Gregor Betz

Chaos, Plurality, and Model Metrics in Climate Science

Commentary on Valerio Lucarini

1 Central findings of climate science are independent of model simulations

Since the reliability of climate models represents a politically highly sensitive issue, I would like to remind us upfront, before I comment on the interesting and illuminating paper by Valerio Lucarini, that many central findings of climate science are entirely independent of Global Climate Models (GCMs). These results include:

1. The atmospheric CO_2-concentration has reached levels unprecedented in at least the past 650.000 years (IPCC 2007, p. 24). More specifically, the CO_2-concentration varied, during the ice age cycles, between 180 and 300 ppm (IPCC 2007, p. 435).
2. The increase of atmospheric CO_2 from a pre-industrial concentration of 280 ppm to 380 ppm in 2005 is caused by human activities, notably by the consumption of fossil fuels (IPCC 2007, p. 25).
3. CO_2 is a greenhouse gas. Absorbing infrared light, it contributes to the natural greenhouse effect that heats the earth – as well as the planet Venus (Rahmstorf and Schellnhuber 2006, p. 32).
4. Increasing the CO_2-concentration is a major intervention into the global climate system, offsetting the earth's radiative equilibrium and thus causing major readjustments of the climate system. These readjustments might consist in global warming or an increase of the earth's albedo.
5. Global average surface temperature has increased by roughly 0.6° during the last century. In the Polar Regions, where climate change, as a consequence of the ice albedo feedback, is expected to be more severe, surface temperature has been increasing at twice the rate of the rest of the world. (IPCC 2007, p. 37)
6. Except for CO_2, known forcings of the climate system exhibit no trend during the last decades of the 20^{th} century. Therefore, at least the latest phase of observed global warming can be attributed to anthropogenic activities. (Rahmstorf and Schellnhuber 2006, pp. 39–40)

These findings alone might constitute a sufficient reason for considering climate change a serious global problem, which has to be addressed by suitable policies. That is why even a radical criticism of GCMs does not automatically lend support to the position of so-called climate sceptics, i.e. the position that the theory of anthropogenic climate change is simply made-up and does not call for any policy measures whatsoever.

This said, reliable Global Climate Models would nevertheless be highly valuable for practical matters, as Lucarini has rightly stressed, because there are some things we apparently cannot estimate without GCMs.

2 Some relevant questions cannot be answered without GCMs

GCMs are required to specify
1. the precise extent and timing of future global warming;
2. the regional patterns of future temperature and precipitation change;
3. the precise degree to which human activity is responsible for already observed climate change;
4. the detailed reconstruction of past climates from (sparse) proxy data.

With regard to the rational deliberation of alternative climate policy decisions, well-founded conditional predictions corresponding to the items 1 and 2 would obviously be extremely helpful. But can we reliably predict the climate? I understand that Lucarini cites three different uncertainties which prevent us from making accurate deterministic forecasts: ignorance about the precise initial conditions, ignorance about future boundary conditions, and ignorance about the causal structure of the climate system, which corresponds – in climate science jargon – to “structural uncertainty”. In his assessment of these uncertainties, Lucarini seems to presuppose that the climate system exhibits sensitive dependence on initial conditions. This prompts my first critical question.

3 Is the climate chaotic?

Is the climate system chaotic; or, indeed, does an error in initial conditions grow exponentially when predicting the evolution of the climate system? This question, I suggest, deserves a careful and differentiated consideration. Granted: The weather is chaotic. But this does not entail that the climate, which is described in terms of *average* weather, depends sensitively on initial conditions, too. It seems to me an obvious fact that some physical systems are chaotic with regard to the microstates they realize, but behave non-chaotically regarding their macrostates (Think, e.g., of boiling water, which is, regarding the location of the first vapour bubble, chaotic, yet is not in terms of the mean temperature when bubbles start to form.). By the way, this is maybe the very reason why reduction of complexity (through devising highly aggregated models) can represent a successful research strategy. So, even if the weather is chaotic, the climate is not necessarily so, and in particular not necessarily with regard to all its state variables. As a philosopher, I am, of course, not in a position to answer the empirical question which climate variables depend sensitively on initial conditions. But I would like empirical scientists to be more specific regarding the chaos hypothesis. Here is a suggestion for how the chaotic character of the climate system might be described in a more nuanced way:

