Efficient estimation and ensemble generation in climate modelling

By J. D. Annan* and J. C. Hargreaves

FRCGC/JAMSTEC, 3173-25 Showa-machi Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan

In this paper, we review progress towards efficiently estimating parameters in climate models. Since the general problem is inherently intractable, a range of approximations and heuristic methods have been proposed. Simple Monte Carlo sampling methods, although easy to implement and very flexible, are rather inefficient, making implementation possible only in the very simplest models. More sophisticated methods based on random walks and gradient-descent methods can provide more efficient solutions, but it is often unclear how to extract probabilistic information from such methods and the computational costs are still generally too high for their application to state-of-the-art general circulation models (GCMs). The ensemble Kalman filter is an efficient Monte Carlo approximation which is optimal for linear problems, but we show here how its accuracy can degrade in nonlinear applications. Methods based on particle filtering may provide a solution to this problem but have yet to be studied in any detail in the realm of climate models. Statistical emulators show great promise for future research and their computational speed would eliminate much of the need for efficient sampling techniques. However, emulation of a full GCM has yet to be achieved and the construction of such represents a substantial computational task in itself.

Keywords: probabilistic climate prediction; climate modelling; ensemble Kalman filter

1. Introduction

Model projections of anthropogenically forced climate change are sensitive to many aspects of model construction, in particular the values of parameters which are not well determined by direct observations or theoretical arguments but which directly impact on the future trajectory of model simulations. Although short-term numerical weather prediction is essentially an initial-value problem, the behaviour of a coupled atmosphere–ocean model in response to a forcing scenario on climatological (multidecadal and upward) timescales is determined much more by the details of its parameterizations rather than the initial state. Therefore, much attention has recently been given to the parameter estimation problem in this context (e.g. Forest *et al.* 2002; Gregory *et al.* 2002; Knutti *et al.*

2002; Hargreaves *et al.* 2004). One initial aim may be to simply try to improve the realism of the model, in order to give the most accurate forecast possible. This model tuning is often seen as part of the model-building process, and has generally been performed in the rather ad hoc and inefficient manner of 'hand-tuning' by the model developers. More generally, we would also like to estimate the uncertainty surrounding this central forecast, typically through an ensemble of model integrations. The methods initially used for estimation and ensemble generation were generally rather simple applications of Monte Carlo sampling (discussed in §2), and thus were computationally inefficient and could only be applied to the cheapest models. Here, we discuss progress in this field.

The computational problem amounts to an application of Bayes' theorem: given prior belief $f(x)$ over any uncertain parameters x, and observational data o which gives rise to a likelihood function $f(o|x)$, how can we efficiently estimate the posterior probability density function (PDF) $f(x|o) = f(o|x)f(x)/f(o)$? In fact, we are unlikely to be directly interested in the full posterior PDF, but will in general hope to calculate statistics of interest such as the expected value of some function g of the model output over the posterior PDF via $E(g(x)) = \int (g(x)f(x|o))$, or marginal and cumulative distributions thereof. Typical examples might be to estimate probabilistic percentiles for climate sensitivity, or calculate a probability for global temperature rising by more than a certain threshold under a specific emissions scenario.

Researchers have used a range of different observational data, including observations of recent and long-term climate change in conjunction with computationally cheap energy-balance Earth system models (Andronova & Schlesinger 2001; Forest *et al.* 2002; Knutti *et al.* 2002), or the use of the quasi-steady-state climates in conjunction with more state-of-the-art general circulation models (GCMs; Murphy *et al.* 2004; Annan *et al.* 2005a) to estimate equilibrium climate changes. In idealized situations where we can assume the model is perfect (i.e. there are correct parameter values to estimate, for which the model behaviour perfectly mirrors the real world), the formulation of the problem is straightforward, and this is a natural environment in which to develop and test methods.

However, the perfect model assumption is rarely tenable in practice. No matter how well tuned, all climate models are readily distinguished from reality, with systematic biases in climate variables, and limited representation of climatic phenomena. The parameters may have no readily identifiable counterparts in reality and this, combined with the need to compensate for missing processes and inevitable biases, means that the concept of a 'correct' parameter value does not apply. Thus, the concept of the 'best input' (Rougier 2007) is introduced, which refers to the input set for which the model provides the 'best' overall simulation of past and future climate. If the model was perfect, the correct parameter values would of course fulfil this role. Estimating this, best input is complicated not only by the fact that the residual between the model and the data is not merely due to observational uncertainty, but also our uncertainty concerning model inadequacy (the discrepancy between the 'best model' and the real climate). This uncertain discrepancy term adds to the observational uncertainty in broadening the posterior distribution. These issues are discussed more fully in Rougier (2007). In the applications presented here, a very simple approach to model inadequacy has been adopted, based on the variance of the

data fields. More detailed specification of this term would help to provide authoritative and credible predictions of climate change.

