

# The co-development of models with expert judgement suppresses model diversity and underestimates risk

Erica Thompson (✉ [e.thompson@lse.ac.uk](mailto:e.thompson@lse.ac.uk))

London School of Economics and Political Science <https://orcid.org/0000-0003-3750-8388>

---

## Research Article

**Keywords:** Climate models, Expert judgement, Climate risk, Climate decision-making, Uncertainty quantification

**Posted Date:** February 25th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-234517/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

It is well-understood that mathematical models are developed with the aid of expert judgement about the relevant real-world processes. Here, I first consider the influence of the mathematical model itself upon our expert judgement. Having established a two-way relationship between model and expert judgement, I then consider some epistemic concerns arising from the difficulty of separating model information from expert judgement, concluding that statistical methods for assessing uncertainty in climate sensitivity are inadequate. I offer two potential solutions to these epistemic concerns. One: pre-registration of model experiments prior to the development of a model, and another: active exploration of other possible model structures based on different starting points and (ideally) different expert input. Next, I offer some practical recommendations for communication of climate information: we must distinguish statements about a model from statements about the real world, and it could be useful to consider ending climate simulations at a threshold of temperature change rather than on an arbitrary time scale. Having identified the climate modelling community (of researchers/modellers and the mathematical and computational entities which are the models themselves) as a cognitive assemblage, I suggest that policy-relevant science and decision-making would benefit from closer examination of the kinds of models and expertise that are privileged within this system and from actively encouraging greater model diversity. The current suppression of diversity results in underestimation of risk. Although I use climate as an example, these recommendations are generalisable to other policy-relevant modelling communities.

## The Effect Of Expert Judgement Upon One'S Model

When a mathematical or computational model of a system is first developed from scratch, by someone who is a domain expert, they will use their expert judgement to construct that model. First, in the decision of what factors are important to include and what structural form the model should take. Second, in the decision of how to assimilate quantitative observations or information into the model (sometimes called calibration or tuning). And third, in the decision of how to evaluate the predictions of the model relative to new observations, in order to get an idea of its predictive capability (sometimes called uncertainty quantification). The expert judgements that go into each of these steps will be formalised in quantitative language and methods. In practice, of course, it is a highly iterative procedure rather than a single initial judgement. Perhaps the expert will begin with a minimal process representation and work up, including further elaborations or more detail until the model becomes “sufficiently good” at representing the outputs or behaviours of interest.

This is uncontroversial, and will be familiar to anyone who has been trained in the physical sciences. The use and formalisation of expert judgement is discussed at length elsewhere. What I wish to do here is to explore further the reverse effect: the way that the model influences one's expert judgement. Once we have a two-way system, then there is also the possibility for feedback loops, negative and positive.

## The Effect Of The Model Upon One'S Expert Judgement

What do you use your model for? Is it a one-way system of construct, calibrate, predict? Or is it actually a sandpit that you use to explore the consequences of different assumptions or scenarios, to test hypotheses and play out counterfactuals? Climate models are certainly in the latter category: they are an important research tool which help develop understanding of the small-scale elements and the overall behaviour of the system. They are not used solely to represent and then predict; they are also driving the scientific process themselves as an inseparable part of the cognitive processes that go into our understanding of climate.

The initial information that goes into such a model is domain expertise gathered in other ways: observations, laboratory experiments, field experiments, the fundamental laws of physics such as conservation laws, the accepted approximations of physics such as ideal-gas laws, mathematical techniques of numerical approximation and integration, understanding of emergent phenomena such as atmospheric waves, and so on. All of those things foster understanding of the system and an ability to make certain kinds of prediction, but the human mind cannot yet perform detailed quantitative extrapolations. To do so, we use the technology of our numerical climate models as a cognitive aid.

Climate models and their component modules have contributed greatly to our understanding of the climate system, in particular the understanding of complex processes such as cloud feedbacks, aerosol effects, and the carbon cycle. The ability to experiment on these processes both separately and in the context of a full climate model is an essential contribution to our understanding, supported by all of the other domain expertise mentioned above. As well as being a *product of* our expert judgement, the model is therefore also *contributing to* our expert judgement.

Like the physical aid of a wheelchair, this cognitive aid both *extends and constrains* our abilities, and both of these effects are significant. The ways in which our abilities are extended are clear, so I will here focus on the constraints, which are less well explored.

## Implications Of This Viewpoint

First, I would say that the increasing accessibility of climate modelling has very positive effects. It is making this particular cognitive aid available to a wider group of climate stakeholders, including those in so-called “developing” countries. With that accessibility comes greater transparency and opportunities for wider criticism and development. These are all good things.