- Some large-scale climate processes such as the thermohaline circulation or the indian monsoon possess, under specific boundary conditions, several equilibria. In such situations, small perturbations might determine whether the respective subsystem ends up in one or the other stable state. These climatic changes might be “abrupt” and trigger global effects (affecting, e.g. global precipitation or temperature patterns).
- Whether, say, average precipitation in northern Germany in the decade 2100–2110 is going to be higher or lower than in the current decade (2000–2010) possibly also depends on the precise climatic initial and boundary conditions such as today’s radiative forcing, heat uptake of the ocean, state and interplay of atmospheric oscillations, etc.
- But whether the emission of another 1000 GtC in the first half of this century causes the earth to warm, in 2100, by 10 or by 2 degree Celsius does not depend sensitively on today’s initial conditions.

4 The plurality of GCMs

Of the three key uncertainties Lucarini enumerated, namely (1) ignorance of initial conditions, (2) ignorance of boundary conditions, and (3) structural uncertainty, it is the last one which is responsible for the plurality of climate models employed in climate science. Unlike in economics, however, climate scientists do understand the basic, small-scale processes in the climate system. The fundamental laws describing these processes, such as the Navier-Stokes equation, are well established. It is only because of limited computational resources that these equations cannot be solved for a system as huge as earth's climate. The computational limitations call for a description of climate processes on a more aggregate scale – and it is precisely on this meso-scale where the causal picture of the climate is still inadequate. When devising a GCM, climate scientists face, as a consequence, a couple of underdetermined choices, and different groups of modellers end up with different climate models (cf. Parker 2006; Betz 2009; Lenhard and Winsberg 2010).

Thus, the 4AR relies for its major predictions as well as for the analysis of past climates on 23 different AOGCMs which are built and run by 17 institutions (IPCC 2007, p. 597). These GCMs comprise sub models of the atmosphere, the ocean, sea ice and land. Their resolutions range from 1,1°×1,1°–4°×5° with 56–12 vertical layers (where, at the equator, one degree of latitude equals a degree of longitude and amounts, roughly, to 111 km).

Given this plurality of models, the question whether one can empirically test, compare and rank these rival models arises quite naturally. This question will eventually lead us to one of Lucarini's main points, namely the proposal of a new metric for climate model evaluation.

5 Epistemic evaluation of GCMs: the role of metrics

Regarding the epistemic assessment of GCMs, it is important to separate the following two questions:

(1) What are the empirical implications of a climate model that ought to be considered during its epistemic assessment at all?

(2) What exactly can one infer from the predictive and explanatory performance of a GCM regarding the relevant empirical indicators?

The second question pertains to the general methodology of the model assessment: Should we try to falsify GCMs and refute those that, for example, give rise to false empirical retrodictions? Or do we have to construe the model evaluation along the lines of inductive modes of reasoning? Or should one assess the models in agreement with a hypothetico-deductive account of confirmation? – We will return to this second question below.

The first question concerns a more rudimentary issue, which has to be addressed before any kind of empirical assessment can be carried out. An answer to the first question, which somehow suggests itself, is to say: The empirical implication E is relevant if and only if E concerns climate variables the model is supposed to predict or to explain. And that is roughly what the IPCC assumes

(IPCC 2007, chapter 8). Accordingly, the aspects of the climate system considered in course of the model evaluation include regional mean surface temperature, the annual variability of surface temperatures, mean and annual variability of precipitation, patterns of cyclone activities, mean temperature and salinity structure of the ocean, strength and geometry of ocean circulations, the extent of sea ice, the severity and frequency of extreme weather events, large-scale processes such as the monsoon or El Niño, etc.

This plurality of relevant climate variables poses a potential problem for the assessment of GCMs since there is no climate model which outperforms its rivals in terms of empirical adequacy and with regard to all the different relevant aspects of the climate system. Every model has some strengths and some deficiencies, and they typically differ from the strengths and deficiencies of its rivals (Heffernan 2010).

