Inferring model structural error from a Multi-Model Ensemble

Jonathan Rougier\textsuperscript{1}  Michael Goldstein\textsuperscript{2}  Leanna House\textsuperscript{2}

\textsuperscript{1}Department of Mathematics  
University of Bristol, UK

\textsuperscript{2}Department of Mathematical Sciences  
University of Durham, UK

IDAG Meeting, Boulder CO, 28-30 January 2008
Outline

1. The basic challenge
2. Statistical modelling
3. Updating calculations
4. Elicitation
5. Discussion
Our climate model

Suppose we have a climate model, with outputs

\[ f^{(0)} = (f_h^{(0)}, f_p^{(0)}) \]

where ‘\( h \)’ denotes ‘historical’ and ‘\( p \)’ denotes ‘predictive’. These outputs are likely to be collections of prognostic quantities, each one indexed by space and time. Consequently there is lots of structure in our judgements about \( f^{(0)} \), prior to observing it.
Our climate model

- Suppose we have a climate model, with outputs

\[ f^{(0)} = (f_h^{(0)}, f_p^{(0)}) \]

where ‘\( h \)’ denotes ‘historical’ and ‘\( p \)’ denotes ‘predictive’. These outputs are likely to be collections of prognostic quantities, each one indexed by space and time. Consequently there is lots of structure in our judgements about \( f^{(0)} \), prior to observing it.

- We would like to use these outputs in an inference about climate system behaviour \( y = (y_h, y_p) \).

The challenge we face is to quantify \( y - f^{(0)} \), the model discrepancy. If we imagine that our model has already been carefully tuned, then this discrepancy can be thought of as structural error.
Eliciting $y - f^{(0)}$ directly is very challenging. But suppose we also had

1. A collection of outputs from models similar to ours,
   \[ F \triangleq (f^{(1)}, \ldots, f^{(m)}) \]

2. Imprecise observations on $y_h$, say $z$.

How can we use these to help us to quantify our uncertainty about $y - f^{(0)}$? In other words, how can we statistically model the relationship between $f^{(0)}$, $y$, $F$, and $z$?
Our sources of quantitative information

- Eliciting \( y - f^{(0)} \) directly is very challenging. But suppose we also had
  1. A collection of outputs from models similar to ours,

  \[ F \triangleq (f^{(1)}, \ldots, f^{(m)}) \]

  2. Imprecise observations on \( y_h \), say \( z \).

- How can we use these to help us to quantify our uncertainty about \( y - f^{(0)} \)? In other words, how can we statistically model the relationship between \( f^{(0)}, y, F, \) and \( z \)?

- One catch: we would like to use the observations \( z \) directly in a subsequent analysis (e.g., to improve our predictions of \( y_f \)), so we would rather not ‘over-use’ \( z \) at this stage.
Second-order exchangeability

- Suppose we thought that each one of the model-outputs

\[ \{ f \} \triangleq \{ f^{(0)}, f^{(1)}, \ldots, f^{(m)} \} \]

were equally informative about \( y \) and \( z \), in the sense that if we had to pick any pair of models to represent \( y \) and \( z \), we would be indifferent between all possible pairs.
Second-order exchangeability

- Suppose we thought that each one of the model-outputs

\[ \{ f \} \triangleq \{ f^{(0)}, f^{(1)}, \ldots, f^{(m)} \} \]

were equally informative about \( y \) and \( z \), in the sense that if we had to pick any pair of models to represent \( y \) and \( z \), we would be indifferent between all possible pairs.

- Examples where we might agree with this judgement:
  - A Multi-Model Ensemble;
  - An Initial Condition Ensemble of coupled models;
  - A Stochastic Forcing Ensemble;
Second-order exchangeability

- Suppose we thought that each one of the model-outputs

\[ \{ f \} \triangleq \{ f^{(0)}, f^{(1)}, \ldots, f^{(m)} \} \]

were equally informative about \( y \) and \( z \), in the sense that if we had to pick any pair of models to represent \( y \) and \( z \), we would be indifferent between all possible pairs.

- Examples where we might agree with this judgement:
  - A Multi-Model Ensemble;
  - An Initial Condition Ensemble of coupled models;
  - A Stochastic Forcing Ensemble;

- And where we might disagree:
  - A Perturbed Physics Ensemble, unless we have forgotten the settings of the parameters for each evaluation.
The simplest representation of this judgement is that
1. The collection \( \{ f \} \) is second-order exchangeable, and
2. Both \( y \) and \( z \) respect exchangeability with \( \{ f \} \).