Having said that, I now raise some second-order concerns. What are the constraints that we experience, as a community of experts, due to the widespread use of climate models to develop our expert judgement? These I separate into academic or epistemic concerns about the quality of knowledge generation, and practical concerns about the real-world implications.

## Epistemic Concerns

If models and expert judgements are not separable, because they are co-developed in the way I described above, then they do not form fully independent sources of information. This has two consequences:

- First, it undermines the semi-formal use of expert judgement to evaluate model output. By this I mean qualitative model evaluations determining that one pattern of behaviour is “more realistic” than another, or that outputs lying outside a particular range are “unlikely”.
- For example, if all of our previous models have shown values of a certain parameter between 1.5 and 4, then we will naturally examine more closely (and perhaps tweak or recalibrate) those that show it to be outside that range. This generates an effective bias towards the “accepted” values.
- Second, it undermines quantitative statistical methods for model analysis which effectively assume that ensembles constitute “random samples” of some informative model space. Because models are based on the same expert judgements, and for other reasons outlined elsewhere, they are not a random sample of anything. In fact, Mauritsen and Roeckner (2020) describe how the climate of one CMIP6 model, MPI-ESM-1.2, was tuned with the specific target of a climate sensitivity of 3.0, in the middle of the expected range and consistent with observation-derived estimates from the twentieth century.

In the words of Donald MacKenzie (2008), this kind of model is “an engine, not a camera,” in the sense that the model itself, which could have been constructed in innumerably many different ways but happens to be constructed in this particular way for a set of particular reasons, is a key driver of the framing and understanding of the situation it models.

We are in a feedback loop, in which the success of the model in reflecting our expert judgement back at us is amplified by the use of expert judgement to assess the quality of the model. Of course, this feedback loop is still connected to reality via observational constraints, so I do not argue that we are in a position of complete epistemic ignorance. My own overall feeling (expert judgement, if you will) is that plausible uncertainty ranges for parameters such as climate sensitivity are simply underestimated, especially at the end of the range which is furthest from observational experience.

## Epistemic recommendations

I offer two potential ways out of this feedback loop, though neither is easy.

The first is to cut it in the middle, by disallowing *all forms of* expert judgement as in-flight model evaluation tools. I propose that this could be achieved by following a procedure such as **pre-registration** of all model experiments prior to the development of a model. Pre-registration would include advance commitment to

the inputs, processes and outputs of the model;

the observational data to be used for tuning and calibration;

the observational data to be used for out-of-sample assessment;

an evaluation metric by which the above data are to be compared with the model output;

“blind” model development, either without running and testing routines, or by subcontracting it to a programmer who will be able to implement the proposal without making domain-expert judgements about the quality of the output;

any other advance stipulations about quality (“disregard all models with a negative climate sensitivity”; “disregard all models where the North Atlantic is frozen”), with clear justification for each;

publication of all results regardless of perceived quality.

No doubt further expert judgements would then follow. This procedure does not prevent us tuning the model to any pre-specified desirable outcome (such as fit to twentieth century climate) but it would prevent us from revisiting tuning parameters multiple times in order to get the best-looking result.

A second way out of the feedback loop would be to **create other possibilities**: accept the contingency on expert judgement, but actively make attempts to explore other possible regions of model space. Previous attempts to do this have been extremely minimal, generally limited to small perturbation of parameters within an existing model. I propose a much wider programme. For example, taking the current suite of state-of-the-art climate models based on atmospheric fluid dynamics as our default, could we instead imagine a “climate model” which is instead primarily based upon highly detailed representations of the biosphere, with the atmosphere parameterised down to a couple of key equations? This would have to be developed from existing ecological models, in the way that present state-of-the-art climate models have developed from numerical weather prediction models.

Another way to create other possibilities would be to fund the creation of a new climate centre, based in say Africa or Bangladesh, staffed with people who have specifically *not* had experience of conventional (European/American-style) climate modelling studies, and tasked with the preparation of policy-relevant climate information in any way they deem appropriate. This is something of a “moon shot” but with time to trust the process and develop a clarity of purpose, I believe it could have a deeply energising effect, generate significant new thought and improve the integration of non-Western viewpoints into global decision processes.

## Climate Modelling Community As Cognitive Assemblage

A cognitive assemblage (Srnicek, 2014; Hayles, 2016) is a collection (assemblage) of interconnected entities which performs an act of information processing (cognition). Since neither human mind nor computer alone could integrate climate information and arrive at the kinds of conclusions that inform climate decision-making, I will say that the community of researcher-modellers and climate models together form a *cognitive assemblage*, part-human and part-model. This assemblage interacts internally in a more-or-less coherent way, via the process of expert judgements described above. Externally, it gains information:

from the non-human environment, via observation and measurement of climate variables and other prior knowledge/assumptions about physics and mathematics;

from the political and social environment, via policy priorities, the UNFCCC process, the IPCC process, national and international research programmes.

