Local High-Resolution Climate Projections: Why a Multi-Model Approach Doesn't Help



Plan

Primer: High Resolution Climate Projection

Part I – The Perils of Model Error

→ Laplace's Demon and the Adventures of His Apprentices, in Philosophy of Science 81(1), 2014, 31–59, with Seamus Bradley, Hailiang Du and Leonard A. Smith.

Part II – The Limits of Post-Processing

- → Short: The Myopia of Imperfect Climate Models: The Case of UKCP09, Philosophy of Science 80(5), 2013, 886–897, with David A. Stainforth and Leonard A. Smith.
- → Long: An Assessment of the Foundational Assumptions in High-Resolution Climate Projections: The Case of UKCP09. Under review.

Outlook: What next?

Plan

Primer: High Resolution Climate Projection

Part I – The Perils of Model Error

→ There is some recognition that models are not truthful reflections of their targets, but there does not seem to be an appreciation of the systematic problem and the extent to which it can affect predictive accuracy.

Part II – The Limits of Post-Processing → Post processing of model outputs with multi-model ensemble methods won't make these problems go away. Outlook: What next?

Part I: In Collaboration With



Part II: In Collaboration With



With Illustrations By





Climate Change Is Real



Climate Change Is Man-Made



IPCC AR5:

Models reproduce:

- "Observed continental-scale surface temperature patterns".
- "Trends over many decades".
- "More rapid warming since the mid-20th century".





One would like to know how the local climate changes because policy is made at the local level.

Make provisions:

- Adaption: flood walls, water provision, etc.
 - Mitigation: implement changes and ideally stop bad things from happening.

Concrete example: UKCP09

The United Kingdom Climate Impacts Program's UKCP09 project aims to answer questions about the local impact of global climate change by making high resolution forecasts of the local climate out to 2100.

The declared aim and purpose of UKCP09 is to provide decision-relevant forecasts, on which industry and policy makers can base their future plans.

The launch document says:

'The projections have been designed as input to the difficult choices that planners and other decision-makers will need to make, in sectors such as transport, healthcare, water-resources and coastal defences, to ensure that UK is adapting well to the changes in climate that have already begun and are likely to grow in future.' (Jenkins et al 2009, 9) The launch document says:

'The projections have been designed as input to the difficult choices that planners and other decision-makers will need to make, in sectors such as transport, healthcare, water-resources and coastal defences, to ensure that UK is adapting well to the changes in climate that have already begun and are likely to grow in future.' (Jenkins et al 2009, 9) The launch document says:

'The projections have been designed as input to the difficult choices that planners and other decision-makers will need to make, in sectors such as transport, healthcare, water-resources and coastal defences, to ensure that UK is adapting well to the changes in climate that have already begun and are likely to grow in future.' (Jenkins et al 2009, 9) In concrete terms:

Probabilistic predictions are given on a 25km grid for finely defined events such as

- Change in mean daily maximum temperature
- Changes in precipitation

It is projected, for instance, that under a medium emission scenario the probability for a 20-30% reduction in summer mean precipitation in central London in 2080 is 0.5



How are these Result Generated?

- 1. GCM: approx. 300 runs of HadCM3/ HadSM3.
- Post-process the GCM outputs and downscale to obtain local predictions on 25km grid.

How are these Result Generated?

Are these results decision-relevant?

Question

- 1. GCM: approx. 300 runs of HadCM3/ HadSM3.
- Post-process the GCM outputs and downscale to obtain local predictions on 25km grid.

GCM: two points matter:

1. Strong simplifications are made to construct the model. So we are faced with model error. 2. As a matter of fact the dynamics is nonlinear.



Central Question:

- (a) Are the outcomes of *nonlinear models with structural model error* trustworthy and reliable?
- (b) Can the outputs of *nonlinear models with structural model error* form the basis of responsible policy making?

Central Question:

- (a) Are the outcomes of *nonlinear models with structural model error* trustworthy and reliable?
- (b) Can the outputs of *nonlinear models with structural model error* form the basis of responsible policy making?