In this case,
1. \( E(f^{(j)}) = \mu^f \) for all \( j \);
2. \( \text{Var}(f^{(j)}) = \Sigma^f \) for all \( j \);
3. \( \text{Cov}(f^{(j)}, f^{(j')}) = \Gamma^f \), for all \( j \neq j' \);
4. \( \text{Cov}(f^{(j)}, y) = \Psi^y \) and \( \text{Cov}(f^{(j)}, z) = \Psi^z \), for all \( j \);

for appropriate choices of \( \mu^f, \Sigma^f, \Gamma^f, \Psi^y, \) and \( \Psi^z \).

Second-order exchangeability is a strong condition, but it is far weaker than full exchangeability, because it only concerns the mean and variance of the collection \( \{ f \} \).
Second-Order Representation Theorem

If \( \{f\} \) is second-order exchangeable then we can introduce additional uncertain quantities \( \mathcal{M}(f) \) and \( \mathcal{R}_j(f) \) such that

\[
f^{(j)} = \mathcal{M}(f) + \mathcal{R}_j(f) \quad j = 0, 1, \ldots, m.
\]

Properties:

1. \( \{\mathcal{R}\} \) is second-order exchangeable with
   
   - \( E(\mathcal{R}_j(f)) = 0 \);
   - \( \text{Cov}(\mathcal{M}(f), \mathcal{R}_j(f)) = 0 \);
   - \( \text{Cov}(\mathcal{R}_j(f), \mathcal{R}_{j'}(f)) = 0 \).

2. Hence
   
   - \( E(f^{(j)}) = E(\mathcal{M}(f)) \);
   - \( \text{Cov}(f^{(j)}, f^{(j')}) = \text{Var}(\mathcal{M}(f)), j \neq j' \).

3. If \( y \) and \( z \) respect exchangeability with \( \{f\} \) then
   
   - \( \text{Cov}(y, \mathcal{R}_j(f)) = 0, \text{Cov}(z, \mathcal{R}_j(f)) = 0 \).
Properties of the discrepancy

- Define the discrepancy of model $j$:

$$d^{(j)} \triangleq y - f^{(j)} \quad j = 0, 1, \ldots, m.$$ 

For simplicity we take $E(d^{(j)}) = 0$.

- $\{d\}$ is also second-order exchangeable, because $y$ respects exchangeability with $\{f\}$. It has the representation

$$d^{(j)} = M(d) + R_j(d) \quad j = 0, 1, \ldots, m$$

where, since $d^{(j)} \equiv y - M(f) - R_j(f)$,

- $M(d) = y - M(f)$, and $E(M(d)) = 0$;
- $R_j(d) = -R_j(f)$, so $\text{Var}(R_j(d)) = \text{Var}(R_j(f))$. 
The sample variance of \( \{f\} \) provides a simple point estimate for \( \text{Var}(\mathcal{R}_j(d)) \), since

\[
f(j) - \bar{f} = \mathcal{R}_j(f) - m^{-1} \sum_{j=1}^{m} \mathcal{R}_j(f)
\]

and, as \( m \) becomes large,

\[
m^{-1} \sum_{j=1}^{m} (f(j) - \bar{f})^2 \approx E((\mathcal{R}_j(f))^2) = \text{Var}(\mathcal{R}_j(f))
\]

Note that the sample variance estimates the variance of the residual, \( \mathcal{R}_j(f) \), because the sample mean absorbs the effect of the common component \( \mathcal{M}(f) \).

**Warning:** the sample variance is likely to be singular, which is unlikely to correspond to our judgements about \( \mathcal{R}_j(d) \).
Data for updating

- We use the *Bayes linear* approach to update $d^{(0)}$ using the observations $F$ and $z$. We restrict the updating information to

$$U \triangleq (z - f_h^{(1)}, \ldots, z - f_h^{(m)}) \equiv (u^{(1)}, \ldots, u^{(m)}).$$

This ‘under-uses’ $z$, since $z$ is not recoverable from $U$.

- Note that the columns of $U$ are second-order exchangeable, because $z$ respects exchangeability with $\{f\}$.

- We choose to model $z$ as

$$z \equiv y_h + e$$

for some observation error vector $e$, where $e$ is uncorrelated with other uncertain quantities. For simplicity we take $E(e) = 0$, and we specify the variance matrix $\Sigma^e$. 

