global warming driven changes in atmospheric circulation and their impact on regional climate (Zappa and Shepherd 2017). This approach also intends to complement or replace ensemble approaches to explore uncertainties in future weather and climate extremes. Ensemble approaches can be problematic because of the sparse data availability, and the fact that changes in these events depend on the understanding of large-scale drivers, as well as regional-to-local feedback processes (Sillmann et al. 2017). This novel approach aims to assess causes of past extreme events to develop plausible storylines about future events (Shepherd et al. 2018). It also follows a distinctively different logic of research than approaches that aim at representing weather events in terms of likelihoods (Lloyd and Shepherd 2020, p. 120).

5. Conclusion

In this paper, we have applied the quality assessment framework developed by Baldissera Pacchetti et al. (2021) to state-of-the-art regional climate information in the form of the UKCP18 land projections. We started by describing the framework, its target, and the components of regional climate information that the framework assesses. We then assessed the UKCP18 probabilistic, global and regional projections along the dimensions of the quality framework.

The assessment produced two major insights that provide key recommendations for future efforts to produce decision relevant information about future regional climate. First, a significant issue that taints the quality of UKCP18 is the lack of transparency. The lack of transparency is particularly problematic if the information is directed towards non-expert users, who would need to develop technical skills to evaluate the quality and epistemic reliability of the information. Second, the probabilistic projections are the projections with lowest quality. This assessment is a consequence of both lack of transparency, and the way the method is used and justified to produce quantified uncertainty estimates about future climate.

The assessment also has some important implications for the application of the quality framework. First, it shows that there are interdependencies among the dimensions. Second, these interdependencies highlight the importance of considering the target of the framework: the evidence and methodology used to derive the statements about future regional climate, and the statements themselves. The way these elements are combined, the choice of variable(s) that the statements address, and the form the statements take, all affect the extent to which different dimensions can or should be satisfied. A quality assessment will therefore look different for a storyline about future regional precipitation by comparison to a probabilistic statement about future regional temperature, for instance.

Looking forward, we ask whether there is state-of-the-art regional climate information that is of high quality. While the quality dimensions of the framework are indeed aspirational, this analysis has shown that UKCP18 does not satisfy several of them for the products analyzed. We have argued that UKCP18 is an exemplar of state-of-the-art regional climate information, so a question that arises in this context is whether, in general, the state of the art needs to include different approaches to achieve high quality. When developing different approaches, the quality framework can be used to inform considerations about use evidence and methodology to derive high quality regional information for climate change adaptation decisions.

There are two different ways in which the above can be explored. First, through a systematic literature review that surveys the most recent research that aims to produce decision-relevant information about future climate at regional scale. Second, the framework could be applied to other products like UKCP18. For example, the Swiss National Centre for Climate Services has also released climate change scenarios (CH2018). The Royal Netherlands Meteorological Institute also releases a suite of scenarios about future regional climate in 2021. Analyses of these products would further demonstrate the value of the quality assessment framework and reveal whether it can detect subtle differences in quality in information produced by different groups of scientists using different methodologies.

Finally, an important yet unexplored aspect of quality is the inclusion of a user perspective. It is increasingly understood that including end-user needs is important for making the information accessible and salient, especially as climate information is incorporated into climate services (Clifford et al. 2020). Understanding how a quality assessment framework might change as the information moves from research and producers to users and centers of knowledge co-production is an important yet unexplored ramification of this research.

Declarations:

Funding:

This research was supported by the U.K. Economic and Social Research Council (ES/R009708/1) Centre for Climate Change, Economics and Policy (CCCEP) and Research England QR-SPF at the University of Leeds.

Author Contributions:

M. Baldissera Pacchetti, S. Dessai, D. A. Stainforth and S. Bradley contributed to the study conception and design. The first draft of the manuscript was written by M. Baldissera Pacchetti and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

Ethics Approval:

We declare no conflict of interest, and that this research did not involve human participants and/or animals.

Consent to participate:

All authors consent to participate.

Consent to publish:

All authors give consent to publish this article.