It is in this situation that climate scientists would like to devise a general quantity, which aggregates all the relevant aspects, and which allows one to express the overall empirical adequacy of a GCM in one single figure. Such an aggregated variable is also referred to as a metric. The IPCC defines a metric as a consistent measurement of an object's or activity's characteristic that is otherwise difficult to quantify. (IPCC 2007, p. 949)

In its Third Assessment Report (TAR), published in 2001, the IPCC was unambiguous about any attempts to design a metric that combines all relevant empirical implications of a GCM:

> It has proved elusive to derive a fully comprehensive multi-dimensional "figure of merit" for climate models. (IPCC 2001, p. 475)

In the 4AR, however, the IPCC has become a bit more optimistic, again:

> The possibility of developing model capability measures ('metrics'), based on the above evaluation methods, that can be used to narrow uncertainty by providing quantitative constraints on model climate projections, has been explored for the first time using model ensembles. While these methods show promise, a proven set of measures has yet to be established. (IPCC 2007, p. 60)

On this background, Lucarini proposes his own, *process-oriented* metrics for model evaluation. As far as I understand, these metrics do not rely on policy-relevant observational variables but try, rather, to capture the key processes of the climate system. They are supposed to describe, based on simulation results or on observational data, the central causal mechanisms that drive the (simulated or real) climate system – what Lucarini also calls the "climatic machine". This would enable us to differentiate, for example, between (a) GCMs that perform well in regard of the reproduction of policy-relevant observational trends but do not get the underlying causal mechanisms right and (b) GCMs with a fairly good representation of key climate processes but a poor performance in terms of policy-relevant variables.

As a philosopher, I cannot judge whether Lucarini's proposal is suited to capture some key climate processes. So that is something I take for granted in the following discussion.

Lucarini's proposal raises the interesting question how the process-oriented metrics relate to the traditional ones based on fields of practical interest. I take it that Lucarini does not mean to replace traditional metrics by one or several process-oriented ones. Still, I see a bunch of questions that deserve further discussion:

– Should the process-oriented metrics be the primary indicator for the reliability of the predictions of a GCM? So, e.g., does the empirical inadequacy of a GCM in terms of a process-oriented metric undermine the credibility of its policy-relevant predictions, even if the model performs well in terms of those policy-relevant variables?
– What are the underlying assumptions that justify the expectation that the improvement of models in terms of process-oriented metrics leads, in the long run, to more accurate predictions in terms of policy-relevant variables?

These points inevitably lead us back to the question what at all one may infer from a good or bad performance of a GCM in terms of the relevant variables; that is back to the second question stated above.

6 Interpreting multi-model ensembles

Every GCM has false empirical implications. This holds for policy-relevant implications as well as, I assume, for process oriented metrics. According to a falsificationist methodology, all GCMs would have to be rejected. The ensemble of GCMs the IPCC relies on can thus not be understood as comprising all models not yet falsified. Falsificationism is of no help to understand the status of GCMs and the way they are assessed given their empirical performance.

Lucarini asserts that the model ensembles must not be interpreted probabilistically, either. (Yet, he seems to be ambiguous on this point, claiming in the concluding section that climate results "must be plural and stated in probabilistic terms".) I wholeheartedly agree that the empirical implications of GCMs, as of today, cannot be used to derive a probabilistic interpretation of the model ensemble (see also Betz 2007; Parker 2010). Let me briefly sketch why: (1.) Assigning probabilities to different GCMs only makes sense in a subjectivist interpretation of probabilities. (2.) Climate data does not significantly constrain the posteriors, which still depend crucially on the prior probabilities. (3.) These priors are really arbitrary, because climate scientists do not possess sufficient tacit knowledge (of 21st century climate change) to justifiably constrain the priors. Probabilistic studies in climatology rely on arbitrary, typically uniform priors. In particular, the catch-all hypothesis that a model not yet devised provides a correct analysis, is typically assigned the value zero.

