

Uncertainty and Climate: A Statistician's View

Jonathan Rougier (A statistician)

Maths Department

Durham University

`J.C.Rougier@durham.ac.uk`



Abstract

Policy-makers and other stakeholders are increasingly asking for *probabilistic* predictions for future climate. In order to provide these we need to understand how climate predictions are made, and to identify and quantify the key sources of uncertainty. In practice it is also necessary to address some misconceptions about probability, and about ‘objectivity’: a concept which is ill-conceived in this context.

What does this mean:

“The chance of the Gulf Stream, which brings warm waters around the British Isles, being halted, sending temperatures plummeting by more than 5C, is now more than 50%, a scientific conference on climate change was told yesterday.”

Paul Brown, Guardian, Wednesday February 2 2005 [\[source\]](#)

Climate prediction

- Climate prediction is an example of *Model-Based Inference for Complex Systems* (M-BICS). Statisticians have been studying this subject for about 25 years, but are only just beginning to tackle the really challenging problems.
- ‘Challenge’ in M-BICS applications can broadly be classified by (1) size of model; (2) model inadequacy; and (3) policy impact. Climate prediction scores highly on all three.
- Our initial contribution to Climate Science is to clarify the nature of the inference, and to highlight some of the undesirable features of current practice.
- Sources of uncertainty:
 1. Measurement errors in climate observations;
 2. The relationship between the climate model and the climate itself;
 3. Budget constraints limiting the number of model evaluations.
- None of these is easy to quantify, but (1) and (3) are at least well-understood by statisticians; (2) requires some deep thinking . . .

So what's going on?

● Basic quantities:

$y \triangleq (y_h, y_f)$ Operationally-defined climate quantities; partitioned into 'historical' and 'future';

z Measurements made on y_h ;

$e \triangleq z - y_h$ 'Measurement error'.

We can write $z \equiv y_h + e$, and a distribution for (y, e) induces a distribution for (y, z) .

So what's going on?

- Basic quantities:

$y \triangleq (y_h, y_f)$ Operationally-defined climate quantities; partitioned into 'historical' and 'future';

z Measurements made on y_h ;

$e \triangleq z - y_h$ 'Measurement error'.

We can write $z \equiv y_h + e$, and a distribution for (y, e) induces a distribution for (y, z) .

- We make *structural* choices to define the framework within which we make our inference, and *tractability* choices to simplify the resulting elicitation and computation:

$$e \perp\!\!\!\perp y \tag{S1}$$

$$e \sim \text{Gau}(\mathbf{0}, \Sigma^e) \quad \Sigma^e \text{ specified;} \tag{T1}$$

Σ^e , the *measurement error variance*, has an interesting hierarchical structure that depends on type of instrument, instrument ID, and proximity (e.g., spatio-temporal).

So what's going on? (cont)

- To predict y_f we condition on the observed value of the climate data, $z = \tilde{z}$:

$$\begin{aligned}\Pr(y \mid z = \tilde{z}) &= c \Pr(z = \tilde{z} \mid y) \Pr(y) && \text{Bayes's theorem} \\ &= c \Pr(e = y_h - \tilde{z} \mid y) \Pr(y) && \text{by definition} \\ &= c \Pr(e = y_h - \tilde{z}) \Pr(y) && \text{using (S1)} \\ &= c \varphi(y_h - \tilde{z}; \mathbf{0}, \Sigma^e) \Pr(y) && \text{using (T1)}\end{aligned}$$

where $c \triangleq \Pr(z = \tilde{z})^{-1}$, and $\varphi(\cdot)$ is the gaussian density function with given mean and variance.

So what's going on? (cont)

- To predict y_f we condition on the observed value of the climate data, $z = \tilde{z}$:

$$\begin{aligned}\Pr(y \mid z = \tilde{z}) &= c \Pr(z = \tilde{z} \mid y) \Pr(y) && \text{Bayes's theorem} \\ &= c \Pr(e = y_h - \tilde{z} \mid y) \Pr(y) && \text{by definition} \\ &= c \Pr(e = y_h - \tilde{z}) \Pr(y) && \text{using (S1)} \\ &= c \varphi(y_h - \tilde{z}; \mathbf{0}, \Sigma^e) \Pr(y) && \text{using (T1)}\end{aligned}$$

where $c \triangleq \Pr(z = \tilde{z})^{-1}$, and $\varphi(\cdot)$ is the gaussian density function with given mean and variance.