Competing Interests:

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Acknowledgements:

Suppressed for blind review.

Availability of data and materials:

All materials used for this research are publicly available.

References

Baldissera Pacchetti M, Dessai S, Bradley S, Stainforth DA (2021) Assessing the quality of regional climate information. Bull Am Meteorol Soc 102:E476-E491. https://doi.org/10.1175/BAMS-D-20-0008.1

Baldissera Pacchetti, M (2020) Structural uncertainty through the lens of model building. Synthese. https://doi.org/10.1007/s11229-020-02727-8

Bhave AG, Conway D, Dessai S, Stainforth, DA (2018) Water resource planning under future climate and socioeconomic uncertainty in the Cauvery River Basin in Karnataka, India. Water Resour Res 54:708-728. https://doi.org/10.1002/2017WR020970

Brohan P, Kennedy JJ, Harris I, Tett SFB, Jones PD (2006) Uncertainty estimates in regional and global observed temperature changes: a new data set from 1850. J Geophys Res Atmos 111:D12106. https://doi.org/10.1029/2005JD006548

Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, Jäger J, Mitchell RB (2003) Knowledge systems for sustainable development. Proc Natl Acad Sci USA 100:8086-8091. https://doi.org/10.1073/pnas.1231332100

Clifford, KR, Travis WR, Nordgren, LT (2020). A climate knowledges approach to climate services. Clim Serv 18:100155. https://doi.org/10.1016/j.cliser.2020.100155

Daron J, Burgin L, Janes T, Jones RG, Jack C (2019) Climate process chains: examples from southern Africa. Int J Climatol 39:4784-4797. https://doi.org/10.1002/joc.6106

Dessai S, Bhave A, Birch C, Conway D, Garcia-Carreras, L, Gosling JP, Mittal N, Stainforth D (2018) Building narratives to characterise uncertainty in regional climate change through expert elicitation. Environ Res Lett 13:074005. https://doi.org/10.1088/1748-9326/aabcdd

Ekström M, Grose MR, Whetton PH (2015) An appraisal of downscaling methods used in climate change research. Wiley Interdiscip Rev Clim Change 6:301-319

Fiedler T, Pitman AJ, Mackenzie K, Wood N, Jakob C, Perkins-Kirkpatrick SE (2021). Business risk and the emergence of climate analytics. Nature Clim Change 11:87-94

Frigg R, Smith LA, Stainforth DA (2015) An assessment of the foundational assumptions in high-resolution climate projections: the case of UKCP09. Synthese 192:3979-4008

Fung F (2018) UKCP18 guidance: data availability, access and formats. Met Office. https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance-data-availability-access-and-formats.pdf Accessed 21 June 2021

Fung F, Lowe J, Mitchell JFB et al (2018) UKCP18 guidance: caveats and limitations. Met Office Hadley Centre, Exeter.

https://www.metoffice.gov.uk/binaries/content/assets/metofficegovuk/pdf/research/ukcp/ukcp18-guidance---caveats-and-limitations.pdf Accessed 21 June 2021

Giorgi F (2020) Producing actionable climate change information for regions: the distillation paradigm and the 3R framework. Eur Phys J Plus 135:435. https://doi.org/10.1140/epjp/s13360-020-00453-1

Goldstein M, Rougier J (2004) Probabilistic formulations for transferring inferences from mathematical models to physical systems. SIAM J Sci Comput 26:467–487 https://doi.org/10.1137/S106482750342670X

Helgeson C, Bradley R, Hill B (2018) Combining probability with qualitative degree-of-certainty metrics in assessment. Clim Chang 149: 517–525 https://doi.org/10.1007/s10584-018-2247-6

IPCC (2012) Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field CB, Barros V, Stocker TF, Qin D, Dokken DJ, Ebi KL, Mastrandrea MD, Mach KJ, Plattner GK, Allen SK, Tignor M, Midgley PM (eds)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp

