Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2009)
58, *Part* 5, *pp.* 679–704

A smoothing algorithm for estimating stochastic, continuous time model parameters and its application to a simple climate model

Lorenzo Tomassini and Peter Reichert,

Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, and Eidgenössiche Technische Hochschule Zürich, Switzerland

Hans R. Künsch, Christoph Buser and Reto Knutti

Eidgenössiche Technische Hochschule Zurich, Switzerland

and Mark E. Borsuk

Dartmouth College, Hanover, USA

[Received October 2006. Final revision February 2009]

Summary. Even after careful calibration, the output of deterministic models of environmental systems usually still show systematic deviations from measured data. To analyse possible causes of these discrepancies, we make selected model parameters time variable by treating them as continuous time stochastic processes. This extends an approach that was proposed earlier using discrete time stochastic processes. We present a Markov chain Monte Carlo algorithm for Bayesian estimation of such parameters jointly with the other, constant, parameters of the model. The algorithm consists of Gibbs sampling between constant and time varying parameters by using a Metropolis–Hastings algorithm for each parameter type. For the time varying parameter, we split the overall time period into consecutive intervals of random length, over each of which we use a conditional Ornstein–Uhlenbeck process with fixed end points as the proposal distribution in a Metropolis–Hastings algorithm. The hyperparameters of the stochastic process are selected by using a cross-validation criterion which maximizes a pseudolikelihood value, for which we have derived a computationally efficient estimator. We tested our algorithm by using a simple climate model. The results show that the algorithm behaves well, is computationally tractable and improves the fit of the model to the data when applied to an additional time-dependent forcing component. However, this additional forcing term is too large to be a reasonable correction of estimated forcing and it alters the posterior distribution of the other, time constant parameters to unrealistic values. This difficulty, and the impossibility of achieving a good simulation when making other parameters time dependent, indicates a more fundamental, structural deficit of the climate model. This is probably related to the poor resolution of the ocean in the model. Our study demonstrates the technical feasibility of the smoothing technique but also the need for a careful interpretation of the results.

Keywords: Climate modelling; Cross-validation; Smoothing; State space models; Stochastic process

1. Introduction

Model structures are rarely detailed or sufficiently accurate to reproduce the behaviour of an environmental system at a level of precision that is comparable with measurements. This means

Address for correspondence: Lorenzo Tomassini, Max Planck Institute for Meteorology, Bundesstrasse 53, D-20146 Hamburg, Germany.
E-mail: lorenzo.tomassini@zmaw.de

 0035–9254/09/58679

that the assumption of independent and identically distributed measurement errors as the only source of discrepancy between deterministic model results and measured data is often invalid. Systematic deviations of model output from data might result from (measurement) errors in model input, deficiencies in model formulation and parameter values or disregard of non-deterministic processes. Such apparently non-deterministic processes can result from true indeterminism, or they might be the consequence of model aggregation and simplification. In any case, the implausibility of the assumption of independent and identically distributed errors implies that parameter uncertainty estimates or model extrapolations based on simple non-linear regression techniques are often unreliable.

This problem has motivated approaches that introduce a model inadequacy, or bias, term to model output, in addition to the usual measurement error term (Craig *et al.*, 1996, 2001; Kennedy and O'Hagan, 2001; Bayarri *et al.*, 2007). In a Bayesian context, such a term is intended to describe our knowledge of the effect of model deficiencies on model output. It is usually formulated in a non-parametric form as a Gaussian process interpolating the (deterministic) model deficit between observation points (Currin *et al.*, 1992; O'Hagan, 1992; Oakley and O'Hagan, 2002). This approach has the virtue of being universally applicable without any detailed knowledge or assumptions about physical causes of the model deficiencies. It can also lead to more reliable estimates of uncertainty of the model parameters. However, as this leads only to a statistical description of bias during the calibration period, the lack of information on underlying mechanisms makes it difficult to apply this concept for prediction.

In the present paper we follow a different, but related, approach. We identify which parameters, when made time variable, would significantly improve the fit and what degree of variation of the parameter is required to do so. This can lead to insight that can be directly used to improve the model by either keeping this parameter stochastic to obtain better estimates of uncertainty (if time variation of the parameter is physically justified and the magnitude of the error seems reasonable), improving the deterministic part of the model through relationships with external or internal influence factors gained from the identified parameter time series or, at least, uncovering more fundamental model deficits. Technically this approach involves replacing a chosen parameter by a dynamic stochastic process, the state of which is estimated jointly with the other (time constant) parameters of the model.

This approach was proposed more than 30 years ago and was successfully applied using discrete time stochastic processes to describe the time-dependent parameter (Beck and Young, 1976; Beck, 1983; Young, 2001; Young *et al.*, 2001; Kristensen *et al.*, 2004). In this paper, we

(a) extend the approach from discrete time to continuous time parameters and
(b) develop a numerical scheme for Bayesian inference that is applicable to non-linear models without relying on linearization.