It is worth mentioning at the outset that there is no reasonable possibility of finding a general solution to the estimation problem that will work efficiently in all applications. The general parameter estimation problem can easily be seen to contain generic optimization problems which are at least non-deterministic polynomial-time hard (Garey & Johnson 1979). For example, an efficient and general parameter estimation method would also be capable of solving quadratic programming problems, which is widely believed to be computationally intractable. Thus, we do not expect to find such a general solution, and our focus is on developing methods which aim to achieve an acceptable level of precision, by making reasonable approximations and taking advantage of any structure that exists in the climate prediction problem.

2. Numerical integration and Monte Carlo sampling

In very low-dimensional problems, the posterior PDF can be approximated by deterministic numerical integration, for example, via Gaussian quadrature. This requires a multifactorial sampling of all adjustable parameters. Clearly, this type of approach is only feasible when the number of dimensions is rather low and/or the model is computationally cheap, since the number of integrations required increases exponentially with the number of dimensions. For example, Forest *et al.* (2000) used 61 integrations to sample two parameters at several levels in an energy balance model. In a more recent example, Stainforth *et al.* (2005) undertook a multifactorial sampling of six parameters of an atmospheric general circulation model (AGCM) at only three values each, and this experiment required massive computational resources. Using the same approach to sample even 10 parameters at 5 levels each would require resources some 10^5 times greater still.

Monte Carlo methods (Hammersley & Handscomb 1964) provide an alternative randomized approach which has been widely used: instead of attempting to exhaustively integrate over the whole parameter space, we can simply draw a set of samples x_i from the prior (note that we are implicitly assuming that this itself is a straightforward operation), calculate their likelihood-based weights $w_i \equiv f(o|x_i)$ and then approximate the required integrals in the obvious manner, e.g.

$$E(g(x)) = \int g(x)f(x|o) \simeq \frac{\sum(w_i g(x_i))}{\sum w_i}.$$

This approach is trivial to implement and very flexible, and has been widely used in climate science (e.g. Knutti *et al.* 2002; Schneider von Deimling *et al.* 2006). A minor variant (rejection sampling) is to reject or accept each sample in a probabilistic manner based on the weight, which produces a smaller posterior ensemble (of accepted samples) for further calculation, and eliminates the need to keep track of a weight for each member. The main drawback of these approaches is that if the weights are low across a large proportion of the prior space, then it may take a large number of samples to populate the posterior and achieve reasonable convergence. In particular, the vast majority of samples may
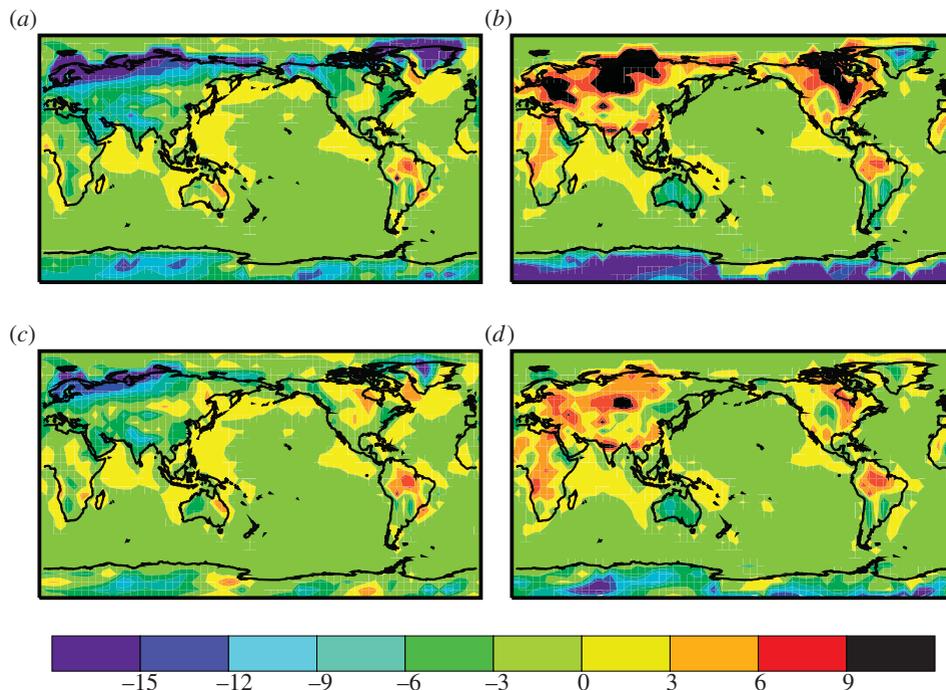
Figure 1. Results obtained by tuning a low-resolution AGCM to NCEP data, allowing five parameters to vary simultaneously. (*a,b*) DJF control SAT—NCEP (K) and JJA control SAT—NCEP (K) show errors in the control run. (*c,d*) DJF tuned SAT—NCEP (K) and JJA tuned SAT—NCEP (K) show greatly reduced errors in posterior ensemble mean. Values are plotted in kelvin. Adapted from Annan *et al.* (2005*b*).