IPCC (2018) Summary for policymakers. In: Masson-Delmotte V, Zhai P, Pörtner HO, Roberts D, Skea J, Shukla PR, Pirani A, Moufouma-Okia W, Péan C, Pidcock R, Connors S, Matthews JBR, Chen Y, Zhou X, Gomis MI, Lonnoy E, Maycock T, Tignor M, Waterfield T (eds) Global

Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization, Geneva, Switzerland, 32 pp

Jack, CD, Marsham J, Rowell DP, Jones RG (2021) Climate information: towards transparent distillation. In: Conway D, Vincent K (eds) Climate risk in Africa, Palgrave Macmillan, Cham, pp 17-35

Jebeile J, Crucifix M (2020) Multi-model ensembles in climate science: mathematical structures and expert judgements. Stud Hist Philos Sci A 83:44-52

Kandlikar M, Risbey J, Dessai S (2005) Representing and communicating deep uncertainty in climate-change assessments. C R Geosci 337:443-455

Kennedy-Asser AT, Andrews O, Mitchell DM, Warren RF (2021) Evaluating heat extremes in the UK Climate Projections (UKCP18). Environ Res Lett 16:014039

Knutti R (2019) Closing the knowledge-action gap in climate change. One Earth 1:21-23

Knutti R, Masson D, Gettelman A (2013) Climate model genealogy: generation CMIP5 and how we got there. Geophys Res Lett 40:1194-1199

Lloyd EA, Shepherd TG (2020) Environmental catastrophes, climate change, and attribution. Ann N Y Acad Sci 1469:105-124. https://doi.org/10.1111/nyas.14308

Lowe JA, Bernie D, Bett P et al (2018) UKCP18 science overview report. Met Office. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Overview-report.pdf

Mahony M, Hulme M (2016) Modelling and the nation: institutionalising climate prediction in the UK, 1988–92. Minerva 54:445-470

Masson D, Knutti, R (2011) Climate model genealogy. Geophys Res Lett 38:L08703. http://dx.doi.org/10.1029/2011GL046864

Mastrandrea MD, Mach KJ, Plattner GK, Edenhofer O, Stocker TF, Field CB, Ebi KL, Matschoss PR (2011) The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. Clim Change 108:675-691.

Millner A, Calel R, Stainforth DA, MacKerron G (2013) Do probabilistic expert elicitations capture scientists' uncertainty about climate change? Clim Change 116:427-436

Moss, RH, Meehl GA, Lemos MC et al (2013) Hell and high water: practice-relevant adaptation science. Science 342:696-698. https://doi.org/10.1126/science.1239569

Murphy JM, Harris GR, Sexton DMH et al (2018) UKCP18 land projections: science report. Met Office. https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf

Nissan H, Muñoz ÁG, Mason SJ (2020) Targeted model evaluations for climate services: a case study on heat waves in Bangladesh. Clim Risk Manag 28:100213

Oakley JE, O'Hagan A (2010) SHELF: the Sheffield elicitation framework (version 2.0). School of Mathematics and Statistics, University of Sheffield, UK. http://tonyohagan.co.uk/shelf.

Otto J, Brown C, Buontempo C et al (2016) Uncertainty: lessons learned for climate services. Bull Am Meteorol Soc 97:ES265-ES269. https://doi.org/10.1175/BAMS-D-16-0173.1

Parker WS (2011) When climate models agree: the significance of robust model predictions. Philos Sci 78:579-600.

Parker WS, Risbey JS (2015) False precision, surprise and improved uncertainty assessment. Phil Trans R Soc A 373:20140453

Parker W (2016) Reanalyses and observations: what's the difference?. Bull Am Meteorol Soc 97:1565-1572

Parker WS (2020) Model evaluation: an adequacy-for-purpose view. Philos Sci 87:457-477

Pirtle Z, Meyer R, Hamilton A (2010) What does it mean when climate models agree? A case for assessing independence among general circulation models. Environ Sci Policy 13:351-361

Porter JJ, Dessai S (2017) Mini-me: why do climate scientists' misunderstand users and their needs? Environ Sci Policy 77:9-14

