Climate simulators and climate projections

Jonathan Rougier[1]
Department of Mathematics
University of Bristol
j.c.rougier@bristol.ac.uk

Michael Goldstein
Department of Mathematical Sciences
University of Durham
michael.goldstein@durham.ac.uk

[1]Corresponding author: School of Mathematics, University of Bristol, University Walk, Bristol BS8 1TW. Draft version, compiled October 7, 2013, prepared for the *Annual Review of Statistics and its Application*, 2014.

Abstract

We provide a statistical interpretation of current practice in climate modelling. This includes: definitions for weather and climate; clarifying the relationship between simulator output and simulator climate; distinguishing between a climate simulator and a statistical climate model; statistical interpretation of the ubiquitous practice of anomaly correction, and a substantial generalisation (the 'best parameter' approach); interpreting simulator/data comparisons as posterior predictive checking, and a simple adjustment to allow for double-counting. We also discuss statistical approaches to simulator tuning, assessing parametric uncertainty, and responding to unrealistic outputs. We finish with a more general discussion of larger themes.

Keywords: Climate modelling, tuning, history matching, anomaly model, best parameter model, model criticism

Contents

1 Introduction

Our purpose in this review is to interpret current practice in climate modelling in the light of statistical inferences about past and future weather. In this way, we hope to emphasise the common ground between our two communities, and to clarify climate modelling practices which may not, at first sight, seem particularly statistical. From this starting-point we can then suggest some relatively simple enhancements, and identify some larger issues. Naturally, we have had to simplify many practices in climate modelling, but not—we hope—to the extent that they are unrecognisable.

1.1 Weather and climate

We define 'weather' to be measurable aspects of our ambient atmosphere, notably temperature, precipitation, and wind-speed. Hence weather is an objective property of the world. We define a climate to be a subjective distribution for weather, represented as a multivariate space-time stochastic process. 'Distribution of weather' is uncontentious, but we believe that much confusion has arisen from attempts to treat climate as an objective property of the world, rather than something associated with a person, and reflecting his disposition to make bets—to adopt a common operationalisation of subjective probability, which we shall use throughout this review to treat all of the uncertainties within a common framework. Thus we write 'your climate' rather than 'the climate'; this seems to be the minimal change that is effective in emphasising the subjective viewpoint.

This subjective definition of climate is not one that many climate scientists will recognise, and so we take a moment to evaluate it. First, one standard definition of 'climate' is 'average weather', often represented as a thirty-year arithmetic mean. Under this type of definition (which may be extended to a much richer summary), climate is an objective property of the world, being simply a known function of weather. Thus one could bet on climate, rather than, as we would have it, climate being the bet one makes on weather. And then we would need a word for 'distribution of climate', because climate has become synonymous with summaries of weather. So we

have chosen to identify weather with its summaries, and reserve climate for 'distribution of weather'.

What about the subjective element? Climate modellers may not be happy about statisticians telling them that the distribution of weather is subjective. They may, for example, point to the histogram of recent past weather as a distribution of weather which is objective. But a histogram is not a distribution. If you make the subjective judgement of temporal exchangeability then the histogram of, say, 1980–2009 weather approximates your distribution of 2010 weather, albeit in a rather lumpy fashion. This is because probabilistic updating of an exchangeable sequence implies convergence to the histogram. So climate modellers who shared the judgement of exchangeability would roughly agree on the distribution of weather in 2010. But this argument is self-defeating, since subjectivity is necessary to turn the histogram into a distribution. It also highlights a common mistake, which is to confuse agreement with non-subjectivity.

The key point is that any probability for a unique event is unavoidably subjective; see, for example, Hacking (2001). The weather event 'there is at least one year of severe drought in England in 2020–2029' is a unique event, about which we cannot be certain before 2020, and may not be certain until 2030. At the moment, you may describe your assessment of this event probabilistically (it is an implication of your climate), but there is no reason to expect you to agree with anyone else. Your information, knowledge, and disposition are yours alone. Of course, a shared judgement of temporal exchangeability extending from 1980 to 2039 would be sufficient for agreement, but this is not a defensible judgement for a well-informed climate modeller, who is aware of the changes that are currently occurring in the earth system.[1]

Finally, our definition of climate seems to be consistent with current practice in climate modelling, as we will describe in more detail in the following sections. A 'climate simulator' is just that, a device for generating a family of distributions for weather. Insofar as the simulator is the outcome of

[1]Exactly the same considerations apply to the weather of the past. Agreement about the climate of the recent past follows from the convergence of exchangeable judgements on the histogram. But where there is no histogram, for example for palaeo-weather, there is no particular reason for agreement.

many judgements, its distribution is subjective. Climate modellers do not accept one of the simulator's climates as their own, but make a subjective adjustment reflecting their judgement about the simulator's limitations. So we find that the practice of climate modellers is inherently subjective, and that defining climate to be a subjective distribution for weather is reasonable not just from a foundational point of view, but also from a naturalistic one.

Just to be absolutely clear, we use the word 'subjective' to indicate only that by our definition climate may vary from one person to another. When judgements are subjective it behoves policymakers and the general public to exercise care when selecting their experts, to ensure that they are qualified and representative. In the case of future weather, it is climate scientists who are the experts, not statisticians like ourselves, and not journalists or bloggers. There is a huge body of climate science which is widely accepted within the climate science community, and this includes that the net effect of human activity since 1750 has been one of warming (Solomon *et al.*, 2007, Summary for policymakers) and that multiple lines of evidence attribute the observed warming to human activity (Hegerl *et al.*, 2007). These conclusions from the IPCC Fourth Assessment Report will shortly be reaffirmed in the Fifth Assessment Report (the IPCC AR5, to be finalised in 2014).

1.2 Climate simulators and their uses

The earth can be represented as a forced system, driven by variations in insolation, by volcanism and other tectonic processes, and by human activity (Peixoto and Oort, 1992). A climate simulator in its most primitive form is a function that maps forcings into weather, after which statistical post-processing of the weather can be used to produce a climate; this is discussed in more detail in section 2.1. We prefer the term 'climate simulator' for the function that does the mapping, reserving 'model' for its use in the statistical sense of a framework designed to simplify the process of specifying your climate (Rougier *et al.*, 2013). But because the word 'model' is heavily overloaded, we will write either 'statistical climate model', or 'XXX model',

where XXX is the name we give to a particular statistical model.[2]

In our sense, statistical climate models will typically encompass climate simulators. This is because the effect of forcing on the earth system has strong constraints that are induced by basic physical principles such as conservation and continuity. The qualitative effects of these constraints can be inferred for a simplified earth; for example, the large-scale atmospheric organisation known as Hadley Cells (see, e.g., Ahrens, 2000, chapter 11). However, a quantitative description on a realistic earth is less amenable to intuition, and must be computed. Thus quantitative statistical climate models are constructed in two stages: (i) develop a climate simulator which represents the physics, and (ii) propose a statistical model which represents your assessment of the simulator's limitations. Climate simulators are discussed in section 2, and statistical climate models in section 3.

Our distinction between 'climate simulator' and 'statistical climate model' is not widely made in climate science, but it exists implicitly, because climate modellers do indeed use statistical models to adjust a simulator's climate. The ubiquitous model is that, for a quantity such as temperature, the simulator's climate is acceptable only up to an unknown fixed offset (the 'simulator bias'), which in practice is estimated and plugged-in. This is the statistical interpretation of 'anomaly correcting', in which, as a matter of course, a simulator's temperatures are vertically shifted so that the simulator mean temperatures over the period 1980–1999 exactly match the mean of observed temperatures over the same period; see Figures 4 and 5 in Guttorp (2014, chapter ??? of this volume). Statisticians will immediately see the opportunities for generalising such a model. For example, the offset might be a spatial-temporal process; offsets from different types of output might be correlated; rather than plugging-in, the offset field might be integrated out, and so on.

[2]The term 'statistical climate model' is also used for a class of statistical-dynamical simulators, in which a separation of scale argument is used to model the large-scale effects directly, and to relegate the small-scale effects to a statistical ensemble; see Hasselmann (1976) for an outline of this approach, and Petoukhov *et al.* (2000) for a description of the CLIMBER-2 statistical-dynamical simulator. We do not consider this type of simulator here.

We make this point right at the start of this review, to stress that encompassing a climate simulator within a statistical climate model is not simply a statistician's conceit. Rather, it is something that already happens, but which, with statistical insight, could be generalised rather easily. It is crucial to appreciate that statistical judgements are necessary to move from climate simulator output to your climate, and these judgements must change as climate simulators evolve.

One role for climate simulators is hypothesis testing and predicting. Hypothesis testing is well-illustrated by 'detection and attribution' (D&A). In D&A, hypotheses compete to explain features such as the spatial-temporal structure of the warming trend in C20th weather. Hypothesis A is that this trend is simply a realisation of the weather's natural variability. Hypothesis B adds solar functions and volcanism. Hypothesis C adds human activities. These hypotheses are statements about the forcing. To compute your likelihood ratio for, say, B versus C you require your climate for a hypothetical earth without humans, as well as your climate for the actual earth. An earth without humans can be implemented in a simulator by fixing the forcing from atmospheric greenhouse gases in the industrial period to be the same as that before the industrial period. Hypothesis tests for D&A are discussed in Rougier (2008a) and reviewed in Hegerl and Zwiers (2011); for reasons of space, they will not be covered further here.

