Model structure uncertainty

A matter of (Bayesian) belief?

16 March 2011 | Peter Janssen & Arthur Petersen
All Models Are Wrong… but anyway, how further?

Selecting an ‘appropriate’ model(structure)?
- Fitness for purpose
- Reflecting current knowledge on most important processes
- Being close to observations

Problematic aspects
- How to express appropriateness?
- What if knowledge base is weak?
- What if data has limited availability, quality and scope?

What uncertainty is involved in selected model(structure)s?
Model selection

- Multiple hypothesis testing
- Penalty Function-based model selection
  AIC, BIC, FIC, MDL
- Cross-validation methods
- Bayesian Model Averaging or Frequentist Model Averaging
- .....?

But how to deal with uncertainty involved in selected model(structure)s?

⇒ Some examples from Climate Change
GLOBAL CLIMATE CHANGE

Piling Up Uncertainties

The gospel of Thomas

In the face of uncertainty, your Bayesian belief comes as a rescue
Issue of model structure uncertainty (12 different AR4 models)

Figure 2.7: Changes (%) in summer (June-August) precipitation by the period 2071-2100 compared to 1961-1990, from 12 climate models, each of which took part in the IPCC AR4, all driven with the same SRES A2 emissions scenario.
MME: multi-model ensemble: CMIP-3 in IPCC-AR4

- **Model ensemble** (carefully selected; all plausible models)

  \[ Y_1 = F_1 (X^*_1), \quad Y_2 = F_2 (X^*_2), \quad \ldots, \quad Y_n = F_n (X^*_n) \]

  - Multi-model averages are shown;
  - Inter-model standard deviation as measure of uncertainty.

- But, what uncertainties are characterized by model ensemble, what is left out (e.g. surprises)?
- Is the sample representative for ‘true’ uncertainties involved?

  **Issue**: scope and relatedness of \( X^*_i \sim X^*_j \) and \( F_i \sim F_j \)
Does MME underestimate uncertainty?

No true model; collection of best guesses;

Mutual dependence (models sharing components/ideas and even sharing systematic and common errors)

Sample is neither random, nor systematic, nor complete
Fig. 2. (a) Absolute bias in 1970–99 average surface temperature from ERA-40, averaged across all CMIP3 models for DJF and JJA. (b) Same as in (a), but bias shown for the multimodel average. In some locations, biases from observations are reduced but the improvement by averaging is very heterogeneous.
Performance of models in this MME

- What *metrics* of performance to be used?
- Is it just to treat models equally?  
  Or should we apply a *weighting*, accounting for their ‘performance’, by confronting them with observed climate?

  E.g. evaluating posterior probability of model, given observations

  \[ P(M_j|Y_{obs}) \sim \exp[-1/2 \text{ BIC}(M_j)] \rightarrow \text{ cf. BMA} \]

- But, what does performance in the *past* tell about *future* performance?
Natural variability

Figure 2.2: The back line shows the observed England and Wales winter precipitation anomaly from 1950–2000, relative to the 1961–1990 average. The three coloured lines show projections of the same variable, from three experiments using the HadCM3 global model. Each experiment was driven with the same (UKCIP02 Medium-High) emissions scenario, but was started with different initial conditions. The differences between the three simulations are due to natural internal variability.
Natural variability

Figure 2.3: Maps of the change in winter precipitation averaged over the period 2071–2100, relative to 1961–1990, taken from the same three model experiments used in Figure 2.2 and described in the caption.
Three different emission scenarios

Seven different timeframes

25km grid, 16 admin regions, 23 river-basins and 9 marine regions

Variables and months

This is what users requested

Uncertainty including information from models other than HadCM3

UKCP09: Towards more complete probabilistic climate projections

David Sexton
Why we cannot be certain...