To oversimplify a little for the sake of clearer discussion, we might define the overall outputs of this cognitive assemblage as being the numerical outputs of the international model intercomparison project (CMIP and its repository of climate model data from simulations performed with pre-industrial, twentieth-century, and possible future forcings) and the Working Group I contribution to the IPCC's fifth Assessment Report. Notably, the IPCC report does much more than describe and summarise model output: it offers also an expert evaluation of the level of confidence in model output which goes beyond simple measures of inter-model consistency and describes also the possible consequences of the models themselves being inadequate. This is another entry of expert judgement into our understanding of the climate system, and a particularly important one.

## Practical Concerns

My practical concerns are related to the widespread use of models and visualisations of model output to **communicate** policy-relevant scientific results and **inform** decision-making, and the way in which these are presented as being “the” evidence.

The headline IPCC projections of expected global increase in temperature due to changes in greenhouse gas concentrations are derived from models. The 90% range of model outcomes is identified with a “likely” range, within which the real-world outcome is deemed to have at least a 66% chance of falling (see discussion in Thompson, Frigg and Helgeson, 2016). This is an expert judgement about the quality of the models and the possibility (by implication, approximately 24%) that they are wrong in a way which would see the global mean temperature fall outside the 90% range of models at the end of this century. The global mean temperature is the variable in which we have most confidence in predicting. Yet it is common to see, even elsewhere in the same IPCC report, model results presented as though frequencies of model occurrences can be directly used as an estimate of the probability of the real-world outcome. Presenting a model-world frequency as a real-world probability is effectively an implicit expert judgement that the model is perfect. This is clearly unreasonable in almost all circumstances.

Modelling can be indefinitely extended (by adding finer detail or further process representations) even within the limits of computational resource, which continues to grow exponentially. Therefore, there is no prospect of reaching a state where any model is “finished”; they can only be abandoned or superseded. This makes it particularly important to judge the adequacy for purpose of a model in any decision-making context. There is a dilemma for researchers in framing the output of the model, with a pressure to provide best-available results for decision support (“the model is adequate for purpose”), and an equal and opposite pressure to request resources for further model development (“the model is not yet adequate for purpose”). The dilemma is easily seen in the abstract, introduction and conclusions of

policy-relevant publications describing model results from climate, finance, epidemiology, and many other fields. The question of whether or not a weather model is adequate for the purpose of deciding how to act next week is answerable (near enough) by recourse to previous data gathered in similar circumstances (the expert judgement is involved only in deciding that past data are a good guide to future success). The question of whether or not a climate model is adequate for the purpose of deciding how to act in the next decades is answerable only by exercising expert judgement in tandem with previous data gathered in moderately-similar circumstances.

Mathematical models, though produced on supercomputers and dressed up with the latest statistical analyses, are social constructs which inherit the attitudes of the system which their creators inhabit, including attitudes towards risk and uncertainty, and value judgements about the importance of different elements. They are shaped by the predict-and-control paradigm of twentieth-century Western science and technology, and the underlying belief that all we need is more information to solve the problem. Now we have an ability to produce detailed visualisations of possible future climatic regimes, policy responses privilege these quantitative predictions as decision-making inputs rather than other perspectives. I do not propose to go into other perspectives in detail here, only to note that there are many ways of conceptualising and describing the relationships of humans with the natural world which do not involve quantitative modelling or even prediction per se, and that those conceptualisations can also be sufficient to motivate and direct action towards shared goals.

## Practical Recommendations

For clear communication and effective decision support, it is critical to “escape from Model Land” and to **distinguish statements about a model from statements about the real world** (see also Thompson and Smith, 2019). It would help to avoid confusion and misinterpretation if the next IPCC report were to commit to making only statements about the real world in the Summary for Policymakers.

Climate models do reasonably well at reproducing the climate of the twentieth century. Therefore, it is plausible that they will do reasonably well in reproducing similar climates in the future, and that they may be adequate for the purpose of informing some kinds of decision-making in futures which are not very different from the present. However, the further we depart from present climate, the less epistemic warrant we have for assuming that the models are a good representation of reality, and the more we are relying on expert judgement rather than quantitative assessment of model against observations to make that assumption. One possible approach would be **to end simulations at a threshold of global temperature change**, say 2 degrees Celsius, rather than at an arbitrary date of 2100. There would be a new question of where to place the threshold, which I believe would stimulate interesting and informative debate about the plausible failure modes of climate models. For example, what kinds of events could occur that would demonstrate that the model was inadequate for purpose in a certain way? It might even be that such debate reveals that experts are *more* confident than implied by the IPCC’s uncertainty language, which would be useful information for decision-makers.