be given negligible weights if the prior is substantially more vague than the posterior, and this situation is likely to arise in high-dimensional problems, especially when (as is common in climate science) the prior is deliberately chosen to be on the vague side so as to minimize the risk of double-counting of data. Latin Hypercube sampling (McKay *et al.* 1979) is generally a rather more efficient approach, but only by a modest factor which is unlikely to completely overcome this disadvantage. For example, in some work which we describe below, a 1000 member Latin Hypercube experiment failed to find a single sample in the support of the posterior distribution defined by the problem (in other words, rejection sampling an ensemble of this size would have almost certainly generated the empty set). Importance sampling, in which samples are drawn not from the prior but from some 'proposal distribution' which is thought to more closely approximate the posterior, could potentially lead to further improvements, but it is not clear how such a proposal distribution could be found.

3. Random walk and gradient-descent methods

There are a wide variety of methods of varying generality, sophistication and efficiency based on random walks or gradient descent in parameter space. Perhaps the simplest is the Monte Carlo Markov Chain (Metropolis *et al.* 1953;
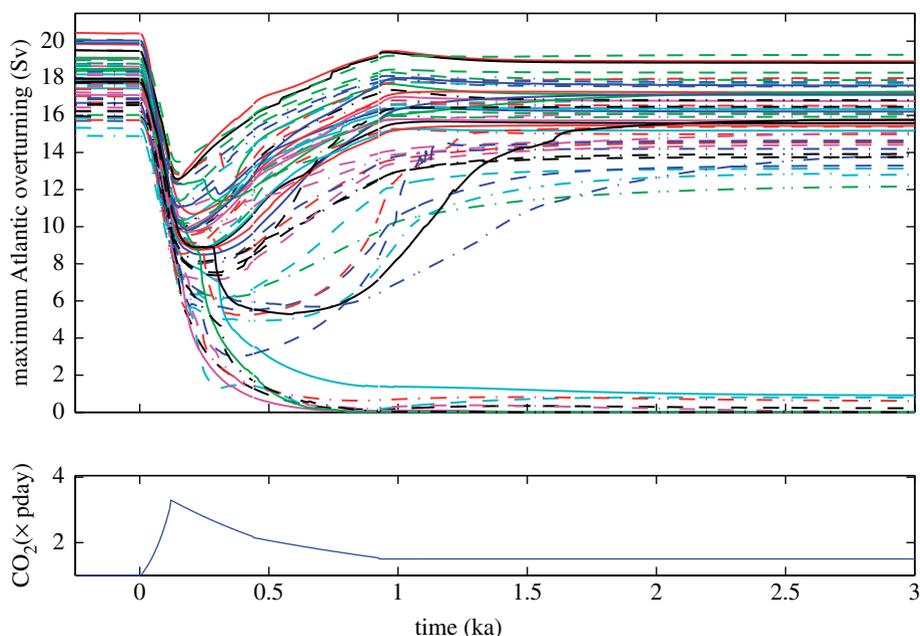
Figure 2. Ensemble response of meridional overturning under a greenhouse gas forcing scenario. Each line represents one of the 54 ensemble members. The probability of a long-term shutdown of the overturning depends strongly on several factors which are poorly understood at present. Adapted from Hargreaves & Annan (2005).

Hargreaves & Annan 2002; Tarasov & Peltier 2005). This method undertakes a directed random walk in parameter space, in which trial steps are generated, and they are accepted or rejected probabilistically depending on how the likelihood of the new parameter values compares to that of the last-visited point. The distribution of points visited in parameter space converges to the posterior PDF. This method does not suffer from the curse of dimensionality, and is also highly flexible and simple to implement, requiring no special assumptions or approximations with the problem. However, since the method is inherently sequential and generally requires (at a minimum) thousands of integrations to converge, it again seems to be limited to computationally cheap models which can be integrated in a matter of seconds.