For predicting, policy interest is in future weather. In current practice the future is represented in terms of scenarios for future forcing, which themselves arise from scenarios for population, economics, technology, and policy interventions.[3] Again, climate simulators provide the means of considering and comparing various hypothetical futures. They provide a platform for what-if intervention studies, such as geo-engineering (e.g. Irvine *et al.*, 2011), and for driving regional simulations for climate impact studies (Parry *et al.*, 2007). Climate prediction is the main focus of this review, discussed in sections 3 and 4.

[3]Williamson and Goldstein (2012) describe an adaptive approach to simulator-based policy assessment, which avoids the use of scenarios.

2 Climate simulators

In this review we focus on large climate simulators, the state-of-the-art simulators that are run at the main climate research centres. The earth system comprises many interacting sub-systems, most notably the atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere, and the same is true of large climate simulators. Ahrens (2000) provides an introduction to weather (the companion volume Stull, 2000, is also helpful), with a more mathematical treatment in Peixoto and Oort (1992). McGuffie and Henderson-Sellers (2005) provide an introduction to climate modelling; Arakawa (1997) is a technical treatment outlining the mathematical issues involved in solving the underlying equations; Watanabe *et al.* (2010) describes some of the pragmatic choices that were made in constructing the MIROC5 simulator.

2.1 The simulator as a dynamical system

We will not consider the precise form of the laws governing the behaviour and interactions of the simulator modules, beyond observing that the laws of the earth's sub-systems are not all known, and that they are not currently solvable at a scale sufficient to resolve all of the interesting processes. Instead, we focus on the nature of a climate simulator, which is a forced non-linear dynamical system (see, e.g., McWilliams, 2007, who also provides a useful overview of the challenges of climate modelling). Therefore, for a particular set of forcings (suppressed in the notation), we consider a climate simulator to be a deterministic function of time

$$x_t = \varphi(t; x_0, \theta) \quad \text{such that} \quad \varphi(t_0; x_0, \theta) = x_0 \text{ for all } \theta,$$

where x_0 is the initial climate state at time t_0.[4] The parameters θ represent coefficients within the simulator code which are imperfectly known, or which are too abstract to have an operational meaning; some are found standing-in

[4]Here, we are simplifying by not distinguishing between the full state vector, and the function of the state vector of interest to us, which would usually be a lower-dimensional summary. Once we dispense with the initial condition, it suffices to treat x_t as the summary.

for processes that are filtered out by the solver (sub-grid-scale processes). Murphy *et al.* (2004) provide a list of about 30 of such parameters, while remarking that there are more than one hundred in a typical large-scale simulator; we return to this in sections 2.2 and 2.3.

The initial value x_0 must be supplied in order for the simulator to run, but it presents a major problem in practice, being very high-dimensional, and largely unknown, even in the case where time t_0 is contemporary—the difficulty is compounded if t_0 represents an historical initialisation date such as 1850. The tendency in climate science has *not* been to specify x_0 directly. Instead, the simulator, whose trajectories are chaotic, is treated as ergodic. A very long 'control run' is made from a specified $\hat{x}_0$ at a preferred set of parameter values (which we will denote $\tilde{\theta}$ below) and with constant or periodic forcing; for example, the forcing of the year 1850 might be repeated again and again. For an ergodic simulator, time averages converge to the stationary measure, and hence an x_0 for 1850, or a sequence of them, can be sampled from the control run after an interval of 'spin-up' to forget $\hat{x}_0$.

In this review we will not consider x_0 any further, but focus instead on the role of θ, the simulator parameters. For simplicity we will focus on the expectation and variance of the simulator's climate, although other features (for example, skewness and extremes) will also be important for climate impact assessment. Figure 1 summarises the exposition of the next few paragraphs. These paragraphs represent a somewhat idealised interpretation of current practice, suggesting an opportunity for increased statistical sophistication, should climate modellers desire.

We write the simulator output at time t as $x_t = \varphi(t; \theta)$, and the full set of simulator outputs as $\boldsymbol{x} := (x_1, x_2, \dots)$; note that x_t might itself represent a large collection of quantities, such as surface temperature, precipitation, and windspeed at every location on a 2° grid. The physics in the simulator suggests that for each t there will be strong relationships among the components of x_t, and that these relationships will be somewhat consistent across t. The chaotic nature of the simulator suggests that $\boldsymbol{x}$ will look like a realisation of a multivariate stochastic process, even when the forcing is smooth in time. For these two reasons, it is uncommon to work with $\boldsymbol{x}$ directly. Instead, a

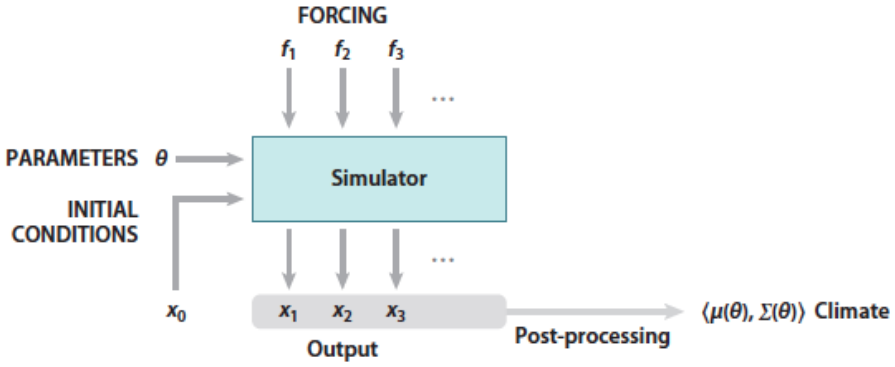


Figure 1: Schematic of a climate simulator. The inputs are forcing, $\boldsymbol{f} := (f_1, f_2, \dots)$, parameter values θ, and an initial value x_0. The simulator is run to produce outputs $\boldsymbol{x} := (x_1, x_2, \dots)$. Using statistical time-series modelling, these are summarised in terms of an expectation and a variance which together represent the simulator's climate, collectively denoted $\mathcal{K}(\theta) := \big\langle \mu(\theta), \Sigma(\theta) \big\rangle$. The forcing ought to be included in the arguments of $\mathcal{K}$, but is suppressed for simplicity; the argument x_0 is also suppressed but in fact $\mathcal{K}$ ought to be nearly invariant to perturbations in x_0, if the simulator is ergodic.

dimensionally-reduced summary is used.

Let $\boldsymbol{x}$ be arranged as a matrix

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \end{bmatrix}$$

with Singular Value Decomposition

$$X - \mathbf{1}\bar{x}^T = UDV^T$$

where $\mathbf{1}$ is the vector of ones, and $\bar{x}$ is the vector of column means (Golub and Van Loan, 1996, chapter 2, describe the SVD and its properties). In climate modelling, the columns of $W := UD$ are known as the Empirical Orthogonal Functions (EOFs, see, e.g., von Storch and Zwiers, 1999, chapter 13); they are a bit like Principal Components, except that the rows of X are not exchangeable—for this reason it is best not to confuse them. In this form the simulator output can be written

$$x_t = \bar{x} + Vw_t \qquad t = 1, 2, \ldots, \tag{1}$$

where w_t^T is one row of W. In practice, both V and w_t would be reduced from their full size to just the first k components, where k is determined empirically. In this case it would be sensible to rescale the columns of X to be dimensionless before taking its decomposition, or else to apply dimensional reduction separately to each type of output.

Following dimensional reduction, the second step is to fit a time-series model to $\boldsymbol{w} := (w_1, w_2 \ldots)$. A simple choice would be to model the components of w_t independently, given that $U^TU = I$ and D is diagonal, and to use a trend-plus-ARIMA-residual model for each component (see, e.g., Chatfield, 2004, chapter 4), including a seasonal component if the frequency is higher than annual. It would also be legitimate and helpful to include current and historical values of the forcing as covariates, in which case the trend may be unnecessary. The time-series model for $\boldsymbol{w}$ can then be used to infer the

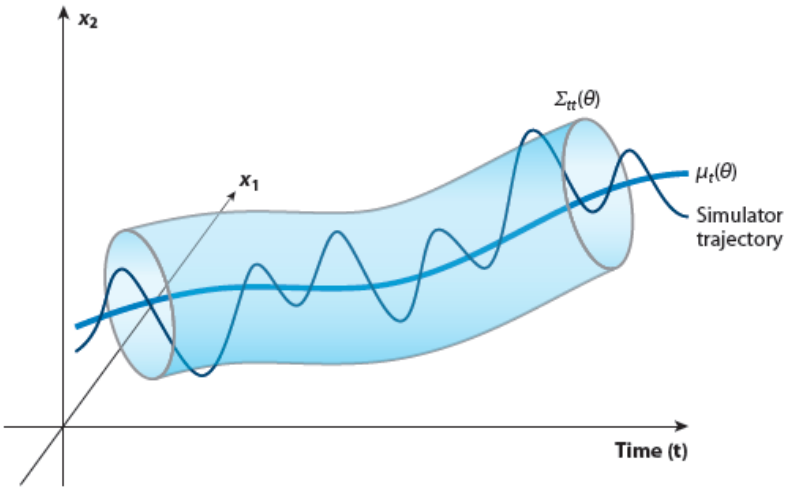


Figure 2: The set of simulator trajectories, and the tube which summarises the simulator's climate at parameter value θ.

second-order structure of $\boldsymbol{x}$ for given θ from (1), which we write here as

$$\boldsymbol{x} \sim \langle \mu(\theta), \Sigma(\theta) \rangle,$$

where $\mu_t(\theta) := \mathrm{E}(x_t; \theta)$ and $\Sigma_{tt'}(\theta) := \mathrm{Cov}(x_t, x_{t'}; \theta)$. Technically both μ and Σ should have hats, to indicate that the parameters of the process for $\boldsymbol{w}$ have been estimated and then plugged-in. Additional simulator runs with the same θ but different initial conditions can be used to improve the estimates of $\mu(\theta)$ and $\Sigma(\theta)$.