Because we concentrate on the application of the proposed technique to environmental systems, we apply the stochastic approach to time-dependent model parameters and inputs only, rather than to modelled state variables. This is because, in contrast with other approaches (Vrugt *et al.*, 2005), we want to avoid violation of conservation equations (Kuczera *et al.*, 2006). We feel that any apparent violations of conservation of mass, heat or momentum are due to imprecise inputs, neglected or inappropriately formulated processes or measurement uncertainty, and not due to true violations of fundamental laws. In our view, violations of fundamental laws should therefore be avoided if possible.

We demonstrate our approach by application to a widely used simple global climate model that was proposed by Wigley and Raper (1987, 1992). For the assessment reports of the Intergovernmental Panel on Climate Change (IPCC) (1995, 2001) this model was used to represent more

complex atmosphere–ocean general circulation models for probabilistic climate projections and the exploration of future scenarios (see also Wigley and Raper (2001) and Meinshausen (2005)). Moreover, it served to diagnose and analyse atmosphere–ocean general circulation model simulations (Raper *et al.*, 2001) to investigate geochemical cycles (Osborne and Wigley, 1994), and it is a part of various integrated assessment models (Edmonds *et al.*, 1994; IMAGE-team, 2001).

The driver of the climate model, radiative forcing, must be reconstructed from the past. This reconstruction usually proceeds in two steps. First a proxy (e.g. the number of sunspots) for a specific component of radiative forcing (e.g. solar radiation) must be reconstructed. Second, a link between the proxy and the effective contribution to radiative forcing must be established. Both steps are affected by uncertainty. This uncertainty cannot simply be represented by an additive or multiplicative constant additional to the historic reconstructions (Crowley, 2000; Joos *et al.*, 2001) but is expected to have a time-dependent structure. We therefore apply our technique to the problem of estimating a time-dependent error term additional to the reconstructed historic radiative forcing. By considering this uncertainty in model input, it is hoped that the estimates of uncertainty of the other, constant, parameters of the climate model will also be improved. Moreover, the estimated time series of forcing corrections can support the identification of other model deficits, e.g. by showing that the forcing error cannot be the cause of certain model–data discrepancies.

This paper is structured as follows. Section 2 describes the governing equations of the general model and the time-dependent parameter. Section 3 describes the techniques to estimate constant model parameters and the time-dependent parameter jointly. In Section 4 a cross-validation criterion to constrain the hyperparameters of the time-dependent process is proposed and an estimator described. Section 5 contains the application to the simple climate model. Finally, in Section 6 conclusions are drawn with respect to our methodology and to the specific application to the climate model.

2. Governing equations

2.1. Differential equation models

We consider an environmental system that is described by a state vector x whose time evolution we want to understand. We assume that there is a physically based model for our system in the form of a set of differential equations (usually derived from conservation laws)

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f\{x(t), \psi\}, \qquad x(0) = x_0, \quad t \in [0, T] =: I. \tag{1}$$

Here the vector ψ denotes unknown parameters and/or inputs to the system.

We assume that our data $y = (y_1, \ldots, y_p)$ consist of noisy observations of some known function η of the state vector at a finite number of discrete time points $t^{\mathrm{ob}} = (t_1^{\mathrm{ob}}, \ldots, t_p^{\mathrm{ob}})$. We call $\eta\{x(\cdot)\}$ the model output. Because of deficiencies in the model, we expect not only random errors but also a bias. As mentioned in Section 1, we try to correct for this bias by allowing a component of ψ to be time varying: $\psi = (\phi, \theta)$ with $\phi = \phi(\cdot) : I \to \mathbb{R}$. We then identify the size and shape of $\phi(\cdot)$ that is required to remove the bias, if possible, and from this we try to learn about the nature of the bias. For ease of presentation, we assume that $\phi(t)$ is a scalar; the multi-dimensional case is analogous, although correlation in the parameters might have to be considered and identifiability problems can occur, depending on the experimental design. The time constant parameter θ is assumed to be multi-dimensional. This means that instead of model (1) with fixed parameters we consider the solution of

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = f\{x(t), \phi(t), \theta\}, \qquad x(0) = x_0, \quad t \in I. \tag{2}$$

In particular, $x(\cdot)$ depends deterministically on t, $\phi(\cdot)$ and θ. The value of $x(t)$ for a fixed t is influenced not only by $\phi(t)$, but also by the whole trajectory of $\phi(\cdot)$ on $[0, t]$ (and by the initial values of x). In what follows x will be considered either as a function of t or as a function of t and all model parameters $\phi(\cdot)$ and θ. It should be clear from the context which viewpoint is adopted.

We assume that the bias of the model is absorbed in the time-dependent parameter, but we still need to consider two kinds of randomness: observation errors, which are denoted by ε_i, and random effects that are not represented in the system, which are denoted by v_i. Hence the measurements are given as

$$y_i = \eta[\{t_i^{\mathrm{ob}}, \phi(\cdot), \theta\}] + v_i + \varepsilon_i. \tag{3}$$

Assuming zero means and a joint Gaussian distribution for all random effects, we obtain

$$y|x(\cdot) \sim \mathcal{N}[\eta\{x(t^{\mathrm{ob}})\}, R]. \tag{4}$$

In our treatment, the covariance matrix R will be estimated independently of the data y and held fixed subsequently (see Section 5.2).