A range of more sophisticated heuristic approaches such as very fast simulated annealing (Jackson *et al.* 2004), genetic algorithms (Price *et al.* 2006) and Proximal-ACCPM (Beltran *et al.* 2005) have also been applied to optimization and estimation with intermediate complexity climate models. These methods have significantly lower computational cost which allows application to at least intermediate complexity models (with integration times of around 1 h). Perhaps the most efficient at present is the Proximal-ACCPM method, which can achieve good results with as few as 20 full model integrations, but it is unclear as to how probabilistic information can be extracted from this method which directly searches for the optimal (maximum-likelihood) parameter set.

The use of an adjoint, which efficiently calculates the gradient of the cost function with respect to all of the control variables, can provide a highly efficient multivariate parameter estimation method in some cases (e.g. Hall *et al.* 1982;

van Scheltinga & Dijkstra 2005). Moreover, it also generates a Jacobian of sensitivities which (at least in the near-linear case) can be used to provide probabilistic information. Although the construction of the adjoint code may be a rather complex task (even similar in magnitude to the creation of the model itself), automatic tools are available to help with this, and in some cases, the adjoint may already exist if a version of the climate model is also used for numerical weather prediction. However, the application of this method to climatological estimation in complex models is problematic, due to the chaotic dynamics of the models (Lea *et al*. 2000, 2002). In general, for a chaotic model integrated towards a climatological steady state, the gradient of the cost function is undefined everywhere, due to the fact that infinitesimal initial perturbations will grow exponentially over a long integration time. This remains the case even when the climatological response of the model to finite perturbations is close to linear. Therefore, various approximations and simplifications have to be developed for particular applications in order to generate usable results. These may include, for example, averaging over an ensemble of shorter integrations, or even direct modifications to the adjoint code through an artificially large diffusion coefficient or the use of a low-resolution non-chaotic approximation to the full model. Furthermore, the Jacobian of sensitivities which an adjoint generates are only locally valid, so in the nonlinear case there is again no clear probabilistic interpretation of this output. No general solutions are known for these problems, and these obstacles appear to have inhibited research in this direction.

4. Ensemble Kalman filter

The ensemble Kalman filter (EnKF) was introduced by Evensen (1994) as an efficient Monte Carlo approximation to the Kalman filter (Kalman 1960), which is the optimal filter for linear systems. The method is now in increasingly widespread use for state estimation problems in initial-value problems such as numerical weather prediction and ocean forecasting. A thorough description of the theory and basic methodology together with a survey of recent applications is provided in Evensen (2003). Although the EnKF has generally been used for sequential initial state estimation, parameter estimation can readily be included in the same framework, by the means of state space augmentation (Derber 1989; Anderson 2001). The principle here is that the parameters can be considered to be part of the model state alongside the conventional variables, and then the covariances sampled by the ensemble members can be used directly to update parameters in exactly the same manner as for the state variables.

The Kalman filter equations are optimal for linear systems, and this method does not suffer from the 'curse of dimensionality'. Later, we investigate the performance in some simple nonlinear cases. The method is trivially parallelizable, at least under the generally realistic assumption that the dominant cost is in the integration of $O(100)$ models.