Preview: not without may qualifications





The Question

A dynamical model has **structural model error** (SME) if its time evolution is relevantly different from that of the target system, possibly due to simplifications and idealisations.

Question: what are the consequences of SME for a model's predictive capacity?

Take-Home Message - Part 1

have.

If **chaotic** models have even the slightest SME, their capacity to make meaningful forecasts is seriously compromised.

This has dramatic consequences for our ability to make the kind of forecasts about the future that policy makers would like to Attention: *not* the same old story.

So far chaos has been studied in connection with uncertainty about *initial conditions*.

We ask what happens if we are uncertain about the correct *model structure*.

These are completely different problems!



Butterfly effect: Error in initial conditions





Butterfly effect: Error in initial conditions

Hawkmoth Effect: Error in the model structure (equations)



Take-Home Message – Part 2

We can mitigate against the butterfly effect by making probabilistic predictions rather than point forecasts.

This route is foreclosed in the case of the hawkmoth effect: nothing can mitigate against that effect!

So structural model error and not uncertainty in the initial conditions is what truly limits predictive power.

Or: butterflies are pretty; hawkmoths are ugly.








A Primer on Models















Literally 10,000s of climate variables for the entire world















Locating the Issues

51









Locating the Issues





Meet Laplace's Demon

- 1. Unlimited computational power
- 2. Unlimited dynamical knowledge
- 3. Unlimited observational power

(Laplace 1814)



The Demon knows everything. Laplace: 'nothing would be uncertain and the future, as the past, would be present to [his] eyes'.

So the Demon's model of the world's climate would be trustworthy because it provides the full truth.

But what happens if we are less capable than the Demon?



Meet the Senior Apprentice



- 1. Unlimited computational power
- 2. Unlimited dynamical knowledge
- 3. No unlimited observational power

How could the limitation of not having unlimited observational power be overcome?



How could the limitation of not having unlimited observational power be overcome?



Generate probabilistic predictions by moving the initial probability distribution forward in time:







Distributions become <u>uninformative</u> as time passes, but they <u>do not become</u> <u>misleading</u>.

The Senior Apprentice realises that this is the limitation that she has to accept.

It is the price to pay for not having unlimited observational power.

Or: butterflies are pretty; hawkmoths are ugly.



Meet the Freshman Apprentice

- 1. Unlimited computational power
- 2. No unlimited dynamical knowledge
- 3. No unlimited observational power



Recall: The Freshman can't formulate the exact dynamics of a system.

Reaction: Distortions and idealisations of all kind are acceptable as long as the resulting model is close enough to the truth.

This is the closeness-to-goodness link

Recall: The Freshman can't formulate the exact dynamics of a system.

Reaction: Distortions and idealisations of all kind are acceptable as long as the resulting model is close enough to the truth.

This is the closeness-to-goodness link

 \rightarrow This is a crucial part!












 $\widetilde{\rho}_{i,i} = 4\widetilde{\rho}_i(1-\widetilde{\rho}_i) \left[(1-\varepsilon) + \frac{4}{5} \varepsilon \left(\widetilde{\rho}_t^2 - \widetilde{\rho}_t + 1 \right) \right]$ $\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\varepsilon)(1-\tilde{\rho}_t) + \varepsilon \frac{16}{5}\tilde{\rho}_t(1-2\tilde{\rho}_t^2 + \tilde{\rho}_t^3)$ where $\varepsilon = 0.1$

The Apprentice remains defiant:



Mathematically:

 $\rho_{t+1} = 4\rho_t(1-\rho_t) + \text{small perturbation}$ $\tilde{\rho}_{t+1} = 4 \tilde{\rho}_t (1-\varepsilon)(1-\tilde{\rho}_t) + \varepsilon \frac{16}{5} \tilde{\rho}_t (1-2\tilde{\rho}_t^2 + \tilde{\rho}_t^3)$ 79

Mathematically:



Mathematically:













If you use your model to offer predictions you get it completely wrong!