- **Internal variability** (initial condition uncertainty)
- **Modelling uncertainty**
  - Parametric uncertainty (land/atmosphere and ocean perturbed physics ensembles, ppe)
  - Structural uncertainty (multimodel ensembles)
  - Systematic errors common to all current climate models
- **Forcing uncertainty**
  - Different emission scenarios
  - Carbon cycle (perturbed physics ensembles, ppe)
  - Aerosol forcing (perturbed physics ensembles, ppe)
Production of UKCP09 predictions

Other models

Equilibrium PPE

Observations

Equilibrium PDF

Time-dependent PDF

Simple Climate Model

4 time-dependent Earth System PPEs (atmos, ocean, carbon, aerosol)

25km PDF

UKCP09

25km regional climate model

David Sexton
Some remarks and innovative aspects (UKCP09)

- **Probability** is considered as ‘Measure of the degree to which a particular outcome is consistent with the evidence’
- Use of ‘emulators’ (metamodels), replacing more computer-intensive models
- Explicit incorporation of *model discrepancy*, estimated from difference multi-model ensemble in relation to HadCM3.
- Explicit incorporation of uncertainty linked to *downscaling* and *timescaling*.
- Also accounting for uncertainties in forcing, due to *carbon cycle* and *aerosol* forcing.
- Honest in explicitly stating underlying assumptions
Assumptions explicitly stated 😊; realistic ? (sometimes 😞)

- That known sources of uncertainty not included in UKCP09 are not likely to contribute much extra uncertainty.
- That structural uncertainty across the current state of the art models is a good proxy for structural error.
- That models that simulate recent climate, and its recent trends well, are more accurate at simulating future climate.
- That single results from other global climate models are equally credible.
- That projected changes in climate are equally probable across a given 30 year time period.
- That local carbon cycle feedbacks are small compared to the impact of the carbon cycle via change in global temperature.
But… (caveats)

Probabilistic projections are *always conditional* on implied assumptions and used methods.

- How far should we go in this?
- How realistic is this sketch of uncertainty?
  - discrepancy assessed on basis of comparison with other models;
  - what about systematic errors common to all state-of-the-art climate models; missing processes, potential surprises, unk-unk?
- How robust, representative and relevant are the results?

- What about non-quantifiable aspects with respect to knowledge quality/underpinning?
Going beyond quantification of uncertainty: integral perspective

Foci and key issues in knowledge quality assessment (ref. 9)

<table>
<thead>
<tr>
<th>Foci</th>
<th>Key issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem framing</td>
<td>Other problem views; interwovenness with other problems; system boundaries; role of results in policy process; relation to previous assessments</td>
</tr>
<tr>
<td>Involvement of stakeholders</td>
<td>Identifying stakeholders; their views and roles; controversies; mode of involvement</td>
</tr>
<tr>
<td>Selection of indicators</td>
<td>Adequate backing for selection; alternative indicators; support for selection in science, society, and politics</td>
</tr>
<tr>
<td>Appraisal of knowledge base</td>
<td>Quality required; bottlenecks in available knowledge and methods; impact of bottlenecks on quality of results</td>
</tr>
<tr>
<td>Mapping and assessing relevant uncertainties</td>
<td>Identification and prioritisation of key uncertainties; choice of methods to assess these; assessing robustness of conclusions</td>
</tr>
<tr>
<td>Reporting uncertainty information</td>
<td>Context of reporting; robustness and clarity of main messages; policy implications of uncertainty; balanced and consistent representation in progressive disclosure of uncertainty information; traceability and adequate backing</td>
</tr>
</tbody>
</table>
CONSENSUS ON KNOWLEDGE

+ Structured problem
  + statistical uncertainty

- Moderately structured (consensus on goals) problem
  - methodological unreliability; recognized ignorance

CONSENSUS ON VALUES

+ Moderately structured (consensus on means) problem
  + value-ladenness

- Unstructured problem
  - recognized ignorance; methodological unreliability; value-ladenness

Model structure uncertainty
Funtowicz and Ravetz, Science for the Post Normal age, *Futures*, 1993