This method has recently been used to solve the parameter estimation problem in a range of models such as the Lorenz model (Annan & Hargreaves 2004), a simple Earth system model (Hargreaves *et al*. 2004), a simplified primitive equation AGCM (Annan *et al*. 2005*b*) and even a modern state-of-the-art AGCM
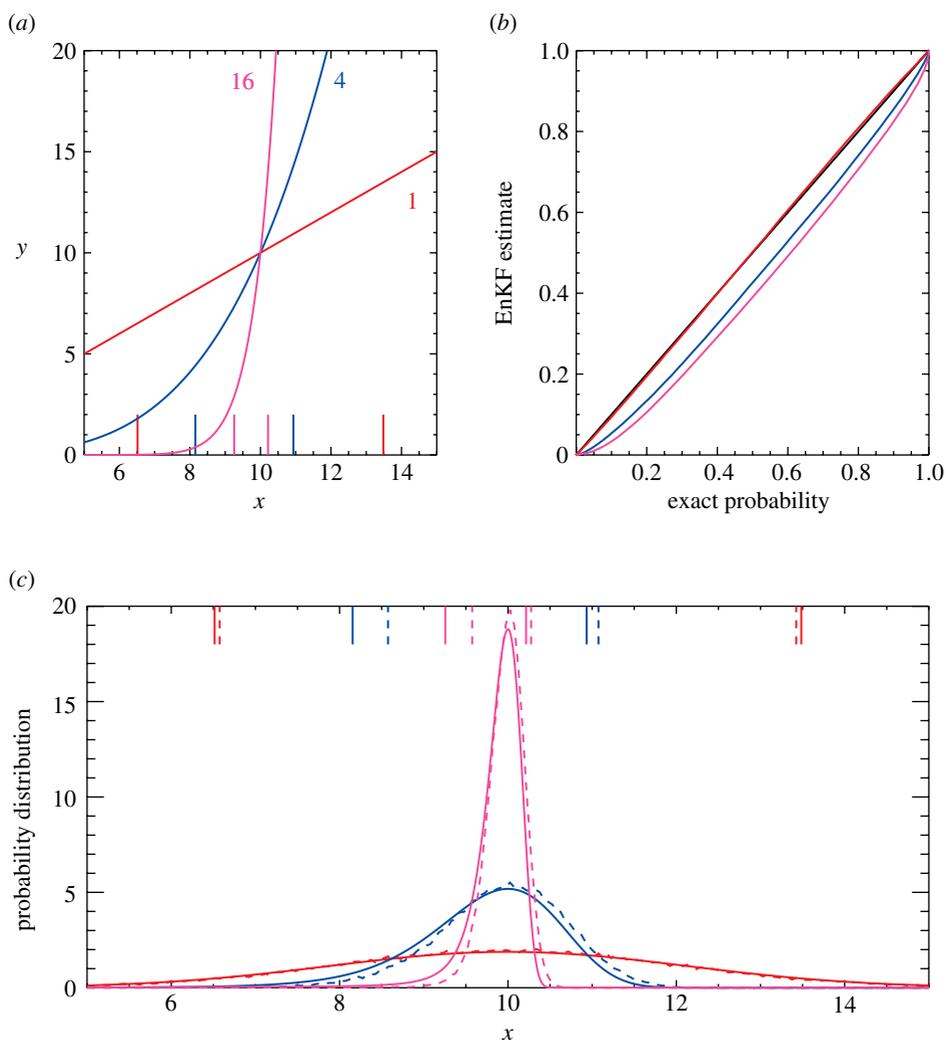
Figure 3. Illustrating the influence of nonlinearity on the EnKF performance. (*a*) Model function showing the function $y = 10 \times (x/10)^k$ for three values of k: red, $k=1$; blue, $k=4$; magenta, $k=16$. (*b*) The probability plot of EnKF output versus correct solution (the red line overlies the exact $x=y$ line, shown in black). (*c*) Distributions showing the correct posterior PDFs (solid lines) with EnKF estimates (dashed lines). Colour coding indicates k (as shown in *a*). (*a,c*) Vertical lines indicate the 5 and 95% thresholds for the posterior PDFs.

with slab ocean, albeit at low resolution (Annan *et al*. 2005*a*). Figure 1 illustrates some results obtained with a simplified T21L7 AGCM (a version of the IGCM of Forster *et al*. 2000). The seasonally averaged differences between the control run and the target of NCEP seasonal surface air temperature fields are shown on at the top. Below, the ensemble mean shows a marked improvement after five parameters were tuned. Precipitation fields were also used as a target in this experiment, but tuning was unable to significantly improve the model fields (not shown), thus demonstrating that there were substantial structural problems and motivating further model development.

Figure 2 illustrates an application of the EnKF to the simultaneous estimation of 12 parameters in an intermediate complexity EMIC in order to investigate the stability and predictability of the long-term response of its meridional overturning circulation to various emissions scenarios (Hargreaves & Annan 2005). Despite sampling a wide range of parameter values, all the model climatologies were of a reasonably high quality in comparison to the observations used, and the root mean square (r.m.s.) error of the ensemble mean was also very similar to the results obtained by other optimization methods which have been applied to this model (Beltran *et al.* 2005; Price *et al.* submitted). Minor differences in model version and problem specification (such as uniform versus Gaussian parameter priors) prevent a direct comparison, but clearly all methods were working well when applied to this problem. In contrast, a 1000 member Latin Hypercube experiment (Edwards & Marsh 2005) with the same model failed to find a single solution comparable to any of these methods, illustrating the limitations of more naive methods. That is, it failed to produce any samples from the posterior defined by our problem specification.

The EnKF approach is only formally correct in the linear case. Of course, the linearity here refers not to the model itself (which in general will exhibit chaotic dynamics) but to the climate of the model, as a function of the control variables. Despite this theoretical limitation, it appears to produce surprisingly good solutions to problems where the exact solution is known to be strongly non-Gaussian (e.g. Zhou *et al.* 2006).