- You regard things that never happen as very likely.
- You regard things that happen very often as unlikely.

Relative Entropy of 2048 initial distributions (t=8)



Conclusion:

wrong!

Even though the model is very close to the truth, it provides ruinous predictions! Hence: If chaotic models have even the slightest model error, their capacity to make meaningful (and policy relevant!) probabilistic forecasts is lost.

The closeness-to-goodness link is

Written Version:

Consequences of this for casino/insurance scenarios are disastrous.



The failure of the closeness-to-goodness link gives raise to the *hawkmoth effect*: the smallest deviation in model structure leads to completely different results, both for deterministic *and* probabilistic forecasts.



Or: butterflies are pretty; hawkmoths are ugly.





Fact: HadCM3 involves strong idealising assumptions

 \rightarrow It has structural model error.

UKCP09 acknowledges the presence of model error and suggests a way of dealing with it.

The message is that the uncertainties due to SME can be estimated and taken into account in projections.

UKCP09 do so with a complex computational scheme.

- \rightarrow "Long paper" for details.
- → Here focus only on the crucial assumptions.

Introduce a so-called discrepancy term:



The discrepancy

'measures the difference between the climate model and the *real climate* [...]. Such differences could arise from processes which are entirely missing from the climate model, or from fundamental deficiencies in the representation of processes which are included [...]' (Sexton et al, 2012, 2515, emphasis added) The discrepancy

'measures the difference between the climate model and the *real climate* [...]. Such differences could arise from processes which are entirely missing from the climate model or from fundamental deficiencies in the representation of processes which are included [...]' (Sexton et al, 2012, 2515, emphasis added)

The discrepancy

'measures the difference between the climate model and the *real climate* [...]. Such differences could arise from processes which are entirely missing from the climate model, or from fundamental deficiencies in the representation of processes which are included [...]'

(Sexton et al, 2012, 2515, emphasis added)

Therefore, the discrepancy term tells us

'what the model output would be if all the inadequacies in the climate model were removed, without prior knowledge of the observed outcome' (Sexton *et al.*, 2012, 2515). Therefore, the discrepancy term tells us

Recall

'what the model output would be if all the inadequacies in the climate model were removed, without prior knowledge of the observed outcome' (Sexton *et al.*, 2012, 2515).

 $c = \varphi(x_0, \alpha^*) + d$

Calculate d and add it to model outputs.



A different route:

2. Informativeness

Assume Gaussianity: *d* is Gaussian

→ Determine: mean and the covariance matrix of the distribution.

Two assumptions needed to do so:

1. Proxy

Not being omniscient, a proxy is introduced:

This is because:

'the effects of structural differences between models can be assumed to provide reasonable *a priori* estimates of possible structural differences between HadSM3 and the real world.' (Murphy *et al.* 2010, 64)

Therefore:

Discrepancy term: 'an appropriate means of quantifying uncertainties in projected future changes' (ibid, 66)


Specifically:

<u>Multi Model Ensemble</u> with 12 models. Main steps:

- Determine the best HadSM3 analogue for each model in the ensemble.
- For each model, calculate the error (the difference between the two model outputs).
 From these the mean and the covariance matrix of are determined.

The Informativeness Assumption

This is the assumption that

'that the climate model *is* informative about the real system' (Sexton *et al* 2012, 2521).



The Informativeness Assumption

This is the assumption that

'that the climate model *is* informative about the real system' (Sexton *et al* 2012, 2521).

And for the best input parameter α^* :

'[α*] is not just a 'statistical parameter', devoid of meaning: it derives its meaning form the physics in the climate model being approximately the same as the physics in the climate.' (Rougier 2007, 253)

Question:

How good are these assumptions?

 \rightarrow Take Gaussianity for granted.