The tuple $\mathcal{K}(\theta) := \langle \mu(\theta), \Sigma(\theta) \rangle$ is synonymous with the climate of the simulator $\varphi(\cdot; \theta)$, at least to second order. $\mathcal{K}(\theta)$ describes an ellipse in the product space of time and the state, but for visualisation purposes much is gained by simplifying this to an elliptical tube in time, i.e. as a sequence $\{\mu_t(\theta), \Sigma_{tt}(\theta)\}$, which suppresses the temporal properties encoded in the covariances $\Sigma_{tt'}(\theta)$; see Figure 2. The mean function $\mu_t(\theta)$ can be used where

a deterministic treatment of the simulator is required: typically it would be similar to a spatially and temporally smoothed version of the raw output. $\Sigma_{tt}(\theta)$ is termed the 'internal variability' of the simulator output at time t. If the time series model for the residual of $\boldsymbol{w}$ is stationary, then $\Sigma_{tt}(\theta)$ is invariant to t, and the tube has the same shape all the way along.

In the special case of a 'time-slice experiment', the forcing is constant or periodic, and the simulator is run until its output stabilises. Typically the mean of the final thirty years of the run is used, which suppresses the internal variability to the point where it can be ignored. Time-slice experiments are used to summarise the climate of different epochs (e.g. the pre-industrial era, the Last Glacial Maximum) and to compute hypothetical quantities such as 'equilibrium climate sensitivity'. They are also used for tuning (see sections 2.2 and 2.3).

For reasons of computational scale it is sensible to perform inferences in the feature space $W = (X - \mathbf{1}\bar{x}^T)V$ rather than the original space, but in this article we will stay with the original space, for simplicity. Rather than take $\bar{x}$ and V from the actual run, they may instead be taken from the control run; in this case $\boldsymbol{w}$ might be modelled with a multivariate time-series model.

2.2 Tuning

'Tuning' is the activity of choosing a preferred value for θ, which we are denoting as $\tilde{\theta}$. The major constraint of tuning is the slow integration time of the simulator. Large-scale simulators can typically compute about one hundred simulator-years per calendar month. To tune to a long time-series, such as a century of regional temperature and precipitation, would additionally require a spin-up for each new choice of parameter values, which could easily require another hundred simulator-years. And so an iterative process with this target would move at about one cycle for every two calendar months, which is hopeless when there are hundreds of uncertain parameters. Therefore climate simulators are, in general, *not* extensively tuned to reproduce the large-scale features of C20th weather, especially as tuning generally happens in the shadow of a looming IPCC deadline. Therefore C20th weather

can be used to assess climate model adequacy (see section 4).

This observation should be tempered, though, in the light of the sequential development of simulators within a research group. Many of the decisions made when up-versioning a climate simulator are based on increased computer power, or better physical understanding; but some will be based on the failure of the current version of the simulator to reproduce C20th weather.[5] Valdes (2011) notes that climate simulators are currently too stable to replicate historical abrupt weather transitions. This might be symptomatic of over-tuning, the Holocene (our current epoch) being unusually stable. One solution is also to tune on previous epochs with different forcing: the difficulty here is that the forcings are much more uncertain, and the histogram of weather must be inferred from proxy measurements (see, e.g. Jones *et al.*, 2009).

There are two camps regarding tuning strategies: (i) that the modules of the climate simulator should be tuned separately, so as to avoid compensatory 'mis-tuning'; (ii) that climate is an emergent property of the interactions of its sub-processes, and so tuning should happen jointly. Typically something of a compromise is reached. Gent *et al.* (2011, section 3) summarises the procedure for CCSM4.[6] First, the modules of the simulator (atmosphere, ocean, land, sea-ice) were each separately tuned to reproduce current behaviour. Danabasoglu *et al.* (2008) illustrate the combination of physical and empirical reasoning that is used to tune one aspect of the ocean module.[7] Module tuning uses both time-slice experiments (to check for long-run stability), and also transient runs, where the fluxes from the other modules are replaced by observations.

Module tuning takes care of most of the parameters. Then Gent *et al.* coupled the modules together into a climate simulator, and the simulator as a whole was tuned on a small number of parameters and a small number of

[5]The genealogy of climate simulators is highly instructive; see Masson and Knutti (2011) and Knutti *et al.* (2013).

[6]This is an open-source climate simulator, which makes it rather unusual, as typically climate simulators are proprietary to climate modelling groups. CCSM4 has now been subsumed within CESM1, see `http://www.cesm.ucar.edu/models/ccsm4.0/`.

[7]Tuning involves much more than just adjusting parameters; often the parameter set itself is changed as chunks of code are swapped.

targets. A cloud parameter was adjusted to achieve a satisfactory radiation balance at the top of the atmosphere, and sea-ice albedo parameters were adjusted to give satisfactory sea-ice thicknesses in the Arctic.

Gent *et al.* also summarise the diagnostic evaluation of CCSM4 at the tuned value $\tilde{\theta}$, using observed C20th weather. Crucially, this evaluation is not just in terms of mean fields, e.g. for temperature and precipitation, although these get checked first, but also in terms of the statistical properties of variability (e.g. the histogram of precipitation) and recurrent events (e.g. the El Niño Southern Oscillation, ENSO). In other words, the purpose of tuning the simulator is not simply to get $\mu(\tilde{\theta})$ about right, but also to get key features of $\Sigma(\tilde{\theta})$ about right as well.

In our experience, the procedure for CCSM4 is unusually ascetic, with most modelling groups tuning jointly on a larger set of parameters and a larger set of targets. Mauritsen (2012) provides a detailed description of the process of tuning the MPI-ESM simulator. Public descriptions of the practice of tuning a large climate simulator are a recent phenomenon.

2.3 History matching

Statisticians have lots of tools to help with the process of tuning a climate simulator. Here we outline an exploratory approach termed 'history matching' (HM), which is much less demanding of expert judgement than fully-probabilistic conditioning.[8] HM is designed to rule out bad choices for the parameter values. 'Not ruled out' values—for which the simulator outputs are consistent with historical observations—do not necessarily have similar simulator outputs under different forcing (such as in future projections); the intention of HM, in contrast to tuning, is to preserve this source of climate uncertainty. HM originated in hydrocarbon reservoir modelling, and is extensively used commercially. It was given its original statistical formulation in Craig *et al.* (1997). Vernon *et al.* (2010) provides a detailed description of HM for a galaxy simulator; Gladstone *et al.* (2012) for the Pine Island glacier; Edwards *et al.* (2011, termed 'pre-calibration') for an intermediate

[8]Sansó *et al.* (2008) and Tokmakian and Challenor (2013) provide examples of fully-probabilistic calibration.

complexity climate simulator; and McNeall *et al.* (2013) for an ice-sheet simulator.

As an illustration, we will take just a single target for tuning, the top of the atmosphere (TOA) mean radiation balance in an 1850 time-slice experiment, with any value θ with an imbalance outside the range $(-0.1, 0.1)\,\mathrm{W/m^2}$ being deemed unacceptable as a candidate for the preferred value (see, e.g., Gent *et al.*, 2011, p. 4977). The width of this target interval should include a component for 'tolerability' (large discrepancies being tolerable for some targets but not for others), and also for measurement error; see, e.g., Vernon *et al.* (2010, section 3.5). HM inverts this constraint to rule out regions of the parameter space. To proceed efficiently it exploits the property that the simulator output for the target is a smooth deterministic function of the parameters. So in this case the simulator output would be, say, a 30-year mean of the radiation imbalance from the end of the run, denoted as $\bar{\mu}(\theta)$, and internal variability can be neglected.

The simulator is treated as an unknown smooth deterministic function of the parameters (or a subset of them), represented by a statistical model termed an *emulator*.[9] An emulator is a sophisticated response surface, typically containing both regressors for global effects, and a stochastic process for local effects. A catalogue of carefully-chosen runs at different points in the parameter space is used to update the emulator, and the result is an expectation and a standard deviation for $\bar{\mu}(\theta)$ at any θ. Note that $\bar{\mu}(\theta)$ is a random quantity even though θ is specified, if the simulator has not been run at θ. In the special case where the simulator *has* been run at θ, to give a value v say, then the smoothness of the emulator ensures that $\mathrm{E}\{\bar{\mu}(\theta)\} = v$ and $\mathrm{Sd}\{\bar{\mu}(\theta)\} = 0$. There are various strategies for choosing the set of runs, but a popular initial choice is a latin hypercube. Santner *et al.* (2003) or Forrester *et al.* (2008) provide more details about emulation and experimental design. Rougier (2008b) develops an emulator for multivariate outputs, such as a spatial field (see Rougier *et al.*, 2009a, for an illustration).

Based on the emulator, any point in the parameter space can be scored,

[9]Emulators are only required for expensive simulators: Gladstone *et al.* (2012), for example, use the simulator directly in the HM procedure.

according to whether the predicted value overlaps with the target. Thus a particular choice θ might be deemed *un*acceptable as a candidate for the preferred value if

$$\big(-0.1, 0.1\big) \cap \big(\mathrm{E}\{\bar{\mu}(\theta)\} \pm 3 \times \mathrm{Sd}\{\bar{\mu}(\theta)\}\big) = \emptyset.$$

Any θ for which the intersection is not empty is 'Not Ruled Out Yet'. Vernon *et al.* (2010) give a detailed description of the process of HM with an emulator, and how it can proceed in successive waves, allowing more and more of the parameter space to be ruled out through additional runs of the simulator and refittings of the emulator. Vernon *et al.* also discuss HM with multiple targets, and low-dimensional visualisations of an 'implausibility' measure defined on the parameter space.