Bayes linear sufficiency

- From our choices, \(d^{(0)}\) respects exchangeability with \(\{u\}\):

\[
\text{Cov}(u^{(j)}, d^{(0)}) = \text{Cov}(e + y_h - f^{(j)}_h, d^{(0)}) = \text{Cov}(d^{(j)}_h, d^{(0)}) = \text{Var}(\mathcal{M}(d))_{ha}
\]

where \(a = (h, p)\): this does not involve \(j\), as required.

- Then it can be proved that the sample mean

\[
\bar{u} \triangleq m^{-1} \sum_{j=1}^{m} u^{(j)}
\]

is Bayes linear sufficient for \(U\) when updating \(d^{(0)}\), which means we only have to update by the vector \(\bar{u}\), rather than the collection \(U\).
Updating the discrepancy of model 0

\[ \tilde{u} = e + M(d)_h + m^{-1} \sum_{j=1}^{m} R_j(d)_h \]

and these three quantities are uncorrelated, so

\[ \text{Var}(\tilde{u}) = \Sigma^e + \text{Var}(M(d))_h + m^{-1} \text{Var}(R_j(d))_h \]
\[ \text{Cov}(\tilde{u}, d^{(0)}) = \text{Cov}(\tilde{u}, M(d)) = \text{Var}(M(d))_{ha}. \]
Updating the discrepancy of model 0

\[ \bar{u} = e + \mathcal{M}(d)_h + m^{-1} \sum_{j=1}^{m} \mathcal{R}_j(d)_h \]

and these three quantities are uncorrelated, so

\[ \text{Var}(\bar{u}) = \Sigma^e + \text{Var}(\mathcal{M}(d))_h + m^{-1} \text{Var}(\mathcal{R}_j(d))_h \]
\[ \text{Cov}(\bar{u}, d^{(0)}) = \text{Cov}(\bar{u}, \mathcal{M}(d)) = \text{Var}(\mathcal{M}(d))_h a_h. \]

The BL updating equations

\[ E_U(d^{(0)}) = \text{Var}(\mathcal{M}(d))_{ah} (\text{Var}(\bar{u}))^{-1} \bar{u} \]
\[ \text{Var}_U(d^{(0)}) = \text{Var}(\mathcal{M}(d)) + \text{Var}(\mathcal{R}_j(d)) \]
\[ - \text{Var}(\mathcal{M}(d))_{ah} (\text{Var}(\bar{u}))^{-1} \text{Var}(\mathcal{M}(d))_{ha} \]
An interesting special case

Suppose that $\text{Var}(\mathcal{M}(d))_h$ dominates $\text{Var}(\bar{u})$, in the sense that

$$(\text{Var}(\bar{u}))^{-1} \approx (\text{Var}(\mathcal{M}(d))_h)^{-1}.$$  

This would happen if $\Sigma^e$ were small and $m$ were large. In this case the updating equations become

$$E_U(d^{(0)}) \approx \left( \begin{array}{c} \bar{u} \\ \text{Var}(\mathcal{M}(d))_{ph} (\text{Var}(\mathcal{M}(d))_h)^{-1} \bar{u} \end{array} \right)$$

$$\text{Var}_U(d^{(0)}) \approx \text{Var}(\mathcal{R}_j(d)) + \left( \begin{array}{cc} 0 & 0 \\ 0 & \text{Var}_h(\mathcal{M}(d)_p) \end{array} \right)$$

where

$$\text{Var}_h(\mathcal{M}(d)_p) \triangleq \text{Var}(\mathcal{M}(d))_{pp} - \text{Var}(\mathcal{M}(d))_{ph} (\text{Var}(\mathcal{M}(d))_{hh})^{-1} \text{Var}(\mathcal{M}(d))_{hp}$$

In this case, $\text{Var}(\mathcal{M}(d))$ is all we need to specify, since we can neglect $\Sigma^e$ and estimate $\text{Var}(\mathcal{R}_j(d))$ from $F$. 