- *In this calculation we need to specify $\Pr(y)$ in order to predict y_f* ; in the optimistic case where we have lots of climate data we *might* be able to get away with only needing to specify $\Pr(y_f \mid y_h)$.
- But y has many thousands of components, indexed by space, time and type, with a complicated interdependency reflecting the climate state vector's evolution in space and time according to 'laws of nature'.

The inferential role of the climate model

- *A climate model is a means of inducing a distribution for y by elaboration from a simpler set of uncertain primitive quantities.*
- The climate model is a deterministic function

$$x \rightarrow g(x)$$

where the model-inputs $x \in \mathcal{X}$ are a collection of:

1. Parameters in the underlying mathematical model;
2. Initial value for the climate state vector;
3. Forcing functions (typically boundary conditions).

The image of \mathcal{X} is $\mathcal{G} \triangleq \{g(x), x \in \mathcal{X}\}$.

- If we want to link our climate model with the climate itself, we must take account of our uncertainty about the appropriate setting for the model-inputs. Even where these model-inputs are operationally defined, it does not follow that the system value is the right value in the model.

A hierarchy of linkages

- What are we prepared to believe about the relationship between our model $g(\cdot)$ and the climate it purports to represent, y ?
 1. *Wow!* There exists a 'correct' value $\hat{x} \in \mathcal{X}$ such that $y = g(\hat{x})$, and \hat{x} is known.
 2. *Phew!* There exists a 'correct' \hat{x} as above, but its value is uncertain.
- These are generally unreasonable; any point in \mathcal{G} is a simplification of climate in the same way that $g(\cdot)$ is a simplification of the processes that are involved in climate. It's hard to believe that $y \in \mathcal{G}$.

A hierarchy of linkages

- What are we prepared to believe about the relationship between our model $g(\cdot)$ and the climate it purports to represent, y ?
 1. *Wow!* There exists a 'correct' value $\hat{x} \in \mathcal{X}$ such that $y = g(\hat{x})$, and \hat{x} is known.
 2. *Phew!* There exists a 'correct' \hat{x} as above, but its value is uncertain.
- These are generally unreasonable; any point in \mathcal{G} is a simplification of climate in the same way that $g(\cdot)$ is a simplification of the processes that are involved in climate. It's hard to believe that $y \in \mathcal{G}$.
- We can believe only that a 'best' value x^* exists (uncertain), and we define the *model discrepancy*

$$\epsilon^* \triangleq y - g(x^*).$$

We do not believe or assert that $\epsilon^* = 0$ a.s..

- Now we have $y \equiv g(x^*) + \epsilon^*$. A distribution for (x^*, ϵ^*) , when combined with the model $g(\cdot)$, induces a distribution for y .

Digression: why aren't climate models better?

- The underlying model comprises (1) differential equations that conserve quantities through space and time; (2) equations of state that constrain how these quantities appear at any instance of space and time; and (3) coupling equations to join all the sub-models together.

Our understanding of both (2) and (3) leaves a lot to be desired.

- Only modelling a subset of the earth introduces tricky boundary conditions. It is much better to model the whole of the earth, which we can treat as a closed system subject to solar forcing. But computing power currently limits the resolution of our finite difference solvers. The result is a mesh-size that is ~ 100 km with current technology.

Features smaller than this would be missed, and unfortunately they are important (e.g., turbulence, which is involved in vertical mixing). They get put back in 'by hand'.

The 'best input' treatment of model inadequacy

- Time for some more *structural* and *tractability* choices:

$$x^* \perp\!\!\!\perp \epsilon^* \perp\!\!\!\perp e \quad (\text{S1}')$$

$$\epsilon^* \sim \text{Gau}(\mathbf{0}, \Sigma^\epsilon) \quad \Sigma^\epsilon \text{ specified} \quad (\text{T2})$$

nb: (S1') \Rightarrow (S1), because given $g(\cdot)$, y is a deterministic function of x^* and ϵ^* .

- The diagonal components of Σ^ϵ represent model-inadequacy. The off-diagonal components represent systematic patterns in the model-error, and would typically have a spatio-temporal structure within component-type.
- There are situations where we would like to relate uncertainty about ϵ^* to x^* . A generalisation of the link between $g(\cdot)$ and y of the form

$$y = g(x^*) + Q(x^*)^T \epsilon^* \quad Q(\cdot) \text{ specified,}$$

which preserves (S1'), can be useful in this case.