It is hard to validate the method in complex applications where the correct solution is not known, but we can readily investigate the effect of nonlinearity in idealized cases. For example, Annan *et al.* (2005*b*) used a simple single-variable problem with a low degree of nonlinearity to demonstrate the benefit of the iterative procedure we have developed. Here, we extend this idea to investigate the degradation of the technique with increasing nonlinearity. Our toy model here takes a single input parameter x for which we have a prior estimate $x = 10 \pm 3$ (the indicated uncertainties are all one standard deviation, Gaussian and assumed independent) and generates the output y via the equation

$$y = 10 \times \left(\frac{x}{10} \right)^k, \tag{4.1}$$

where k will vary between experiments.

We have a single observation $y_o = 10 \pm 3$ with which we aim to improve our estimate of x. For all k, the posterior maximum-likelihood value is given by $(x, y) = (10, 10)$, but for $k > 1$, the model is nonlinear and the posterior is skewed. The results, shown in figure 3, are the average of 10 experiments with 10 000 ensemble members each, so as to focus solely on the influence of nonlinearity. For the case $k = 1$, the problem is linear and trivial, and the EnKF generates an accurate solution. As k increases, the posterior PDF becomes more strongly skewed and the accuracy of the EnKF solution (which is more symmetrical in shape) decreases. For the case $k = 16$, the bias in the PDF is comparable in magnitude (and of course additional to) the 2σ uncertainty that would also arise due to sampling error with use of a typical practical ensemble size of 100 members, as indicated by the standard formula for the sampling error of an event

of probability p: $\sigma^2 = p(1-p)/N$. As was noted by Zhou *et al.* (2006), the EnKF generally gives fairly good estimates for the mean even for nonlinear problems, with the accuracy of higher-order moments being more noticeably degraded.

We can evaluate the nonlinearity in the posterior automatically, by comparing the analysis output to the model output at the subsequent iteration. The EnKF automatically generates an ensemble of 'analysis' values (x_i^a, y_i^a) from linear combinations of the previous iteration's 'forecast' (x_i^f, y_i^f). If the model is linear, the (x_i^a, y_i^a) pairs should be in perfect balance. Therefore, we examine the r.m.s. error function

$$R = \sqrt{\frac{\sum_i \left(y_i^a(t) - y_i^f(t+1)\right)^2}{n}},$$

where $y_i^a(t)$ is the analysed output for y from the tth iteration and $y_i^f(t+1)$ is the forecast of y for the subsequent iteration. For the linear case $k=1$, R is essentially the machine precision. For $k=4$, R is approximately 0.30, and for $k=16$ it grows to around 0.45. The bias of the analysis versus forecast is negligible in all cases. These results suggest that it might be possible to diagnose the presence of nonlinearity and determine to what extent the results from particular experiments can be considered reliable. However, practical considerations may limit the integration time to the point that the model's internal variability and lack of true convergence contaminates the calculation.

For strongly bimodal PDFs, such as would be generated by the function $y = x^2$ with an observation of $y_o = 10$ and a prior that assigns significant probability to both high probability regions $x = \pm\sqrt{10}$, the EnKF method fails to converge, with the main mass of the ensemble instead oscillating between the two modes of the posterior. Such behaviour is easy to demonstrate in artificial situations, but has not yet been observed in any of the practical experiments we have undertaken with realistic models.

5. Particle filters

Particle filtering is a sequential Monte Carlo method in which the PDF is, as with the EnKF, sampled by an ensemble of models (particles). Unlike the EnKF, the particles are not directly repositioned in light of observational data, but instead likelihood-based weights are assigned to them. To compensate for the inevitable concentration of weights on very few members, regular resampling is generally required. For linear problems, the unequal weights result in the method being generally less accurate than an EnKF of similar size, but it has the advantage of being able in principle to handle nonlinear problems with no loss of precision. Zhou *et al.* (2006) illustrate the superior accuracy of a particle filter method over the EnKF for a highly nonlinear problem when the ensemble size is not a limiting factor. Particle filtering methods are not ideally suited for parameter estimation, since there is no direct means by which replicated particles (generated in the resampling step) become separated. However, there are techniques such as jittered resampling Liu & West (2001) which address this problem. The hybrid kernel resampling method of Anderson & Anderson (1999) offers another related avenue by which nonlinearity can be handled within the

particle/Kalman filter framework, without the need for such a large ensemble as the particle filter typically requires. Such methods have not been widely tested.