Fast approximate simulators (FAS). The emulation approach is particularly powerful for simulators for which there are fast approximations. For climate simulators, these would typically be simulators with lower resolution, with prognostic variables replaced by diagnostic variables (effectively, removing feedback from some of the state variables), or with shorter spin-ups. In order to exploit this approach, it must be relatively easy to run the simulator in 'fast' mode, and this is something that must be designed in from the start.

For example, emulators can include arbitrary smooth functions of the parameters as regressors, and the FAS could be one such. In this way an emulator can be thought of as a statistical approach to correcting an FAS. After having built the emulator, which requires paired runs of both the full simulator and the FAS, the parameter space can then be explored at the speed of the FAS. Intuitively, if the FAS is a poor approximation that is difficult to correct, then not much of the parameter space will be ruled out, because $\mathrm{Sd}\{\bar{\mu}(\theta)\}$ will tend to be large.

In practice, a more sophisticated use of emulators is possible, linking simulators through the coefficients in their emulators, see Cumming and Goldstein (2009). In climate science, Rougier *et al.* (2009b) use a FAS to provide prior information for the HadSM3 climate simulator, and Williamson *et al.* (2012)

link the low-resolution FAMOUS climate simulator with the high-resolution HadCM3 simulator.

3 Statistical climate models

In this section an important transition is made, from the climate simulator, thought of as a function of the parameters θ, to your climate. In passing from one to the other we pass into the realm of subjective judgements, as explained in section 1.1. If this subjectivity is not obvious, it must be due to the widespread acceptance of conventional judgements. And, indeed, the 'anomaly model' described in section 3.1 is exactly this: a conventional judgement for passing from a simulator's climate to your climate, which conceals the essential subjectivity of this step.

Conventions can be supremely useful, of course. For example, symmetry-breaking conventions, such as agreeing to drive on the lefthand side (in the UK). Conventional *simplifications*, on the other hand, must continually be reappraised as our understanding and our tools develop. Thus many of the conventional simplifications in climate modelling have now been relaxed, and parts of the earth system previously ignored or treated as diagnostic have become prognostic (e.g. the sulphur cycle, vegetation). The anomaly model is a conventional simplification of judgements which goes back to the very start of climate modelling, and the time has come to relax it too. For example, the 'best parameter' model that we discuss in section 3.2 is a useful first step in this direction.

In this section we contrast two different approaches to inferences about future weather under particular future forcings, termed climate *projections*. The time domain is divided into the Past ($t \in \mathcal{P}$), for which there are observations, and the Future ($t \in \mathcal{F}$), for which we would like to make a projection. For the time being we treat the past forcing as known (but see section 4.2); the future forcing is specified by the projection scenario. For concreteness, $\mathcal{P}$ might be the period 1850–2013 and $\mathcal{F}$ the period 2014–2100, and the simulator output might be annual global mean temperature. The time-period $\mathcal{P}$ does not have to be contiguous with that of $\mathcal{F}$, and for simplicity we take

$\Sigma_{\mathcal{PF}}(\theta) = \mathbf{0}$ for each θ.[10]

Let $Y = [Y_{\mathcal{P}}, Y_{\mathcal{F}}]$ be the weather, and let

$$\boldsymbol{z}^{\text{obs}} = Y_{\mathcal{P}} \oplus \boldsymbol{e}$$

be the statistical model for past weather observations, where $\boldsymbol{e}$ is measurement error with expectation zero and known variance matrix E, and '$\oplus$' indicates the addition of uncorrelated components.[11] Then the objective is to make inferences about $Y_{\mathcal{F}}$ based on $\boldsymbol{z}^{\text{obs}}$, and on runs of the simulator. We restrict ourselves to the single simulator run at the preferred parameter value $\tilde{\theta}$, represented in terms of the simulator's climate $\mathcal{K}(\tilde{\theta}) = \langle \mu(\tilde{\theta}), \Sigma(\tilde{\theta}) \rangle$. Section 3.3 discusses the important issue of alternative choices for θ.

3.1 The anomaly model

Our statistical interpretation of climate modellers' current behaviour is that they take a classically Frequentist approach to climate inference, proposing a strongly parametric statistical model linking the climate simulator and their climate, estimating the parameters of this statistical model, and then plugging-in the estimated values to make probabilistic projections.

We will term the climate modellers' current statistical model the 'anomaly model'. It asserts the existence of parameters (θ^*, α^*) with the property that

$$Y \mid \theta^*, \alpha^* \sim \langle \mu(\theta^*) + \alpha^* \mathbf{1}, \Sigma(\theta^*) \rangle \tag{2}$$

where α^* is a scalar *anomaly correction*. This statistical model asserts that $\mathcal{K}(\theta^*)$ adequately represents your climate, except for an unknown translation.

[10] In practice, a much more detailed set of outputs would be used—generalising the approach described below is straightforward. Where $\mathcal{P}$ and $\mathcal{F}$ are contiguous, $\Sigma_{\mathcal{PF}}(\theta) \approx \mathbf{0}$ is implied by short (say, not more than a decade) correlation lengths in the residual component of the time-series model for $\boldsymbol{w}$. Again, generalising is straightforward.

[11] We are skipping over the nature of these observations. Many weather observations start out as indirect measurements, e.g. from weather satellites, which measure radiances at different wavelengths, which are then processed ('inverted') to give temperatures. Common sources of uncertainty in the forward relationship from temperature to radiance would induce systematic errors in the observations; see section 4.2. But it would be unusual for the weather and the measurement error to be correlated.

Eq. (2) is the simplest version, and can easily be generalised to allow the anomaly correction to have multiple components, which depend on the type of weather output, and possibly the spatial location. Some outputs may need to be transformed in order that an additive shift is appropriate (e.g. precipitation, which is non-negative on its natural scale).

The tuned value $\tilde{\theta}$ is taken to be an estimate of θ^*, and α^* is then estimated as

$$\tilde{\alpha} = (n_p)^{-1} \sum_{t \in \mathcal{P}} \left(z_t^{\text{obs}} - \mu_t(\tilde{\theta}) \right) \tag{3}$$

in the simplest version, where n_p is the number of time points in the period $\mathcal{P}$. Thus the anomaly correction is a function of $\boldsymbol{z}^{\text{obs}}$, even if $\tilde{\theta}$ is not. The projection is then found by plugging-in the estimate $(\tilde{\theta}, \tilde{\alpha})$ for the unknown (θ^*, α^*), to give

$$Y_{\mathcal{F}} \sim \left\langle \mu_{\mathcal{F}}(\tilde{\theta}) + \tilde{\alpha}\mathbf{1}, \Sigma_{\mathcal{F}\mathcal{F}}(\tilde{\theta}) \right\rangle. \tag{4}$$

This is a rather long-winded way of saying "Having found, by other means, a preferred value for θ, translate the simulator climate so that it matches, on average, the historical observations".[12] But we emphasise that the parametric model in (2) is a subjective assessment of the relationship between the simulator's climate and your climate, notwithstanding the aura of objectivity that arises from the apparent absence of any explicit quantification of uncertainty. This is discussed further in section 3.2.

3.2 The 'best parameter' model

The anomaly model of section 3.1 is rather simple. The statistical field of computer experiments has developed a richer set of models for simulator-based inference for complex systems (see, e.g., Kennedy and O'Hagan, 2001; Craig *et al.*, 2001; Goldstein and Rougier, 2004, 2006), for which Rougier (2007) provides an illustration in climate science. These are based around a 'best parameter' statistical model, which in this context asserts that there exists a θ^* such that $\mathcal{K}(\theta^*)$ is second-order sufficient for your climate. This

[12] In fact, as mentioned in section 1, the convention is to match the means over the period 1980–1999, rather than the whole of $\mathcal{P}$, as $\mathcal{P}$ might differ from one experiment to another.

model can be written as

$$Y \mid \theta^*, \boldsymbol{\varepsilon} \sim \big\langle \mu(\theta^*) + \boldsymbol{\varepsilon}, \Sigma(\theta^*) \big\rangle$$

where $\boldsymbol{\varepsilon}$ is an additive 'discrepancy', probabilistically independent of θ^*, with $\mathrm{E}(\boldsymbol{\varepsilon}) = m$ and $\mathrm{Var}(\boldsymbol{\varepsilon}) = T$, both m and T being specified. We write 'discrepancy' instead of 'anomaly correction' because $\boldsymbol{\varepsilon}$ is a random vector, not a scalar shift. It is common to take $m = \mathbf{0}$, because known translations can be incorporated directly into the simulator, and we will do this from now on. Integrating out $\boldsymbol{\varepsilon}$ gives

$$Y \mid \theta^* \sim \big\langle \mu(\theta^*), \Sigma(\theta^*) + T \big\rangle.$$

The discrepancy variance T captures your remaining uncertainty about climate, were you able to run the simulator at its best parameter. First, you would not expect $\mathcal{K}(\theta^*)$ to be in quite the right place, and so there ought to be a term that provides for translations, like the anomaly correction. Second, $\mathcal{K}(\theta^*)$ is typically under-dispersive with respect to your climate, due to the limited resolution of the solver, which acts as a filter.[13] Thus, in the simplest version of the Best Parameter model

$$\boldsymbol{\varepsilon} = \alpha^* \mathbf{1} \oplus \boldsymbol{\varepsilon}_2 \quad \text{implying} \quad T = \sigma_1^2 \mathbf{1}\mathbf{1}^T + T_2 \tag{5}$$

where $\alpha^* \sim \big\langle 0, \sigma_1^2 \big\rangle$, and $T_2 := \mathrm{Var}(\boldsymbol{\varepsilon}_2)$ might be as simple as the diagonal matrix $\sigma_2^2 I$. But T also has the capacity to encode changes in the shape of $\mathcal{K}(\theta^*)$, to give a richer and more appropriate description of the simulator's discrepancies at different time points.