Model validation

- *Model validation*: Is the model about as good as we think it is?
- Our primitives are $g(\cdot)$, $\Pr(x^*)$, Σ^ϵ and Σ^e .
- The natural diagnostic would be to compare the data $z = \tilde{z}$ with its predictive distribution:

$$\begin{aligned}\Pr(z) &= \int \Pr(z \mid x^*) \Pr(x^*) dx^* && \text{'law of total probability'} \\ &= \int \Pr(\epsilon_h^* + e = z - g_h(x^*) \mid x^*) \Pr(x^*) dx^* && \text{by definition} \\ &= \int \Pr(\epsilon_h^* + e = z - g_h(x^*)) \Pr(x^*) dx^* && \text{using (S1')} \\ &= \int \varphi(z - g_h(x^*); \mathbf{0}, \Sigma_{hh}^\epsilon + \Sigma^e) \Pr(x^*) dx^* && \text{using (S1'), (T1) and (T2).}\end{aligned}$$

- Diagnostics would be based on $\Pr(z \leq \tilde{z})$. Physically-meaningful subsets of z could also be used to 'drill down' to identify systematic problems.

Model Calibration

● *Model calibration* is learning about x^* using $z = \tilde{z}$:

$$\Pr(x^* \mid z = \tilde{z}) = c \Pr(z = \tilde{z} \mid x^*) \Pr(x^*)$$

Bayes's theorem

$$= c \Pr(\epsilon_h^* + e = \tilde{z} - g_h(x^*) \mid x^*) \Pr(x^*)$$

by definition

$$= c \varphi(\tilde{z} - g_h(x^*); \mathbf{0}, \Sigma_{hh}^\epsilon + \Sigma^e) \Pr(x^*)$$

using (S1'), (T1) and (T2)

where $c \triangleq \Pr(z = \tilde{z})^{-1}$ as before.

Model Calibration

- *Model calibration* is learning about x^* using $z = \tilde{z}$:

$$\begin{aligned}\Pr(x^* \mid z = \tilde{z}) &= c \Pr(z = \tilde{z} \mid x^*) \Pr(x^*) && \text{Bayes's theorem} \\ &= c \Pr(\epsilon_h^* + e = \tilde{z} - g_h(x^*) \mid x^*) \Pr(x^*) && \text{by definition} \\ &= c \varphi(\tilde{z} - g_h(x^*); \mathbf{0}, \Sigma_{hh}^\epsilon + \Sigma^e) \Pr(x^*) && \text{using (S1'), (T1) and (T2)}\end{aligned}$$

where $c \triangleq \Pr(z = \tilde{z})^{-1}$ as before.

- Model calibration should not be confused with *tuning*: trying to find an $x \in \mathcal{X}$ that minimises some metric defined on $\tilde{z} - g_h(x)$. All big computer models are tuned in this way. But:
 - We have to decide (1) what metric to use; and (2) what to do in the case of multiple 'good' choices, or no 'good' choices.
 - The probabilistic approach answers: (1) the metric may be inferred from our judgements, e.g., about ϵ_h^* and e (see above); and (2) we average over candidate values for x^* giving more weight to those that fit better (see below).

Calibrated Prediction

- *Calibrated prediction* is learning about y_f using $z = \tilde{z}$:

$$\begin{aligned}\Pr(y_f \mid z = \tilde{z}) &= \int \Pr(y_f \mid x^*, z = \tilde{z}) \Pr(x^* \mid z = \tilde{z}) dx^* && \text{'law of total probability'} \\ &= \int \varphi(y_f; \mu_{f|z}(x^*), \Sigma_{f|z}) \Pr(x^* \mid z = \tilde{z}) dx^* && \text{see below}\end{aligned}$$

where $\Pr(x^* \mid z = \tilde{z})$ comes from the calibration distribution and

$$\begin{aligned}\mu_{f|z}(x) &\triangleq g_f(x) + \Sigma_{fh}^\epsilon (\Sigma_{hh}^\epsilon + \Sigma^e)^{-1} (\tilde{z} - g_h(x)), \\ \Sigma_{f|z} &\triangleq \Sigma_{ff}^\epsilon - \Sigma_{fh}^\epsilon (\Sigma_{hh}^\epsilon + \Sigma^e)^{-1} \Sigma_{hf}^\epsilon,\end{aligned}$$

based on $(y, z) \mid x^*$ being jointly gaussian, according to (S1'), (T1) and (T2).

- Relative to $\Pr(y)$, most of the reduction in uncertainty comes from concentrating the distribution for x^* ; some also comes from learning about ϵ_f^* , provided that $\Sigma_{fh}^\epsilon \neq \mathbf{0}$, which offers a type of 'bias correction' for systematic model discrepancies.

