

Curation of Climate Information

Climate model runs are becoming publicly available through easy-access repositories. Although this freedom of information is undoubtedly good for science and society, it raises questions around the curation of the information to help prospective users make use of not just the data but also their context, value and limitations.

Historically, output from state-of-the-art climate models was available only to the institution generating it, with elaborate archival procedures on tape and smaller summary datasets being provided to other institutions on request. More recently, the Coupled Model Intercomparison Project (CMIP) provides a structured framework for constructing cross-institutional data sets which are now available online to registered users. As models continue to increase in complexity, the sheer volume of data is still an issue. Software innovations allow interactive plotting and visualisation tools and selective download, which are extending the reach and potential use of model data.

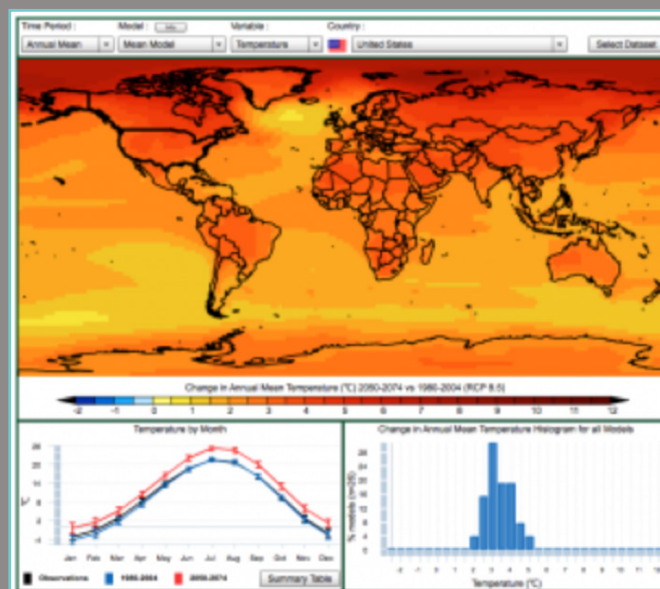
When such data can be accessed at the click of a mouse, there is a risk that it may be misinterpreted or used

inappropriately. Climate models of scientific value do not necessarily make useful real-world predictions – presentation of such results to decision makers could lead to maladaptive choices. For example, a consultancy firm engaged by a company to assess climate risk over the next decades could misinterpret high resolution data from a single model and make over-precise predictions about future climate hazards. These may then be used to determine the company's response, resulting in maladaptation.

Making raw simulation data available in this way is risky unless accompanied by clear

information about the ways in which it should and should not be used. Open discussion of this guidance itself would prove of value to climate science, climate modelling, and climate decision support.

Some suggestions for responsible data curation are provided overleaf.



Sensible use of default settings

For example, automatically masking projection results for all areas where the change is insignificant, or where the range of alternative model results is very wide. Alternatively, a hatching/stippling approach could be considered, as used by the IPCC. Default viewing resolution should be coarser than the model's native grid-scale, consistent with standard good practice of numerical methods.

Deprecating selection of a single model

Users accessing datasets from a single model could directly be referred to all other similar runs from alternative models. A “multi-model mean” should not be provided.

Integrated treatment of uncertainty

Where visualisation interfaces are provided, a range of outputs is always shown rather than a single quantity.

Use of histograms rather than “probability distributions”

Clear distinctions must be made between real quantities and quantities of the model.



Explicit categorization of all forward-projection data

We suggest a simple categorisation of any data with a real-world timestamp in the twenty-first century, according to expert judgement about its expected fidelity. This could be achieved by asking the source to tick a box and provide comments when uploading the data, consistent with the IPCC requirement for a “traceable account” of uncertainty.

● **Green:** the ranges of projected quantities are reasonably expected to be informative about the outcome, given the forcing scenarios used. Some form of quantitative evaluation and/or subjective expert judgement backs up this claim. These data can be used for decision-making purposes.

● **Amber:** the ranges of projected quantities may be useful in an indicative sense, but are not expected to be quantitatively informative, OR single values only are available, and are expected to be informative, but no ranges are provided, OR ranges are provided but their informativeness has not yet been evaluated. These data may be useful as context for decision-making but should not be used as direct quantitative input.

● **Red:** the projected quantities are not expected to reflect real-world outcomes, for example if the experiment was designed to test model behaviour in a theoretical situation. Alternatively, an evaluation may have shown this simulation to be significantly unrealistic, or to be suffering from a significant known bug.

The colour code could then appear on analyses, graphics and publications using the data.



The Centre for the Analysis of Time Series (CATS) is a research centre of the London School of Economics and Political Science. We focus on nonlinear analysis methods for decision support in situations of economic and physical significance.

Lead Author: **Erica Thompson**

lse.ac.uk/CATS/Home.aspx

Series Editor: **Leonard Smith**