 \rightarrow Scrutinise proxy and informativeness

Scrutinising the Proxy Assumption

First argument in support:

'Indeed, the multimodel ensemble mean has been shown to be a more skilful representation of the present-day climate than any individual member' (Sexton et al 2012, 2526)

Scrutinising the Proxy Assumption

First argument in support:

'Indeed, the multimodel ensemble mean has been shown to be a more skilful representation of the present-day climate than any individual member' (Sexton et al 2012, 2526)

But this is sleight of hand:

Is "more skilful" close to being "skilful"?
 No evidence is given that this is the case.

Second argument in support:

'the structural errors in different models can be taken to be independent' (Sexton et al 2012, 2526.)

This is needed to avoid that models in the ensembles have systematic bias.

Second argument in support:

Jun,

. . .

'the structural errors in different models can be taken to be independent' (Sexton et al 2012, 2526.)

This is needed to avoid that models in the ensembles have systematic bias.

But: Models aren't independent, and common errors are widely acknowledged (Knutti, Parker, Bishop and Abramowitz, Further worry:

- Current MME's 'ensembles of opportunity', grouping together existing models.
- They are not designed to systematically explore all possibilities.
- There could be vast classes of models that produce entirely different results.

Notice:

- The IPCC acknowledge this limitation and downgrade the assessed likelihood of ensemble-derived confidence intervals.
- Example: the 5-95% range of model results for GMT change in 2100, under forcing scenario RCP8.5 (2.6 to 4.8 degrees) is not deemed "very likely" (90%) chance), which would correspond to a direct use of model frequencies as probabilities; instead, it is deemed only "likely" (66% chance).

Hence:

- Ensemble information is used, but supplemented with expert judgement about the chance that models are misinformative.
- In effect, 24% of the probability mass has been reassigned in an undetermined manner, which we might interpret as a 1-in-4 chance that something occurs which the models are incapable of simulating.
 (NB: This is for GMT!)

Conclusion:

- For these reasons, the assumption that the use of an MME will accurately quantify the distance to our true target is unjustified.
- It produces a distribution that is more consistent with the diversity of current models, but which need not reflect the uncertainty in the true future climate.

- Worry: the distribution may simply be in the wrong place.
- Analogy: Trying to predict the true climate with structurally wrong models is like trying to predict the trajectory of Mercury with Newtonian models. These models will invariably make false projections for some lead time, and these errors cannot be removed by adding linear discrepancy term derived from other Newtonian models.

Scrutinising Informativeness

Recall that is the assumption that

'that the climate model *is* informative about the real system' (Sexton *et al* 2012, 2521).

And for the best input parameter α^* :

'[α*] is not just a 'statistical parameter', devoid of meaning: it derives its meaning form the physics in the climate model being approximately the same as the physics in the climate.' (Rougier 2007, 253)

Scrutinising Informativeness

Recall that is the assumption that

'that the climate model *is* informative about the real system' (Sexton *et al* 2012, 2521).

And for the best input parameter α^* :

'[α*] is not just a 'statistical parameter', devoid of meaning: it derives its meaning form the physics in the climate model being approximately the same as the physics in the climate.' (Rougier 2007, 253)

Scrutinising Informativeness

Recall that is the assumption that

'that the climate model *is* informative about the real system' (Sexton *et al* 2012, 2521).

And for the best input parameter α^* :

'[α*] is not just a 'statistical parameter', devoid of meaning: it derives its meaning form the physics in the climate model being approximately the same as the physics in the climate.' (Rougier 2007, 253)
 → Closeness-to-goodness link!

Statistics "backup": CPI (Murphy et al 2004)

- Take 32 climate variables (precipitation,...)
- For each variable there is a time series of past observations.
- Calculate mean and variance.
- Make 53 runs of HadCM3 retrodicting the values of the 32 variables.
- Calculate how many standard deviations the model is away from the observations.

CPI = average for the 32 variables of the difference between model and data

- Finding:
- For all 53 runs CPI is between 5 and 8.

 \rightarrow Is this really informative?