Specifying $\text{Var}(\mathcal{M}(d))$

From our definitions and modelling, $\mathcal{M}(d) \equiv y - \mathcal{M}(f)$, so $\text{Var}(\mathcal{M}(d))$ is our quantification of the ‘common’ error across our ensemble. Thus it’s the limit of the variance of the mean discrepancy, $\bar{d}_m$:

$$\lim_{m \to \infty} \text{Var}(\bar{d}_m) = \text{Var}(\mathcal{M}(d)).$$
Specify $\text{Var}(\mathcal{M}(d))$

- From our definitions and modelling, $\mathcal{M}(d) \equiv y - \mathcal{M}(f)$, so $\text{Var}(\mathcal{M}(d))$ is our quantification of the ‘common’ error across our ensemble. Thus it’s the limit of the variance of the mean discrepancy, $\bar{d}_m$:

$$\lim_{m \to \infty} \text{Var}(\bar{d}_m) = \text{Var}(\mathcal{M}(d)).$$

- What would it mean to choose $\text{Var}(\mathcal{M}(d)) \propto \text{Var}(\mathcal{R}_j(d))$?

Basically,

When the individual models are in disagreement, then the models taken together are less reliable.

This is a common judgement for climate MMEs: the attraction of second-order exchangeability is that we can statistically model it in a simple way.
One simple framework

- The simplest framework would be

\[
\text{Var}(\mathcal{M}(d)) = \frac{\rho}{1 - \rho} \text{Var}(\mathcal{R}_j(d))
\]

for some specified scalar \( \rho \in [0, 1) \).
One simple framework

- The simplest framework would be
  \[ \text{Var}(\mathcal{M}(d)) = \frac{\rho}{1 - \rho} \text{Var}(\mathcal{R}_j(d)) \]
  for some specified scalar \( \rho \in [0, 1) \).

- This implies
  \[ \text{Corr}(d_i^{(j)}, d_{i'}^{(j')}) = \rho \text{Corr}(\mathcal{R}_j(d))_{ii'} \quad j \neq j', \]
  where \( \text{Corr}(\mathcal{R}_j(d)) \) is the correlation matrix of \( \mathcal{R}_j(d) \). Therefore
  \[ \text{Corr}(d_i^{(j)}, d_i^{(j')}) = \rho \quad j \neq j', \]
  i.e. the correlation between the same outputs in two different ensemble members is always \( \rho \), no matter what the outputs.
One simple framework

- The simplest framework would be

\[ \text{Var}(\mathcal{M}(d)) = \frac{\rho}{1 - \rho} \text{Var}(\mathcal{R}_j(d)) \]

for some specified scalar \( \rho \in [0, 1) \).

- This implies

\[ \text{Corr}(d_i^{(j)}, d_i^{(j')}) = \rho \text{Corr}(\mathcal{R}_j(d))_{ii'} \quad j \neq j', \]

where \( \text{Corr}(\mathcal{R}_j(d)) \) is the correlation matrix of \( \mathcal{R}_j(d) \). Therefore

\[ \text{Corr}(d_i^{(j)}, d_i^{(j')}) = \rho \quad j \neq j', \]

i.e. the correlation between the same outputs in two different ensemble members is always \( \rho \), no matter what the outputs.

- **Not ideal:** may want different correlations for different types, and also \( \text{Var}(\mathcal{M}(d))_i > 0 \) when \( \text{Var}(\mathcal{R}_j(d))_i = 0 \).
A more flexible approach

The generalisation is

\[ \text{Var}(\mathcal{M}(d)) = A + Q\text{Var}(\mathcal{R}_j(d)) Q^T \]

for SNND matrix \( A \).
A more flexible approach

- The generalisation is

\[
\text{Var}(\mathcal{M}(d)) = A + Q \text{Var}(\mathcal{R}_j(d)) Q^T
\]

for SNND matrix \(A\).

- Most likely, \(A\) and \(Q\) will be diagonal. In this case, we can parameterise as

\[
Q_{ii'} = \delta_{ii'} \sqrt{\frac{\rho_i}{1 - \rho_i}} \quad \text{and} \quad A_{ii'} = \delta_{ii'} \frac{a_i}{\sqrt{(1 - \rho_i)(1 - \rho_{i'})}}
\]

where \(\delta_{ii'}\) is the Kronecker delta, and then

\[
\text{Corr}(d_i^{(j)}, d_i^{(j')}) = \frac{a_i \delta_{ii'} + \sqrt{\rho_i \rho_{i'}} \text{Var}(\mathcal{R}_j(d))_{ii'}}{\sqrt{a_i + \text{Var}(\mathcal{R}_j(d))_i} \sqrt{a_{i'} + \text{Var}(\mathcal{R}_j(d))_{i'}}}
\]

for \(j \neq j'\).
A more flexible approach (cont)

- One special case is where we partition the outputs, and set the same values of $a_i$ and $\rho_i$ for every output in the same component of the partition. The obvious partition in climate models is by type.

\[ \text{Simple elicitation} \]

- Fix the partition (e.g., by type). For each component of the partition, and before observing $\text{Var}(R_j(d_i))$, 
  1. Set $\rho_i$ as though $\text{Var}(R_j(d_i))$ were substantial, considering the correlation;
  2. Set $a_i$ using $\rho_i$ and $\text{Var}(R_j(d_i))$, considering both the implied marginal standard deviations, and the correlations.
A more flexible approach (cont)

- One special case is where we partition the outputs, and set the same values of $a_i$ and $\rho_i$ for every output in the same component of the partition. The obvious partition in climate models is by type.