Computation

- Computing the probability of a climate event typically involves a double integral.
- For example, let Q be the subset of y_f values for which a doubling of atmospheric CO_2 causes global mean temperature to increase by at least 2°C :

$$\begin{aligned}\Pr(y_f \in Q \mid z = \tilde{z}) &= \int 1_Q(y_f) \Pr(y_f \mid z = \tilde{z}) dy_f \\ &= \int 1_Q(y_f) \int \varphi(y_f; x^*) \Pr(x^* \mid z = \tilde{z}) dx^* dy_f \\ &= \int \left\{ \int 1_Q(y_f) \varphi(y_f; x^*) dy_f \right\} \Pr(x^* \mid z = \tilde{z}) dx^* \\ &\equiv \int f(x^*) \Pr(x^* \mid z = \tilde{z}) dx^*\end{aligned}$$

where $\varphi(y_f; x) \triangleq \varphi(y_f; \mu_{f|z}(x), \Sigma_{f|z})$, and $f(x) \triangleq \int 1_Q(y_f) \varphi(y_f; x) dy_f$.

- With our choices (S1'), (T1), and (T2), computing $f(x)$ can be straightforward, and the challenge is integrating over x^* with the weighting function $\Pr(x^* \mid z = \tilde{z})$.

/...

Computation (cont)

- One simple method is *Monte Carlo integration*. If

$$I^{(n)} \triangleq n^{-1} \sum_{i=1}^n f(X_i) \quad X_i \stackrel{\text{iid}}{\sim} \Pr(x^* \mid z = \tilde{z})$$

then $\lim_{n \rightarrow \infty} I^{(n)} = \Pr(y_f \in Q \mid z = \tilde{z})$, by the Strong Law of Large Numbers.

- But we cannot sample easily from $\Pr(x^* \mid z = \tilde{z})$ so instead we sample from $\Pr(x^*)$ and then re-weight:

$$I^{(n)} \triangleq \sum_{i=1}^n w_i f(X_i) \quad X_i \stackrel{\text{iid}}{\sim} \Pr(x^*)$$

where

$$\begin{aligned} w_i &\propto \Pr(z = \tilde{z} \mid x^* = X_i) \\ &= \varphi(\tilde{z} - g_h(X_i); \mathbf{0}, \Sigma_{hh}^\epsilon + \Sigma^e) \quad \text{and} \quad \sum_{i=1}^n w_i = 1. \end{aligned}$$

- Better methods might include importance sampling with variance reduction techniques like antithetic variables.

Summary

Here are some questions for climate stakeholders to ask climate scientists, and for climate scientists to ask each other.

0. *Probability*. What do your probability statements about future climate represent? Why should we believe that your probability is a better guide to the future than someone else's?
1. *Measurements*. Do you have exact observations on historical and current climate data? If not, how have you quantified the measurement errors? (Σ^e in our treatment.)
2. *Model inadequacy*. Do you believe your model is so good that that there exists a model-input \hat{x} such that $y = g(\hat{x})$? If not, how have you quantified the model's imperfections? (Σ^e in our treatment.)
3. *The 'best' input*. How did you elicit ranges for the best model-input? Are the components probabilistically independent of each other? Are extreme values as likely as central values? ($\Pr(x^*)$ in our treatment.)

/ ...

Summary (cont)

4. *Computation*. How did you compute your probability? What uncertainty do you have about your estimate? [Possible follow-up: Isn't it rather reckless to use a random design if the model evaluations are very expensive?]

P.S. Craig, M. Goldstein, J.C. Rougier and A.H. Seheult (2001), Bayesian Forecasting for Complex Systems Using Computer Simulators, *Journal of the American Statistical Association*, **96**, 717-729.

M. Goldstein and J.C. Rougier (2005), Probabilistic Formulations for Transferring Inferences from Mathematical Models to Physical Systems, *SIAM Journal on Scientific Computing*, **26**, 467-487.

M. Goldstein and J.C. Rougier (2005), Reified Bayesian Modelling and Inference for Physical Systems, *Journal of Statistical Planning and Inference*, forthcoming as discussion paper.

M. Goldstein and J.C. Rougier (2005), Bayes Linear Calibrated Prediction for Complex Systems, *Journal of the American Statistical Association*, forthcoming.

J.C. Rougier (2005), Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations, *Climatic Change*, forthcoming.