In this model, the projection is made by updating using the historical observations, giving

$$Y_{\mathcal{F}} \mid \theta^*, \boldsymbol{z}^{\mathrm{obs}} \sim \big\langle \mu_{\mathcal{F}|\mathcal{P}}(\theta^*), \Sigma_{\mathcal{F}|\mathcal{P}}(\theta^*) \big\rangle$$

[13] This is not the only source of under-dispersion: there are also simplifications in the modelling of processes such as sea-ice and vegetation, reflecting both computational constraints and lack of knowledge.

where, taking $\Sigma_{\mathcal{PF}} = \mathbf{0}$ and suppressing θ^* (which is an argument for all terms of the form μ or Σ)

$$\mu_{\mathcal{F}|\mathcal{P}} := \mu_{\mathcal{F}} + T_{\mathcal{FP}}(\Sigma_{\mathcal{PP}} + T_{\mathcal{PP}} + E)^{\dagger}(\boldsymbol{z}^{\text{obs}} - \mu_{\mathcal{P}})$$
$$\Sigma_{\mathcal{F}|\mathcal{P}} := \Sigma_{\mathcal{FF}} + T_{\mathcal{FF}} - T_{\mathcal{FP}}(\Sigma_{\mathcal{PP}} + T_{\mathcal{PP}} + E)^{\dagger}T_{\mathcal{PF}}.$$

These are the usual second-order updating equations (Goldstein and Wooff, 2007; Rougier *et al.*, 2013), where † denotes the Moore-Penrose inverse.[14] For a plug-in projection, θ^* can be replaced by its estimate $\tilde{\theta}$.

Interestingly, this best parameter model includes the anomaly model as a special case. We illustrate in the simplest version, given in (5). If

$$\sigma_1^2 \mathbf{1}\mathbf{1}^T \gg T_2 \quad \text{and} \quad \sigma_1^2 \mathbf{1}\mathbf{1}^T \gg \Sigma_{\mathcal{PP}} + E, \tag{6}$$

then $(\Sigma_{\mathcal{PP}} + T_{\mathcal{PP}} + E)^{\dagger} \approx T_{\mathcal{PP}}^{\dagger} = n_p^{-2}\sigma_1^{-2}\mathbf{1}\mathbf{1}^T$. Simple arithmetic then shows that the projection is the same as in the anomaly model: $\mu_{\mathcal{F}|\mathcal{P}} = \mu_{\mathcal{F}} + \tilde{\alpha}\mathbf{1}$ and $\Sigma_{\mathcal{F}|\mathcal{P}} = \Sigma_{\mathcal{FF}}$, where $\tilde{\alpha}$ was defined in (3). The assertions in (6) state that your concern for the mis-location of the simulator's climate dominates all other uncertainties. A climate modeller who did not believe (6) would regard the anomaly-corrected projection (4) as over-fitted. This modeller would attribute part of the systematic difference between $\boldsymbol{z}^{\text{obs}}$ and $\mu_{\mathcal{P}}(\theta^*)$ to internal variability, and so adjust the location of the simulator climate by less, and decrease the projection uncertainty by less.

Compared to the anomaly model, the best parameter model allows much more flexibility for the discrepancy, including that it is a stochastic process. For example, $\alpha_t^* = \rho\alpha_{t-1}^* + \eta_t$, with $\alpha_0^* \sim \langle 0, \sigma_1^2 \rangle$ and $\boldsymbol{\eta}$ a sequence of uncorrelated innovations with $\eta_t \sim \langle 0, (1-\rho^2)\sigma_1^2 \rangle$, for which the anomaly model is the special case of $\rho = 1$. This model for α_t^* allows the anomaly to be both uncertain and time-varying, with the value ρ controlling the temporal correlation length. The main difficulty for climate modellers is that it seems more acceptable to specify $\rho = 1$ rather than choose a value for ρ less than one. Likewise, it seems more acceptable to specify $\sigma_2 = 0$ after (5) rather

[14] They are also the conditioning expressions for the multivariate Gaussian distribution.

than choose a value greater than zero.[15] While there are much more detailed judgements that can be represented in T, if the starting-point is the anomaly model, then $\rho < 1$ and $\sigma_2 > 0$ are very simple extensions.

Sexton *et al.* (2012) is the one study in climate modelling which has explicitly included a full discrepancy variance. They use an approach based on an ensemble of simulators from different modelling groups, generating one realisation of the discrepancy per simulator, and estimating T from the result. The small number of simulators in their ensemble would produce a severely rank-deficient value for T, but Sexton *et al.* perform the entire inference in the much lower-dimension feature space (see section 2.1). While hesitant to endorse this particular approach, our view is that any approach that helps climate modellers to propose a T which adjusts both the location and the dispersion of the simulator's climate is welcome, provided that the results are not inconsistent with the modellers' judgements.

3.3 Uncertainty about the best parameter

The projections in section 3.1 and 3.2 were both made for a specific parameter value, the preferred value $\tilde{\theta}$, thought of as a plug-in point-estimate for θ^*. The fact that $\tilde{\theta}$ is not θ^*, and that θ^* remains uncertain, is a concern in climate modelling, and there have been several experiments to assess the sensitivity of simulator output to parameter perturbations, reviewed by Murphy *et al.* (2011). So far, though, there has been no attempt to quantify the effect of parameter uncertainty on climate projections, for the current generation of climate simulators.

Formally, this quantification is straightforward: $\mathcal{K}(\theta)$ denotes the expectation and variance conditional on $\theta^* = \theta$, and so integrating out θ^* gives

$$\mathcal{K}^* = \big\langle \mathrm{E}[\mu(\theta^*)], \mathrm{E}[\Sigma(\theta^*)] + \mathrm{Var}[\mu(\theta^*)] \big\rangle. \tag{7}$$

The expectations and variance can be replaced by finite approximations. In

[15] Boundary values such as $\rho = 1$ and $\sigma_2 = 0$ give the appearance of objectivity, but are of course just as subjective as any other values, and less defensible than many. Box (1980, p. 384) commented on "the curious idea that an outright assumption does not count as a prior belief".

the simplest case these would be from an ensemble of runs sampled from the prior distribution for θ^* (Rougier, 2007), but much more sophisticated approaches are possible using emulators (Craig *et al.*, 2001; Rougier and Sexton, 2007; Rougier *et al.*, 2009b). The expectation in $\mathcal{K}^*$ will not be the same as $\mu(\tilde{\theta})$, even in the case where $\tilde{\theta}$ is the prior expectation of θ^*, because μ is a non-linear function, perhaps extremely so in some parts of the parameter space (McWilliams, 2007). And $\Sigma(\tilde{\theta})$ will typically understate the variance in $\mathcal{K}^*$, because of missing the $\text{Var}[\mu(\theta^*)]$ term.

However, there is a computational price to pay for exploring the effect of uncertainty in θ^*, because today's climate simulators are about as large as computing resources will allow, and there are no spare CPU cycles for replication. So running different candidate values for θ^* is only possible by reducing the simulator's resolution.[16] Halving the resolution of today's simulators would allow about ten low-resolution simulator runs instead of one high-resolution simulator run, putting aside concerns about spinning-up.

In their choices, climate modellers currently reveal a strong preference for one high-res run, rather than, say, ten low-res ones.[17] This reluctance to use low-res runs for assessing projection uncertainty has implications for the statistical model. The replacement of θ^* by the plug-in point estimate $\tilde{\theta}$ could be compensated by increasing the variance of the discrepancy in the best parameter model, i.e. letting T_2 in (5) represent $\text{Var}[\mu(\theta^*)]$. But this would require an explicit departure from the anomaly model, for which $T_2 \approx \mathbf{0}$. So climate modellers find themselves in a statistical *impasse*. As we perceive it, the choice to do one high-res run instead of a set of low-res runs is incompatible with the use of the anomaly model to link the climate simulator and actual climate, unless one is prepared to defend the judgement

[16] Or by other approaches to speeding up the simulator, as discussed in section 2.3; but here we focus on resolution.

[17] It is not for us to speculate on why this is so. But as (statistical) modellers ourselves, we are familiar with the vexed issue of 'realism'. After a successful 'up-versioning', today's climate simulator looks appreciably more realistic than before. For example, Gent *et al.* (2011, section 5d) document the improvement of CCSM4 over CCSM3 in representing ENSO. ENSO is an important feature of the earth system, and a milestone for climate simulation, and it must be painful for climate modellers who have finally achieved this milestone to then reduce resolution and lose it again. Salt (2008) provides an interesting reflection on modelling culture.

that $\mathrm{Var}[\mu(\theta^*)] \approx \mathbf{0}$, i.e. that perturbing the parameters has a negligible effect on the expectation of the simulator's climate.