6. Emulators

A statistical emulator is a computationally efficient approximation to a complex computer code. A simple approach based on a linear interpolation was presented by Murphy *et al.* (2004), and a more sophisticated version is described by Rougier & Sexton (2007). Owing to the computational efficiency of the emulator, it is possible to evaluate its PDF with high precision, even using simple Monte Carlo sampling. This approach appears to have great potential for more accurate evaluation of probabilities, although the construction of the emulator is itself a complex task. Success in this research would have strong implications for the need to find efficient optimization techniques, but the ability to emulate the full multivariate output (i.e. fields of climatological variables) of a complex GCM, which would be needed for calibration and prediction of spatially resolved climate change, has not yet been achieved.

7. Conclusions

Most methods of uncertainty analysis in common use in climate science rely on large numbers of model evaluations sampled directly from prior distributions on the input parameters, and thus can only be applied to low-dimensional problems and simplified models. Even within this setting, significant improvements in efficiency can often be achieved by careful attention to sampling strategy such as the use of a Latin Hypercube.

More sophisticated estimation and data assimilation methods are based on gradient descent and other heuristics attempt to sample more effectively from the posterior (defined by some measure of fit to observational data). The methods are generally flexible, optimal (in the sense of correctly solving an arbitrary problem, given adequate computational resources) and simple to implement, but are still usually too expensive for application to state-of-the-art models. The EnKF is highly parallelizable and requires relatively few model integrations, but it is only approximate in the nonlinear case, and the validity and accuracy of its approximation in real applications remains rather uncertain. However, it has proven very effective in at least finding well-calibrated solutions across a wide range of applications, and the more general particle filter or variants such as kernel resampling may have the potential to address its main limitations. The Bayesian emulator approach demonstrated by Rougier & Sexton (2007) appears to be a highly promising alternative for the future.

References

Anderson, J. L. 2001 An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **129**, 2884–2902. (doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2)

Anderson, J. L. & Anderson, S. L. 1999 A Monte Carlo implementation o the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Weather Rev.* **127**, 2741–2758. (doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2)

Andronova, N. G. & Schlesinger, M. E. 2001 Objective estimation of the probability density function for climate sensitivity. *J. Geophys. Res.* **108**, 22 605–22 611. (doi:10.1029/2000JD000259)

Annan, J. D. & Hargreaves, J. C. 2004 Efficient parameter estimation for a highly chaotic system. *Tellus A* **56**, 520–526. (doi:10.1111/j.1600-0870.2004.00073.x)

Annan, J. D., Hargreaves, J. C., Ohgaito, R., Abe-Ouchi, A. & Emori, S. 2005*a* Efficiently constraining climate sensitivity with paleoclimate simulations. *SOLA* **1**, 181–184. (doi:10.2151/sola.2005-047)

Annan, J. D., Lunt, D. J., Hargreaves, J. C. & Valdes, P. J. 2005*b* Parameter estimation in an atmospheric GCM. *Nonlin. Process. Geophys.* **12**, 363–371.

Beltran, C., Edwards, N. R., Haurie, A., Vial, J.-P. & Zachary, D. S. 2005 Oracle-based optimization applied to climate model calibration. *Environ. Model. Assess.* **11**, 31–43. (doi:10.1007/s10666-005-9024-4)

Derber, J. 1989 A variational continuous assimilation scheme. *Mon. Weather Rev.* **117**, 2437–2446. (doi:10.1175/1520-0493(1989)117<2437:AVCAT>2.0.CO;2)

Edwards, N. R. & Marsh, R. 2005 Uncertainties due to transport-parameter sensitivity in an efficient 3-D ocean-climate model. *Clim. Dynam.* **24**, 415–433. (doi:10.1007/s00382-004-0508-8)

Evensen, G. 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10 143–10 162. (doi:10.1029/94JC00572)

Evensen, G. 2003 The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**, 343–367. (doi:10.1007/s10236-003-0036-9)

Forest, C. E., Allen, M. R., Stone, P. H. & Sokolov, A. P. 2000 Constraining uncertainties in climate models using climate change detection techniques. *Geophys. Res. Lett.* **27**, 569–572. (doi:10.1029/1999GL010859)

Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R. & Webster, M. D. 2002 Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science* **295**, 113–117. (doi:10.1126/science.1064419)

Forster, P. M., Blackburn, M., Glover, R. & Shine, K. P. 2000 An examination of climate sensitivity for idealised climate change experiments in an intermediate general circulation model. *Clim. Dynam.* **16**, 833–849. (doi:10.1007/s003820000083)

Garey, M. R. & Johnson, D. S. 1979 *Computers and intractability: a guide to the theory of NP-completeness.* San Francisco, CA: W. H. Freeman.