# Conclusions

The quantification of uncertainty in projections of the global mean temperature under significant forcing scenarios, informed by climate model simulations, is essentially an impossible task. Simulations are extended beyond the domain of calibration, and physical intuition is extended beyond the domain of experience. This does not mean we know nothing. It means we must be honest about the types and confidence levels of the knowledge we do have. Putting arbitrary numbers onto uncertainty levels is only of value in order to put those numbers into a cost-benefit analysis, to try to optimise a certain outcome. And cost-optimisation is only one limited frame for decision-making about the future. Many other decision frameworks are available, which do not require detailed quantitative projections of all variables. I have above suggested some specific options for changing or adding to climate modelling procedures, to help understand and communicate the robustness of our model results.

I must also conclude that the methods used by the IPCC's Working Group 1 (Stocker, 2014) to assign probabilities to future global mean temperatures are inadequate. Since models are tuned (either directly or implicitly) to produce a climate sensitivity within an acceptable range, the existence of that range is entirely an expert judgement and not a model quantity. If the expert judgement of the IPCC authors (Stocker, 2014) is that there is a 24% chance of the real-world outcome falling outside the 90% confidence interval of model runs, then the models are not sampling a wide enough region of space: they are overfitted. Framing ensembles of climate model outcomes as being statistically independent estimates of a true value is misleading, whether it is done verbally, by presentation of model outcomes on a scatterplot, or by the application of statistical techniques which make that assumption.

The consequence of underestimating the dynamic interplay between models and expert judgement is overconfidence. Overconfidence leads to underestimation of uncertainty; underestimation of uncertainty leads to underestimation of risk. By suppressing model diversity, it is likely that we are underestimating the potential physical risks of climatic change over the next centuries.

Finally, I emphasise that modelled quantitative projections of future climate states are only one of many inputs to "climate decision-making" over the next century. All "climate decisions" are also political decisions about which industries to support or restrain, which goals to prioritise, which voices to amplify or to ignore. All "climate decisions" are also moral decisions about whose lives matter; what species matter; what levels of risk we are prepared to live with and accept on behalf of future generations. In framing climate decisions as technical decisions primarily to be answered by modelling studies, it is imperative to consider the political and ethical dimensions of that framing and what interests are served by doing so. I have demonstrated above that climate models and expert judgements are intimately linked as a cognitive assemblage which processes climate data to arrive at policy-relevant information and which derives confidence from expert judgement as well as by comparison with observation. Suppressing diversity results in underestimation of risk, and therefore in poor risk management. It is urgent, therefore, to decide what kinds of models and what kinds of expertise should be privileged within that system, actively encourage diversity of modelling approaches, and shape the cognitive assemblage so that it can

provide the kinds of information which are most useful for the purposes of generating a shared vision of a future we can work towards.

## Declarations

**Funding:** This work was supported by the EPSRC CRUISSE Network ([EP/P016847/1](#)), by NERC (NE/R01423X/1), and results from a stimulating workshop sponsored by the University of Leeds and the UK's ESRC Center for Climate Change Economics and Policy (ES/K006576/1)

**Conflicts of interest/Competing interests:** Not applicable

**Availability of data and material:** Not applicable

**Code availability:** Not applicable

**Authors' contributions:** ET is sole author

## References

1. Hayles N, Katherine (2016) Cognitive assemblages: Technical agency and human interactions. *Critical Inquiry* 43(1):32–55
2. MacKenzie D. *An engine, not a camera: How financial models shape markets*. MIT Press, 2008
3. Mauritsen T, and Erich Roeckner. "Tuning the MPI-ESM1. 2 Global Climate Model to Improve the Match With Instrumental Record Warming by Lowering Its Climate Sensitivity." *Journal of Advances in Modeling Earth Systems* 12, no. 5 (2020) e2019MS002037
4. Smicek N (2014) "Cognitive assemblages and the production of knowledge". In: *Reassembling International Theory*. Palgrave Pivot, London, pp 40–47
5. Stocker T, ed. *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 2014
6. Thompson E, Frigg R, Casey Helgeson. "Expert judgment for climate change adaptation." *Philos Sci* 83, 5 (2016): 1110–1121
7. Thompson E, Smith LA (2019) Escape from Model-Land. *Economics: The Open-Access, Open-Assessment E-Journal*, 13 (2019-40): 1–15
8. Declarations