We also believe that the reluctance to perform low-res runs is misplaced. Crudely, today's low-res simulator 'φ_{low}' is the previous IPCC report's state-of-the-art simulator. At the time of the previous report, perturbations of φ_{low} were thought to be informative about your climate. Today perturbations of φ_{high} are thought to be informative about your climate. Accepting the premise—which is not disputed—that climate simulators are currently much more like each other than any one climate simulator is like your climate, you must accept that perturbations in φ_{low} are informative about perturbations in φ_{high}. Statistical models for sequences of simulators are discussed in Goldstein and Rougier (2009), and statistical models for a collection of climate simulators of roughly equal fidelity are discussed in Rougier *et al.* (2013).

We hope that by the time of the sixth IPCC report (around about 2020), climate modellers will be exploring the limitations of the tuning process, and quantifying the effect of parametric uncertainty. Ideally this would take the form of carefully designed experiments combining runs of low- and high-res simulators, which we believe is the most efficient way to exploit a fixed budget of CPU cycles (Cumming and Goldstein, 2009).

4 Model criticism

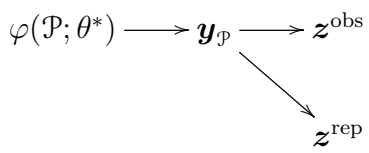
Model criticism, also termed 'model validation' by engineers, attempts to evaluate whether the statistical climate model is adequate for its purpose.

4.1 Computing residuals

There is a lot of informal model criticism in current climate modelling, but, at first glance, very little formal statistical model criticism. Informal model criticism tends to proceed by comparing $\boldsymbol{z}^{\mathrm{obs}}$ with the anomaly-corrected simulator output $\mu_{\mathcal{P}}(\tilde{\theta}) + \tilde{\alpha}\mathbf{1}$, and seeing whether the differences are large relative to the standard deviations in $\Sigma_{\mathcal{P}\mathcal{P}}(\tilde{\theta}) + E$; see the statement in Randall *et al.* (2007, section 8.1.2.3). Interestingly, under the anomaly model

this approach is precisely the posterior predictive checking (PPC) approach originally proposed by Rubin (1984), and advocated by Gelman *et al.* (2003, chapter 6).

In the PPC approach, one notionally replicates the observations, which in our case gives rise to the second-order Directed Acyclic Graph (DAG)



where $\boldsymbol{z}^{\text{rep}}$ are the notionally replicated observations. Then $\boldsymbol{z}^{\text{obs}}$ is evaluated with respect to the expectation and variance of $\boldsymbol{z}^{\text{rep}}|\boldsymbol{z}^{\text{obs}}$. Under the anomaly model, now thought of as a special case of the best parameter model as discussed in section 3.2,

$$\boldsymbol{y}_\mathcal{P} \mid \theta^*, \boldsymbol{z}^{\text{obs}} \sim \left\langle \mu_\mathcal{P}(\theta^*) + \tilde{\alpha}\mathbf{1}, \Sigma_{\mathcal{P}\mathcal{P}}(\theta^*) \right\rangle$$

and the result then follows by taking $\boldsymbol{z}^{\text{rep}} = \boldsymbol{y}_\mathcal{P} \oplus \boldsymbol{e}^{\text{rep}}$ where $\boldsymbol{e}^{\text{rep}}$ is uncorrelated with $\boldsymbol{e}$ but has the same expectation and variance, and plugging-in $\tilde{\theta}$ for θ^*.

Gelman *et al.* advocate using graphical summaries of $\boldsymbol{z}^{\text{obs}}$ in the distribution of $\boldsymbol{z}^{\text{rep}} \mid \boldsymbol{z}^{\text{obs}}$; in climate these might be plots of rescaled prediction errors,

$$r_t := \frac{z_t^{\text{obs}} - \big(\mu_t(\tilde{\theta}) + \tilde{\alpha}\big)}{\sqrt{\Sigma_{tt}(\tilde{\theta}) + E_{tt}}} \qquad t \in \mathcal{P}, \tag{8}$$

which can also be very effective if $\boldsymbol{z}^{\text{obs}}$ is a spatial map. These residuals are not standardised to have variance one when the model is adequate, due to the double-counting of $\boldsymbol{z}^{\text{obs}}$, which is used to estimate $\tilde{\alpha}$. The adjustment is straightforward. If $H := I - (n_p)^{-1}\mathbf{1}\mathbf{1}^T$, termed the 'centering matrix' (Mardia *et al.*, 1979, chapter 1) then

$$\boldsymbol{z}^{\text{obs}} - \big(\mu_\mathcal{P}(\tilde{\theta}) + \tilde{\alpha}\mathbf{1}\big) = H\big(\boldsymbol{z}^{\text{obs}} - \mu_\mathcal{P}(\tilde{\theta})\big),$$

and so, letting $V := \Sigma_{\mathcal{P}\mathcal{P}}(\tilde{\theta}) + E$, the denominator of the residuals in (8) should be not $\sqrt{V_{tt}}$, but $\sqrt{(HVH)_{tt}}$. Slight modifications to H would be required for different anomaly conventions (see footnote 12).

Multivariate information including covariances can be harder to visualise; Bastos and O'Hagan (2009) give a simple approach based on the pivoted Choleski decomposition of the predictive variance.

4.2 Diagnostic warnings

How should the climate modeller respond to a subset of diagnostics that are large in absolute size?[18] Suppose, for example, that many of the standardised precipitation residuals are larger than 3 in absolute size, in a region such as western Europe, where precipitation changes are an important feature of climate change impact. There are several options:

1. Acknowledge that the simulator does not yet 'do' precipitation, and request a more powerful computer, hoping that higher resolution will reduce the residuals.

2. Acknowledge that the simulator has been badly tuned, and restart the tuning process, hoping that a better choice for $\tilde{\theta}$ will reduce the residuals.

3. Acknowledge that the anomaly model is a rather simplistic representation of judgements about the simulator and your climate, hoping that a better specification for the discrepancy variance T will reduce the residuals.[19]

In all three options the climate modeller should decline to make projections for precipitation, given the failure of his model. Instead, he should wait for

[18] We can also imagine situations where some linear combinations of the residuals are surprisingly small, because the climate model has failed to respect physical constraints. However, at the moment climate modellers are mainly concerned with residuals which are too large.

[19] A more sophisticated variant of this option is to use statistical downscaling to adjust the simulator climate; see Maraun *et al.* (2010).

his new computer, or for the simulator to be re-tuned, or while he refines his judgements about T.

We would strongly recommend exploring option 3 first! Once T is explicitly specified, the climate modeller can compute *prior* predictive residuals

$$r'_t := \frac{z_t^{\text{obs}} - \mu_t(\tilde{\theta})}{\sqrt{\Sigma_{tt}(\tilde{\theta}) + T_{tt} + E_{tt}}} \qquad t \in \mathcal{P} \tag{9}$$

(Box, 1980). These are standardised to have expectation zero and variance one when the model is adequate.

The idea that T might be adjusted retrospectively to improve the residuals $\boldsymbol{r}'$ needs to be clearly motivated. First, at a pragmatic level, the climate modeller might have decided to use the 'objective' anomaly model if he can, despite it not being a defensible representation of his judgements. So large residuals under the anomaly model indicate that T is an area of the inference where he must make an additional effort.

Second his adjustment ought to be to the whole of T, not just to the submatrix $T_{\mathcal{P}\mathcal{P}}$. He might, for example, reflect on whether $\rho = 1$ and $\sigma_2 = 0$ were really appropriate choices (see section 3.2). Setting $\rho < 1$ and/or $\sigma_2 > 0$ will increase both the historical uncertainty about $\boldsymbol{z}$, and the projection uncertainty about $Y_{\mathcal{F}}$. It is self-evident that a failure of the simulator to match historical observations increases uncertainty about climate projections. Adjusting the whole of T in order to improve the residuals in $\mathcal{P}$ is a simple implementation of this.

Third, from its position in (9) it is clear that T can take on some of the burden of specifying $\Sigma_{\mathcal{P}\mathcal{P}}(\tilde{\theta})$ and E. Both of these are challenging. $\Sigma_{\mathcal{P}\mathcal{P}}(\tilde{\theta})$ is expressed for a specified forcing, yet C20th forcing is not well-known (Forster *et al.*, 2007). So some of the increase in T might reflect your assessment of uncertainty in the forcing. For E, Guttorp (2014, chapter ??? of this volume) describes some of the issues with climate observations. Statisticians will appreciate that common uncontrolled sources of variation in the collection and processing of instrumental readings introduce non-zero off-diagonal elements into the observation error variance E; no attempt has been made, so far, to

tackle this. So, again, some of the covariances in T might reflect common sources of variation in the observations.

Finally, we note that there are conceptual difficulties in the best parameter model, particularly in specifying judgements which are coherent for a sequence of simulators. Goldstein and Rougier (2009) introduce a more general approach, termed 'reified modelling'. Thus although modifying T might be the incremental response to poor residuals, the reified modelling approach might turn out to be a better representation of a climate modeller's actual judgements.

5 Summary and prospects

There are two issues which should engage statisticians working in climate research. (i) Given where we are, what is the pragmatic statistical analysis which will best complement/enhance current climate practice? (ii) Given an ideal position, what is the analysis that we would like to carry out, to best inform us about future weather?

This review has been largely concerned with aspects of the first issue. We briefly summarise the main sources of uncertainty that we have identified, for a given climate projection:

1. Input uncertainty, which is not knowing the historical and future boundary conditions;

2. Parametric and structural uncertainty, which arise from limitations in the climate simulator (and subsume other uncertainties such as the effect of code errors and numerical noise);

3. Observational error, including common components which induce covariances in the observation error variance;

4. Code uncertainty, which is being unable to run the simulator at any desired parameter setting.

These sources of uncertainty are all familiar to statisticians, but not routinely addressed in climate projections. We hope that our review has identified

some simple extensions of current practice, and also the opportunity for more detailed treatment, where judgements allow.