- For $i$ and $i'$ in the same component of the partition and $a_i$ small relative to both $\text{Var}(\mathcal{R}_j(d))_i$ and $\text{Var}(\mathcal{R}_j(d))_{i'}$,

  $$\text{Corr}(d_i^{(j)}, d_{i'}^{(j')}) \approx \rho_i \text{Corr}(\mathcal{R}_j(d))_{ii'}, \quad j \neq j'.$$
A more flexible approach (cont)

- One special case is where we partition the outputs, and set the same values of \(a_i\) and \(\rho_i\) for every output in the same component of the partition. The obvious partition in climate models is by type.

- For \(i\) and \(i'\) in the same component of the partition and \(a_i\) small relative to both \(\text{Var}(R_j(d))_i\) and \(\text{Var}(R_j(d))_{i'}\),

\[
\text{Corr}(d_i^{(j)}, d_{i'}^{(j')}) \approx \rho_i \text{Corr}(R_j(d))_{ii',} \quad j \neq j'.
\]

Simple elicitation

Fix the partition (e.g., by type). For each component of the partition, and before observing \(\text{Var}(R_j(d))\),

1. Set \(\rho_i\) as though \(\text{Var}(R_j(d))_i\) were substantial, considering the correlation;

2. Set \(a_i\) using \(\rho_i\) and \(\text{Var}(R_j(d))_i\), considering both the implied marginal standard deviations, and the correlations.
The second-order exchangeability approach outlined here is likely to be sufficiently general for judgements about climate models. The main obligation is to think carefully about whether all the models in the collection are really equally useful.
Discussion

- The second-order exchangeability approach outlined here is likely to be sufficiently general for judgements about climate models. The main obligation is to think carefully about whether all the models in the collection are really equally useful.

- The specification of $\text{Var}(M(d))$ and $\text{Var}(R_j(d))$ is perfectly general (subject to these being valid variance matrices). The elicitation approach outlined here, which is relatively straightforward, is advocated only if judgements are quite vague, and time is pressing.
The second-order exchangeability approach outlined here is likely to be sufficiently general for judgements about climate models. The main obligation is to think carefully about whether all the models in the collection are really equally useful.

The specification of $\text{Var}(M(d))$ and $\text{Var}(R_j(d))$ is perfectly general (subject to these being valid variance matrices). The elicitation approach outlined here, which is relatively straightforward, is advocated only if judgements are quite vague, and time is pressing.

The outcome is a mean vector and a variance matrix for the discrepancy of model 0. This can be used to turn the point value $f^{(0)}$ into a statistical prediction for $y$. It can also be used in calibration, to learn about the values of model 0’s parameters.
Next step

- There is more information in the data $[F, z]$ than we have used. In particular, the sample variance of $U$ has not been used, and this ought to be highly informative for the variance matrices of $\mathcal{M}(d)$ and $\mathcal{R}_j(d)$.

- The next step is to do a two-stage analysis, where in the first stage we learn about the variances using the sample variance of $U$, and then, in the second stage, plug the updated variances into the analysis outlined here.

- This is a fourth-order analysis (learning about variances), which requires additional second-order exchangeability judgements. It is more complicated, and does not scale well without simplifications, because the variance of a variance is a four-tensor.
Further reading

- The Bayes linear approach is outlined in Goldstein (1999), and described in detail in Goldstein and Wooff (2007), where all the pertinent results can be found. Its use in Computer Experiments is described in Craig et al. (1996, 1997), Craig et al. (2001), Goldstein and Rougier (2006).

- The standard statistical framework for combining model evaluations with system behaviour and observations is described in Goldstein and Rougier (2004) and Rougier (2007, climate example). It is critiqued and extended in Goldstein and Rougier (2007).

- A thorough and perceptive analysis of the use of MMEs in climate is provided by Tebaldi and Knutti (2007).
References