Gregory, J. M., Stouffer, R. J., Raper, S. C. B., Stott, P. A. & Rayner, N. A. 2002 An observationally based estimate of the climate sensitivity. *J. Clim.* **15**, 3117–3121. (doi:10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2)

Hall, M. C. G., Cacuci, D. G. & Schlesinger, M. E. 1982 Sensitivity analysis of a radiative–convective model by the adjoint method. *J. Atmos. Sci.* **39**, 2038–2050. (doi:10.1175/1520-0469(1982)039<2038:SAOARC>2.0.CO;2)

Hammersley, J. M. & Handscomb, D. C. 1964 *Monte Carlo methods.* London, UK: Methuen & Co Ltd.

Hargreaves, J. C. & Annan, J. D. 2002 Assimilation of paleo-data in a simple Earth system model. *Clim. Dynam.* **19**, 371–381. (doi:10.1007/s00382-002-0241-0)

Hargreaves, J. C. & Annan, J. D. 2005 Using ensemble prediction methods to examine regional climate variation under global warming scenarios. *Ocean Model.* **11**, 174–192. (doi:10.1016/j.ocemod.2004.12.004)

Hargreaves, J. C., Annan, J. D., Edwards, N. R. & Marsh, R. 2004 Climate forecasting using an intermediate complexity Earth system model and the ensemble Kalman filter. *Clim. Dynam.* **23**, 745–760. (doi:10.1007/s00382-004-0471-4)

Jackson, C., Sen, M. K. & Stoffa, P. L. 2004 An efficient stochastic Bayesian approach to optimal parameter and uncertainty estimation for climate model predictions. *J. Clim.* **17**, 2828–2841. (doi:10.1175/1520-0442(2004)017<2828:AESBAT>2.0.CO;2)

Kalman, R. E. 1960 A new approach to linear filtering and prediction problems. *J. Basic Eng. D* **82**, 33–45.

Knutti, R., Stocker, T. F., Joos, F. & Plattner, G.-K. 2002 Constraints on radiative forcing and future climate change from observations and climate model ensembles. *Nature* **416**, 719–723. (doi:10.1038/416719a)

Lea, D. J., Allen, M. R. & Haine, T. W. N. 2000 Sensitivity analysis of the climate of a chaotic system. *Tellus A* **52**, 523–532. (doi:10.1034/j.1600-0870.2000.01137.x)

Lea, D. J., Haine, T. W. N., Allen, M. R. & Hansen, J. A. 2002 Sensitivity analysis of the climate of a chaotic ocean circulation model. *Q. J. R. Meteorol. Soc.* **128**, 2587–2605. (doi:10.1256/qj.01.180)

Liu, J. & West, M. 2001 Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo methods in practice* (eds A. Doucet, J. F. G. De Freitas & N. J. Gordon), pp. 197–223. Berlin, Germany: Springer.

McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245. (doi:10.2307/1268522)

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. 1953 Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1091. (doi:10.1063/1.1699114)

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M. & Stainforth, D. A. 2004 Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772. (doi:10.1038/nature02771)

Price, A. R., Voutchkov, I. I., Pound, G. E., Edwards, N. R., Lenton, T. M., Cox, S. J. & The GENIE Team. 2006 Multiobjective tuning of grid-enabled earth system models using a non-dominated sorting genetic algorithm (nsga-ii). In *2nd Int. Conf. on e-Science and Grid Computing.* Amsterdam, The Netherlands.

Rougier, J. C. 2007 Probabilistic inference for future climate using an ensemble of simulator evaluations. *Clim. Change* **81**, 247–264. (doi:10.1007/s10584-006-9156-9)

Rougier, J. & Sexton, D. M. H. 2007 Inference in ensemble experiments. *Phil. Trans. R. Soc. A* **365**, 2133–2143. (doi:10.1098/rsta.2007.2071)

Schneider von Deimling, T., Held, H., Ganopolski, A. & Rahmstorf, S. 2006 Climate sensitivity estimated from ensemble simulations of glacial climate. *Clim. Dynam.* **27**, 149–163. (doi:10.1007/s00382-006-0126-8)

Stainforth, D. A. *et al.* 2005 Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature* **433**, 403–406. (doi:10.1038/nature03301)

Tarasov, L. & Peltier, W. R. 2005 Arctic freshwater forcing of the Younger–Dryas cold reversal. *Nature* **435**, 662–665. (doi:10.1038/nature03617)

Terwisscha van Scheltinga, A. D. & Dijkstra, H. A. 2005 Nonlinear data-assimilation using implicit models. *Nonlin. Process. Geophys.* **12**, 515–525.

Zhou, Y., Mclaughlin, D. & Entekhabi, D. 2006 Assessing the performance of the ensemble Kalman filter for land surface DTA assimilation. *Mon. Weather Rev.* **134**, 2128–2142. (doi:10.1175/MWR3153.1)