We now turn to the second issue. Changes in weather over the present century have the capacity to threaten literally hundreds of millions of people. To give just one peril, Nicholls (2011) discusses the effect of sea-level rise: currently, over 200 million people are vulnerable to flooding during extreme storms, and the probability of a catastrophe will increase as sea-levels continue to rise through the century. Among the huge uncertainties affecting the risk of flooding are the behaviour of the Greenland and West Antarctic ice-sheets, the intensification of tropical and extra-tropical storms, and changes to surge propagation (*ibid.*, p. 147–148). Our society's response to this peril, and others like it, might come to be seen as one of the defining features of political and social activity for the current century.

Addressing the original four uncertainties is clearly already a large challenge for climate modellers. But when we consider the impact of the weather, we have to allow for additional uncertainties, principally:

5. downscaling uncertainty, which maps from the large grid-cell of global climate simulators to the small grid-cell that is necessary to evaluate losses, taking account of local topography and bathymetry;

6. loss uncertainty, which is valuing the harm and damage caused by climate-related hazards;

7. decision uncertainty, which is the uncertain consequence of an intervention, which depends on social and economic factors.

All of these are single-simulator concerns, and so we must add:

8. multi-simulator uncertainty, accounting for the sequence of simulators within each research group, and the different simulators across research groups.

Now a full uncertainty assessment for climate policy appears doubly daunting. However, that is the wrong way to look at the problem. The really

daunting task is to make and successfully implement a climate policy without doing a careful uncertainty analysis.

While it is true that these sources of uncertainty are challenging to assess, they are no more challenging than other parts of climate modelling (doing the basic science, formulating mathematical models, constructing simulators that run efficiently on super-computers, designing satellite missions to collect observations, and so on). The difference is that the climate modelling challenges are addressed by well established communities. If climate policy is a genuine concern, then scientifically-leading countries such as the UK need to develop a similar community of climate statisticians, working alongside the other communities, and funded at a sufficient level, with the same access to computing facilities.

In this case we would hope to see rapid development along the lines outlined in this review. This would include the replacement of tuning with history matching, the incorporation of parametric uncertainty into projections, and the use of fast approximate simulators in experimental design and emulation.

In a few years we would hope to see a gradual acceptance among climate modellers that the distribution of weather is subjective, and that current approaches based on suppressing that subjectivity using boundary choices for statistical parameters, as in the anomaly model, are indefensible. We will need creative approaches to specifying the discrepancy variance matrix. Following on from this, we hope for the development of more powerful statistical modelling approaches for linking multiple simulators, in order to put climate policy at the heart of the inference (our own suggestion is reified modelling, see Goldstein and Rougier, 2009).

Within the decade, we look for the development of new statistical techniques which modularise inference for future weather impacts in the same way that climate modelling itself is modularised. We expect these to be based on dynamical graphical models, where the main challenge is the very high level of interconnectivity between the vertices. These developments in statistical methodology would be widely applicable, useful for any complex systems wherever uncertainty about real world consequences is informed by

families of computer simulators.

Finally, we address the political dimension of the issues that we have discussed. It would be relatively straightforward to consider what perils we should protect against, and to what degree, were we to know precisely what the future held in store. However, our world is far too complex to offer such certainty. A common line of argument is that the element of subjectivity involved in climate projections justifies taking a 'wait and see' attitude towards actions for climate change mitigation and adaptation. This argument has a potentially paralysing effect on informed discussion over climate policy, and it is understandable that scientists, concerned that such arguments will be used to discount the utility of their assessment, may be tempted to downplay the uncertainty associated with their projections.

This is doubly unfortunate. First, downplaying or 'objectifying' uncertainty makes much valuable and informative work in climate science an easy target for groups with a vested interest in preserving the *status quo* (see, for example, Oreskes and Conway, 2010). Second, this suppression of uncertainty masks much of the case for taking action now, because climate projections are by no means worst-case scenarios. Rather, they are located near the centre of a range of possible future climates, all of which ought to be considered in order to develop appropriate climate policies. Such policies will, inevitably, be constructed in a context of uncertainty, and this uncertainty can only be assessed in terms of expert judgements, based on a synthesis of all the available evidence.

That there is some disagreement between scientists studying climate is natural and inevitable. However, as mentioned in the Introduction, the broad lines of the argument for human-induced climate change are clear (and we should not forget that this is just one of the perils we face, see Rockström *et al.*, 2009). Rational choice of action in complex problems requires careful consideration of both uncertainties and consequences. Analysis of the most careful and detailed climate projections that are possible within current computational constraints consistently suggests the potential for human activity to lead to disastrous real world consequences. The case for acting now is

not diminished by a plurality of expert judgements about probabilities and consequences.

Acknowledgements

We would like to thank James Annan, Philip Brohan, Richard Chandler, Peter Challenor, Mat Collins, Tamsin Edwards, Lindsay Lee, Reto Knutti, and Danny Williamson for very helpful comments on an earlier draft of this review, and absolve them completely of any responsibility for the views expressed herein.

Glossary

These definitions are by necessity very crude. In some cases more detail is given in the text; in most, it would be advisable to consult a standard source such as the IPCC glossary or the WMO glossary. Statistical definitions introduced or clarified in this paper are indicated with asterisks.

Anomaly model* A statistical climate model in which the dominant limitation of the simulator is mis-location of your climate.

Best parameter model* A generalisation of the anomaly model, in which the limitations of the simulator involve the location, size, and shape of your climate.

Climate* your distribution for weather, represented as a multivariate spatial-temporal process (inherently subjective).

Climate simulator* Computer code which maps forcing into weather, given also an initial condition and parameter values.

Climate sensitivity The difference in equilibrium global mean temperature between two time-slice experiments, one with forcing using pre-industrial levels of CO_2, and a second with forcing using double CO_2.

Control run A long time-slice experiment, typically with pre-industrial forcing.

Detection and attribution (D&A) Examining the role of anthropogenic effects in C20th weather patterns.

Discrepancy* In the best parameter model, the uncertain additive difference between the simulator output at its best parameterisation and your climate.

Emulator* Statistical model for a simulator, allowing prediction of the simulator output at untried values of the parameters.

Forcing Boundary conditions for the simulator, usually implemented in terms of radiation (solar forcing and greenhouse gas forcing) and optical depth (aerosol forcing).

GCM, AO-GCM, ESM Acronyms for large climate simulators: General Circulation (or Global Climate) Model, Atmosphere-Ocean GCM, Earth System Model.

Grid-scale The horizontal length of a spatial grid-cell in a finite difference approximation to the underlying partial differential equations. About 100 km for current large climate simulators.

History Matching* Statistical approach for ruling out poor choices of the simulator parameters.

Internal variability The inherent variability of the weather within a climate simulator.

Over-fitting* A consequence of using the anomaly model, resulting in too-small projection uncertainties.

Over-tuning* Tuning which emphasises a match with the histogram of C20th weather.

Parameters Adjustable coefficients in the simulator which represent sub-grid-scale processes and incompletely understood processes.

Projection A climate prediction along a specified scenario.

Scenario A future described by specified forcing, representative of trends in population, economics, technology, and policy interventions.

Spin-up A time-slice experiment in which the simulator 'forgets' its initial condition.

Statistical climate model* A statistical framework designed to simplify the process of specifying your climate.

Sub-grid-scale processes Processes in the mathematical model with length-scales comparable to or smaller than the grid-scale of the simulator.

Time slice experiment Experiment where the simulator is given constant or periodic forcing, and run at least until its output stabilises.

Tuning Choosing a preferred value for the simulator's parameters.

Weather Measurable aspects of the ambient atmosphere, notably temperature, precipitation, and wind-speed.

References

C. Donald Ahrens, 2000. *Meteorology Today: An Introduction to Weather, the Climate, and the Environment.* Pacific Grove, CA: Brooks/Cole, 6th edition.

A. Arakawa, 1997. Adjustment mechanisms in atmospheric models. *Journal of the Meteorological Society of Japan*, **75**(1B), 155–179.

L.S. Bastos and A. O'Hagan, 2009. Diagnostics for Gaussian Process emulators. *Technometrics*, **51**(4), 425–438.

G.E.P. Box, 1980. Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Series A*, **143**(4), 383–430. With discussion.

C. Chatfield, 2004. *The Analysis of Time Series.* Boca Raton, FL: Chapman & Hall/CRC.

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.

P.S. Craig, M. Goldstein, A.H. Seheult, and J.A. Smith, 1997. Pressure matching for hydrocarbon reservoirs: A case study in the use of Bayes Linear strategies for large computer experiments. In C. Gatsonis, J.S. Hodges, R.E. Kass, R. McCulloch, P. Rossi, and N.D. Singpurwalla, editors, *Case Studies in Bayesian Statistics III*, pages 37–87. New York: Springer-Verlag. With discussion.

J.A. Cumming and M. Goldstein, 2009. Small sample Bayesian designs for complex high-dimensional models based on information gained using fast approximations. *Technometrics*, **51**(4), 377–388.

G. Danabasoglu, R. Ferrari, and J.C. McWilliams, 2008. Sensitivity of an ocean general circulation model to a parameterization of near-surface eddy fluxes. *Journal of Climate*, **21**, 1192–1208.