Risbey JS, Kandlikar M (2007) Expressions of likelihood and confidence in the IPCC uncertainty assessment process. Clim Change 85:19-31

Risbey, JS, O'Kane TJ (2011) Sources of knowledge and ignorance in climate research. Clim Change 108:755-773

Rummukainen M (2010) State-of-the-art with regional climate models. Wiley Interdiscip Rev Clim Change 1:82-96. https://doi.org/10.1002/wcc.8

Rummukainen M (2016) Added value in regional climate modeling. Wiley Interdiscip Rev Clim Change 7:145-159. https://doi.org/10.1002/wcc.378

Sexton DMH, Murphy JM, Collins M, Webb MJ (2012) Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. Clim Dyn 38:2513-2542. https://doi.org/10.1007/s00382-011-1208-9

Sexton DMH, Karmalkar AV, Murphy JM, Williams KD, Boutle IA, Morcrette CJ, Stirling AJ, Vosper SB (2019) Finding plausible and diverse variants of a climate model. Part 1: establishing the relationship between errors at weather and climate time scales. Clim Dyn 53:989-1022. https://doi.org/10.1007/s00382-019-04625-3

Sillmann J, Thorarinsdottir T, Keenlyside N et al (2017) Understanding, modeling and predicting weather and climate extremes: challenges and opportunities. Weather Clim Extremes 18:65-74. https://doi.org/10.1016/j.wace.2017.10.003

Singh R, AchutaRao K (2020) Sensitivity of future climate change and uncertainty over India to performance-based model weighting. Clim Change 160:385-406. https://doi.org/10.1007/s10584-019-02643-y

Shepherd TG, Boyd E, Calel RA et al (2018) Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. Clim Change 151:555-571. https://doi.org/10.1007/s10584-018-2317-9

Stainforth DA, Aina T, Christensen C et al (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature 433:403-406. https://doi.org/10.1038/nature03301

Stainforth DA, Allen MR, Tredger ER, Smith LA (2007a) Confidence, uncertainty and decision-support relevance in climate predictions. Phil Trans R Soc A 365:2145-2161. https://doi.org/10.1098/rsta.2007.2074

Stainforth DA, Downing TE, Washington R, Lopez A, New M (2007b) Issues in the interpretation of climate model ensembles to inform decisions. Phil Trans R Soc A 365:2163-2177. https://doi.org/10.1098/rsta.2007.2073

Thompson E, Frigg R, Helgeson C (2016) Expert judgment for climate change adaptation. Philos Sci 83:1110-1121

Thompson EL, Smith LA (2019) Escape from model-land. Economics Discussion Papers, No 2019-23, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2019-23.

Titchner HA, Rayner NA (2014) The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. J Geophys Res Atmos 119:2864-2889. https://doi.org/10.1002/2013JD020316

USGCRP (2018) Impacts, risks, and adaptation in the United States: Fourth National Climate Assessment, Volume II [Reidmiller DR, Avery CW, Easterling DR, Kunkel KE, Lewis KLM, Maycock TK, Stewart BC (eds)]. U.S. Global Change Research Program, Washington, DC, USA, 1515 pp.

Voosen P (2020) Why weather systems are apt to stall. Science 367:1062-1063.

Winsberg E (2006) Models of success versus the success of models: reliability without truth. Synthese 152:1-19. https://doi.org/10.1007/s11229-004-5404-6

Watterson IG, Bathols J, Heady C (2014) What influences the skill of climate models over the continents? Bull Am Meteorol Soc 95:689-700. https://doi.org/10.1175/BAMS-D-12-00136.1

Williams KD, Bodas-Salcedo A, Déqué M et al (2013) The transpose-AMIP II experiment and its application to the understanding of southern ocean cloud biases in climate models. J Clim 26:3258-3274. https://doi.org/10.1175/JCLI-D-12-00429.1

Zappa G, Shepherd TG (2017) Storylines of atmospheric circulation change for European regional climate impact assessment. J Clim 30:6561-6577. https://doi.org/10.1175/JCLI-D-16-0807.1