The agnostic gospel
<table>
<thead>
<tr>
<th>Location</th>
<th>Ecological, technological, economic, social and political representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert judgement</td>
<td>Narratives, storylines, advice</td>
</tr>
<tr>
<td>Model structure</td>
<td>Relations</td>
</tr>
<tr>
<td>Technical model</td>
<td>Software &amp; hardware implementation</td>
</tr>
<tr>
<td>Model parameters</td>
<td></td>
</tr>
<tr>
<td>Model inputs</td>
<td>Input data, driving forces, input scenarios</td>
</tr>
<tr>
<td>Data (in general sense)</td>
<td>Measurements, monitoring data, survey data</td>
</tr>
<tr>
<td>Outputs</td>
<td>Indicators, statements</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Location</th>
<th>Statistical uncertainty (range of chance)</th>
<th>Scenario uncertainty (range as ‘what if’ option)</th>
<th>Recognized ignorance</th>
<th>Knowledge-related uncertainty</th>
<th>Variability-related uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>Fair</td>
<td>Strong</td>
<td>Small</td>
<td>Strong</td>
<td>Large</td>
</tr>
<tr>
<td>Small</td>
<td>Fair</td>
<td>Strong</td>
<td>Small</td>
<td>Strong</td>
<td>Large</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Qualification of knowledge base (backing)</th>
<th>Value-ladeness of choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak</td>
<td>Fair</td>
</tr>
<tr>
<td>Strong</td>
<td>Small</td>
</tr>
<tr>
<td>Large</td>
<td>Medium</td>
</tr>
</tbody>
</table>

**Uncertainty Matrix**

**Level of uncertainty**
(from determinism, through probability and possibility, to ignorance)

**Nature of uncertainty**

**16 March 2011 | Peter Janssen & Arthur Petersen**
Model structure uncertainty
Graphical display of qualitative uncertainty

**Figure 4A**
Level of Confidence for Increasing Atmospheric GHG Concentrations Caused by Human Activities

**Figure 4B**
Level of Confidence that Climate Sensitivity Lies in 1.5-4.5 °C Range

**Figure 4C**
Level of Confidence for More Frequent Episodes of Non-coastal Flooding by 2050
Pedigree matrix for assumptions

<table>
<thead>
<tr>
<th>Criteria Score</th>
<th>Plausibility</th>
<th>Inter-subjectivity peers</th>
<th>Inter-subjectivity stakeholders</th>
<th>Choice space</th>
<th>Influence situational limitations</th>
<th>Sensitivity to view and interests of the analyst</th>
<th>Influence on results</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>the assumption is plausible</td>
<td>many would have made the same assumption</td>
<td>many would have made the same assumption</td>
<td>hardly alternative assumptions available</td>
<td>choice assumption hardly influenced</td>
<td>choice assumption hardly sensitive</td>
<td>the assumption has only local influence</td>
</tr>
<tr>
<td>1</td>
<td>the assumption is acceptable</td>
<td>several would have made the same assumption</td>
<td>several would have made the same assumption</td>
<td>limited choice from alternative assumptions</td>
<td>choice assumption moderately influenced</td>
<td>choice assumption moderately sensitive</td>
<td>the assumption greatly determines the results of the step</td>
</tr>
<tr>
<td>0</td>
<td>the assumptions is fictive or speculative</td>
<td>few would have made the same assumption</td>
<td>few would have made the same assumption</td>
<td>ample choice from alternative assumptions</td>
<td>totally different assumption had there not been limitations</td>
<td>choice assumption sensitive</td>
<td>the assumption greatly determines the results of the indicator</td>
</tr>
</tbody>
</table>
Conclusions (I)

1. Conditional character of probabilistic projections requires being clear on assumptions and potential consequences (e.g. robustness, things left out)

2. Room for further development in probabilistic uncertainty projections: how to deal decently with model ensembles, accounting for model discrepancies
   - Beyond the Bayesian paradigm → e.g. Dempster-Shafer
   - Second order uncertainty → imprecise probabilities

3. There is a role to be played for knowledge quality assessment, as complementary to more quantitative uncertainty assessment
Conclusions (II)

4. Recognizing ignorance often more important than characterizing statistical uncertainty

5. Communicate uncertainty in terms of societal/political risks
Some references