N. Edwards, D. Cameron, and J.C. Rougier, 2011. Precalibrating an intermediate complexity climate model. *Climate Dynamics*, **37**, 1469–1482.

A.I.J. Forrester, A. Sóbester, and A.J. Keane, 2008. *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester, UK: John Wiley & Sons.

P. Forster, V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D.W. Fahey, J. Haywood, J. Lean, D.C. Lowe, G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schulz, and R. Van Dorland, 2007. Changes in atmospheric constituents and in radiative forcing. In Solomon *et al.* (2007), chapter 2.

A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin, 2003. *Bayesian Data Analysis*. Boca Raton, Florida: Chapman and Hall/CRC, 2nd edition.

P.R. Gent, G. Danabasoglu, L.J. Donner, M.M. Holland, E.C. Hunke, S.R. Jayne, D.M. Lawrence, R.B. Neale, P.J. Rasch, M. Vertenstein, P.H. Worley, Z.-L. Yang, and M. Zhang, 2011. The Community Climate System Model Version 4. *Journal of Climate*, **24**, 4973–4991.

R.M. Gladstone, V. Lee, J.C. Rougier, A.J. Payne, H. Hellmer, A. Le Brocq, A. Shepherd, T.L. Edwards, J. Gregory, and S.L. Cornford, 2012. Calibrated prediction of Pine Island Glacier retreat during the 21st and 22nd centuries with a coupled flowline model. *Earth and Planetary Science Letters*, **333–334**, 191–199.

M. Goldstein and J.C. Rougier, 2004. Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM Journal on Scientific Computing*, **26**(2), 467–487.

M. Goldstein and J.C. Rougier, 2006. Bayes linear calibrated prediction for complex systems. *Journal of the American Statistical Association*, **101**, 1132–1143.

M. Goldstein and J.C. Rougier, 2009. Reified Bayesian modelling and inference for physical systems. *Journal of Statistical Planning and Inference*, **139**, 1221–1239. With discussion, pp. 1243–1256.

M. Goldstein and D.A. Wooff, 2007. *Bayes Linear Statistics: Theory & Methods*. John Wiley & Sons, Chichester, UK.

G.H. Golub and C.F. Van Loan, 1996. *Matrix Computations*. Baltimore: Johns Hopkins University Press, 3rd revised edition.

P. Guttorp, 2014. Statistics and climate. In S.E. Fienberg, editor, *Annual Review of Statistics and its Application*. Annual Reviews, Palo Alto, CA, USA.

I. Hacking, 2001. *An Introduction to Probability and Inductive Logic*. Cambridge, UK: Cambridge University Press.

K. Hasselmann, 1976. Stochastic climate models part I: Theory. *Tellus*, **28**, 473–485.

G. Hegerl and F. Zwiers, 2011. Use of models in detection and attribution of climate change. *WIREs Climate Change*, **2**, 570–591.

G.C. Hegerl, F.W. Zwiers, P. Braconnot, N.P. Gillett, Y. Luo, J.A. Marengo Orsini, N. Nicholls, J.E. Penner, and P.A. Stott, 2007. Understanding and attributing climate change. In Solomon *et al.* (2007), chapter 9.

P.J. Irvine, A. Ridgwell, and D.J. Lunt, 2011. Climatic effects of surface albedo geoengineering. *Journal of Geophysical Research*, **116**, D24112.

P.D. Jones *et al.*, 2009. High-resolution palaeoclimatology of the last millennium: a review of current status and future prospects. *The Holocene*, **19**(1), 3–49.

M.C. Kennedy and A. O'Hagan, 2001. Bayesian calibration of computer models. *Journal of the Royal Statistical Society, Series B*, **63**, 425–450. With discussion, pp. 450–464.

R. Knutti, D. Masson, and A. Gettelman, 2013. Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, **40**, 1194–1199.

D. Maraun *et al.*, 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, **48**, RG3003.

K.V. Mardia, J.T. Kent, and J.M. Bibby, 1979. *Multivariate Analysis*. Harcourt Brace & Co., London.

D. Masson and R. Knutti, 2011. Climate model genealogy. *Geophysical Research Letters*, **38**, L08703.

T. Mauritsen, 2012. Tuning the climate of a global model. *Journal of Advances in Modeling Earth Systems*, **4**, M00A01.

K. McGuffie and A. Henderson-Sellers, 2005. *A Climate Modelling Primer*. Chichester: John Wiley & Sons, 3rd edition.

D.J. McNeall, P.G. Challenor, J.R. Gattiker, and E.J. Stone, 2013. The potential of an observational data set for calibration of a computationally expensive computer model. *Geoscietific Model Development Discussions*, **6**, 2369–2401.

J.C. McWilliams, 2007. Irreducible imprecision in atmospheric and oceanic simulations. *Proceedings of the National Academy of Sciences*, **104**(21), 8709–8713.

J. Murphy, R. Clark, M. Collins, C. Jackson, M. Rodwell, J.C. Rougier, B. Sanderson, D. Sexton, and T. Yokohata. Perturbed parameter ensembles as a tool for sampling model uncertainties and making climate projections. In *Proceedings of ECMWF Workshop on Model Uncertainty, 20-24 June 2011*, pages 183–208, 2011. Available online, `http://www.ecmwf.int/publications/library/ecpublications/_pdf/workshop/2011/Model_uncertainty/Murphy.pdf`.

J.M. Murphy, D.M.H. Sexton, D.N. Barnett, G.S. Jones, M.J. Webb, M. Collins, and D.A. Stainforth, 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772.

R.J. Nicholls, 2011. Planning for the impacts of sea-level rise. *Oceanography*, **24** (2), 144–157.

N. Oreskes and E.M. Conway, 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press, New York, USA.

M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, and C.E. Hanson, editors, 2007. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK.

J.P. Peixoto and A.H. Oort, 1992. *Physics of Climate*. New York: Springer.

V. Petoukhov, A. Ganapolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf, 2000. CLIMBER-2: a climate system model of intermediate complexity. Part I: model description and performance for present climate. *Climate Dynamics*, **16**, 1–17.

D.A. Randall, R.A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R.J. Stouffer, A. Sumi, and K.E. Taylor, 2007. Climate models and their evaluation. In Solomon *et al.* (2007), chapter 8.

J. Rockström *et al.*, 2009. A safe operating space for humanity. *Nature*, **461**, 472–475.

J.C. Rougier, 2007. Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, **81**, 247–264.

J.C. Rougier, 2008a. Annotated bibliography: Climate change detection and attribution. *The ISBA Bulletin*, **15**(4). Available online, `http://bayesian.org/sites/default/files/fm/bulletins/0812.pdf`.

J.C. Rougier, 2008b. Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, **17**(4), 827–843.

J.C. Rougier, M. Goldstein, and L. House, 2013. Second-order exchangeability analysis for multi-model ensembles. *Journal of the American Statistical Association*. Forthcoming; currently available at `http://amstat.tandfonline.com/doi/full/10.1080/01621459.2013.802963`.

J.C. Rougier, S. Guillas, A. Maute, and A. Richmond, 2009a. Expert knowledge and multivariate emulation: The Thermosphere-Ionosphere Electrodynamics General Circulation Model (TIE-GCM). *Technometrics*, **51**(4), 414–424.

J.C Rougier and D.M.H. Sexton, 2007. Inference in ensemble experiments. *Philosophical Transactions of the Royal Society, Series A*, **365**, 2133–2143.

J.C. Rougier, D.M.H. Sexton, J.M. Murphy, and D. Stainforth, 2009b. Analysing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. *Journal of Climate*, **22**(13), 3540–3557.

D.B. Rubin, 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, **12**(4), 1151–1172.

J.D. Salt, 2008. The seven habits of highly defective simulation projects. *Journal of Simulation*, **2**, 155–161.

B. Sansó, C. Forest, and D. Zantedeschi, 2008. Inferring climate system properties using a computer model. *Bayesian Analysis*, **3**(1), 1–38. With discussion, pp. 39–62.

T.J. Santner, B.J. Williams, and W.I. Notz, 2003. *The Design and Analysis of Computer Experiments*. New York: Springer.

D.M.H. Sexton, J. Murphy, M. Collins, and M.J. Webb, 2012. Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Climate Dynamics*, **11–12**, 2513–2542. DOI:10.1007/s00382-011-1208-9.

S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, editors, 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 University Press, Cambridge, UK.

Roland B. Stull, 2000. *Meteorology for Scientists and Engineers*. Pacific Grove, CA: Brooks/Cole, 2nd edition.

R. Tokmakian and P. Challenor, 2013. Uncertainty in modeled upper ocean heat content change. *Climate Dynamics*, **In press**. Available on-line, DOI:10.1007/s00382-013-1709-9.

P. Valdes, 2011. Built for stability. *Nature Geoscience*, **4**, 414–416.

I. Vernon, M. Goldstein, and R.G. Bower, 2010. Galaxy formation: a Bayesian uncertainty analysis. *Bayesian Analysis*, **5**(4), 619–670.

Hans von Storch and Francis W. Zwiers, 1999. *Statistical Analysis in Climate Research*. Cambridge: Cambridge University Press.

M. Watanabe *et al.*, 2010. Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate*, **23**, 6312–6335.

D. Williamson and M. Goldstein, 2012. Bayesian policy support for adaptive strategies using computer models for complex physical systems. *Journal of the Operational Research Society*, **63**, 1021–1033.

D. Williamson, M. Goldstein, and A. Blaker, 2012. Fast linked analyses for scenario-based hierarchies. *Applied Statistics*, **61**(5), 665–691.