 Model runs for individual variables can be as much as much as 24 standard deviations away from observations. Further Assumptions in the scheme:

- Use of an Emulator
- Choice of a trapezoid prior probability distributions:
 - Principle of indifference
 - Robustness of posteriors
- Downscaling
- Initial condition uncertainty

Conclusion

- There is no evidence for interpreting UKCP09's projections as trustworthy information for quantitative decision support.
- NB: questioning the evidence for a result does not amount to proving it wrong
- The concern is that the premises of the argument do not warrant trust in the results.



Observations

- Better decisions could be made with better understanding of the scientific uncertainties even if they were presented in a less quantitative fashion.
- The detailed probabilistic projections might be expected to change substantially in future assessments, thus undermining the user communities trust in scientific outputs.

Think Different

For Science:

- Don't aim to reduce uncertainty.
- Instead better understand, classify and communicate uncertainty.
- For Policy:
- Renounce the first-predict-then-act rule.
 Decisions <u>can</u> be made under uncertainty.

Shifting Paradigm



Expert Elicitation



A Way Forward

- Better actionable information at local level.
- Cheaper and faster.
- Informs resilience planning.



Take Home Message

For the purpose of local decision support:

- Climate model outputs can be misleading.
- New models and faster computers will not make the problem go away.
- We should focus on understanding and assessing uncertainty.

This is best done using polling methods.





And if you use your model to offer bets (or insurance policies) on certain events, you are losing money!

Probability *p* on event *E*: p(E)Odds on *E*: $o(E) = 1/p \rightarrow pay-out if$ *E*occurs And if you use your model to offer bets (or insurance policies) on certain events, you are losing money!

Probability p on event E: p(E)Odds on E: $o(E) = 1/p \rightarrow pay-out if E occurs$ Example: coin p or heads is $\frac{1}{2}$. Odds on heads is 2. If you bet £1 on heads and head occurs you get £2 back.

"Lower Half" against "Upper Half



"Lower Half" against "Upper Half





Nine punters with £1000 each.

In every round they bet 10% of their wealth on events with probability in the interval:




Result:

- 7 out of the 9 punters make enormous gains!
- The casino runs up huge losses.

→ Insurance companies ...

But: is this just a bad "bad luck event"?







Feigenbaum's classical discussion:



Time series for different parameter values:



Time series for different parameter values:



Time series for different parameter values:







This is a study of **parameter variation**.

It provides information about what happens if we are uncertain about parameter values.

But: it provides <u>no</u> information about what happens when we are **uncertain about the model structure**.

What if the true equation is not exactly

 $\rho_{t+1} = \alpha \rho_t (1 - \rho_t)?$



Recall our conclusion: the closeness to goodness link is not an adequate means to deal with structural model error.

Why is this a *general* problem and not just a problem of our example?

There is an elaborate mathematical theory of structural stability:

Andronov and Pontrjagin, Peixoto, Palis, Smale, Mañé, Hayashi.

But:

Stability proofs are forthcoming only for two-dimensional flows!

But that is a very special kind of system!

In general the situation is more involved:



Axiom A: the system is uniformly hyperbolic.
Strong transversality condition: stable and unstable manifolds must intersect manifolds must intersect

Palis and Smale (1970) conjectured that a system is structurally stable iff it satisfies Axiom A and the strong tranversality condition.

Proofs:

Mañé (1988) for maps

Hayashi (1997) for flows.



What do Axiom A and the strong transversality condition mean for physical models?

162

11144

1 61

What do Axiom A and the strong transversality condition mean for physical models?

Physical models? What are you talking about?

163

1.1.1.4

But:

Smale (1966): structural stability is not generic in the class of diffeomorphisms on a manifold: the set of structurally stable systems is open but not dense.

Smith (2002) and Judd and Smith (2004): if the model's and the system's dynamics are not identical, then 'no state of the model has a trajectory consistent with observations of the system' (2004, 228). Minimal conclusion: shift of the onus of proof!

Those using non-linear models for predictive purposes owe us an argument that they are structurally stable, not *vice versa*!