This said, what does an ensemble of GCMs tell us? A group of climatologists from the Hadley Centre, who basically share the above diagnosis, have put forward an interesting proposal. In an article published in 2007, Stainforth et al. suggest:

> Today's ensembles give us a lower bound on the maximum range of uncertainty. (Stainforth, Allen et al. 2007, p. 2156)

So, in other words, whatever happens in a model simulation might actually happen in the future. Still, the future evolution of the climate system might also follow a dynamic that is not predicted by any GCM yet. That is, the range of possible evolutions of the climate system, given our current understanding, comprises *at least* the predictions of the model ensemble.

This is arguably a very modest interpretation of the epistemic status of GCMs. I would say, however, that this is the correct interpretation.

One might wonder whether, according to this interpretation of model ensembles, climate models are assessed in terms of their empirical implications at all. Is not the methodological outlook of Stainforth et al. flawed, or at least incomplete, as long as it does not explain how GCMs are evaluated on the basis of relevant climate data? We can address this challenge as follows: The empirical data regarding relevant climate variables already enters the process of constructing climate models. Model versions that perform definitely very poorly are excluded by so-called tuning. It is the calibration of model parameters that makes sure that only GCMs with a comparatively high empirical adequacy enter the model ensemble. And maybe it is exactly here, in the calibration process, where Lucarini's process-oriented metrics might have a major role to play. But that remains a further open question.

7 Improving our epistemic situation

As a final remark, I would like to draw our attention to the question how to construe scientific advancement given the specific interpretation of GCMs. Obviously, climate scientists should keep on trying to establish reliable and justified probability forecast, and succeeding in doing so would count as a major scientific breakthrough. But provided these attempts fail and the model ensemble remains merely a lower bound on the range of uncertainty, what sorts of changes count as improvement of our epistemic situation? What does scientific progress, in such a situation, mean at all?

Counter-intuitively, progress might consist in widening the range of models and their predictions. This could be achieved through devising ever new models, for instance by systematically varying all uncertain model assumptions and including ever new, relevant processes into the models. And that is, partly, what happens in the climate community. According to a recent News Feature in *Nature*, some climate scientists expect that this process will lead to a significant extension of the span of climate predictions, proving the ranges of previous IPCC reports too narrow.

> It's very likely that the generation of models that will be assessed for the next IPCC report will have a wider spread of possible climate outcomes as we move into the future. (Jim Hurrell, National Center Atmospheric Research in Boulder, Colorado, quoted in (Heffernan 2010, p. 1014))

Given the interpretation of model ensembles by Stainforth et al., such an extension of the scenario range is nothing climate scientists should fear – but rather strive for.

References

Betz, G. (2007). Probabilities in Climate Policy Advice: A Critical Comment. *Climatic Change* 85(1–2). 1–9.

Betz, G. (2009). Underdetermination, Model-ensemble, and Surprises – On the Epistemology of Scenario-analysis in Climatology. *Journal for General Philosophy of Science* 40(1). 3–21.

Heffernan, O. (2010). The Climate Machine. *Nature* 463(7284). 1014–1016.

IPCC (2001). *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge: Cambridge University Press.

IPCC (2007). *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge; New York: Cambridge University Press.
Lenhard, J. & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Modern Physics* 41(3). 253–262.
Parker, W. S. (2006). Understanding Pluralism in Climate Modeling. *Foundations of Science* 11. 349–368.
Parker, W. S. (2010). Predicting weather and climate: Uncertainty, ensembles and probability. *Studies in History and Philosophy of Modern Physics* 41(3). 263–272.
Rahmstorf, S. & Schellnhuber, H. J. (2006). *Der Klimawandel*. München: C. H. Beck.
Stainforth, D. A., Allen, M. R., et al. (2007). Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society A – Mathematical Physical and Engineering Sciences* 365(1857). 2145–2161.

Jun.-Prof. Dr. Gregor Betz
Karlsruhe Institute of Technology (KIT)
Institute of Philosophy
Kaiserstraße 12
Geb. 20.12
76131 Karlsruhe
Germany
gregor.betz@kit.edu