Our approach is Bayesian: we consider $\phi(\cdot)$ as a random process and θ as a random vector with a prior distribution. Our goal is to estimate the posterior probability densities of their distributions, given the observations, y.

2.2. The prior for the time varying parameter

In our development, we shall take as the prior for the time-dependent parameter $\phi(\cdot)$ a Gaussian process with constant mean value $\bar{\phi}$ and covariance function

$$\mathrm{cov}\{\phi(t), \phi(s)\} = \sigma^2 \exp(-\gamma|t-s|). \tag{5}$$

Hence σ^2 is the variance and $\tau := 1/\gamma$ is the characteristic correlation time of the process.

This process is a continuous time auto-regressive process: for any $h > 0$, we can write

$$\phi(t+h) = \bar{\phi} + \exp(-\gamma h)\{\phi(t) - \bar{\phi}\} + \varepsilon_{t,h}, \tag{6}$$

where $\varepsilon_{t,h}$ has variance $\sigma^2\{1 - \exp(-2\gamma h)\}$ and is uncorrelated with past values $\phi(s)$, $s \leqslant t$. Because of Gaussianity $\varepsilon_{t,h}$ is also independent of the past, and hence $\phi(\cdot)$ is Markovian. Moreover it can be shown to have continuous trajectories (see section 9.2 in Cramér and Leadbetter (1967)).

In our Markov chain Monte Carlo algorithm to be discussed later, we must simulate $\phi(\cdot)$ on an interval (s, u) given the values on the intervals $[0, s]$ and $[u, T]$. This is greatly simplified because the conditional distribution of $\phi(\cdot)$ on (s, u) given the values on $[0, s]$ and $[u, T]$ depends only on the two values $\phi(s)$ and $\phi(u)$. In discrete time, this follows from general results on graphical models, and it continues to hold in the limit if we approximate a continuous time process by a discrete time process. Moreover, because the process is Gaussian, this conditional distribution is again Gaussian, and we can compute the mean

$$\begin{aligned} E[\phi(t)|\phi(s), \phi(u)] = \bar{\phi} &+ \frac{\exp\{-\gamma(t-s)\}[1 - \exp\{-2\gamma(u-t)\}]}{1 - \exp\{-2\gamma(u-s)\}}\{\phi(s) - \bar{\phi}\} \\ &+ \frac{\exp\{-\gamma(u-t)\}[1 - \exp\{-2\gamma(t-s)\}]}{1 - \exp\{-2\gamma(u-s)\}}\{\phi(u) - \bar{\phi}\} \end{aligned} \tag{7}$$

and the variance

$$\mathrm{var}\{\phi(t)|\phi(s),\phi(u)\} = \sigma^2 \frac{[1-\exp\{-2\gamma(t-s)\}][1-\exp\{-2\gamma(u-t)\}]}{1-\exp\{-2\gamma(u-s)\}}. \tag{8}$$

(For $t-s=u-t$, for instance, the result is given as formula (1.6) in Elerian *et al.* (2001).) In particular, for given start and end values, conditional simulation of $\phi(\cdot)$ on a fine grid is straightforward: we can proceed sequentially in time, always using the last value that we generated and the end value.

In the context of modelling systems by differential equations, it is important to note that ϕ can also be written as the solution of the following linear Itô stochastic differential equation:

$$\mathrm{d}\phi(t) = -\gamma\{\phi(t)-\bar{\phi}\}\,\mathrm{d}t + \sqrt{(2\gamma)}\sigma\,\mathrm{d}w(t) \tag{9}$$

(see for example Kloeden and Platen (1995), or take the limit $h \to 0$ in equation (6)). The first term on the right-hand side of equation (9) represents a drift of $\phi(t)$ towards the mean $\bar{\phi}$; the second term describes the random fluctuations that are induced by the increments of a Brownian motion $w(t)$, i.e. white noise. In the literature, this process goes under the name of a mean reverting Ornstein–Uhlenbeck process.

Equations (2) and (9) imply that $(x(\cdot),\phi(\cdot))$ together are the solution of a system of stochastic differential equations. Since, in many subject areas, modelling by differential equations is the basic paradigm, we believe that it is an advantage to formulate the model extension in the same paradigm. If we want to use a more general prior for $\phi(\cdot)$, it is then natural to consider a more general class of stochastic differential equations instead of Gaussian processes with other covariance functions. For simulation with given start and end values, we can then use the algorithm of Beskos *et al.* (2006).

3. Estimation techniques

3.1. Markov chain Monte Carlo simulation

Let $\xi=(\sigma,\tau)$ denote the two-dimensional vector of hyperparameters of the covariance function (5) for $\phi(\cdot)$. The dependence structure of our model can then be summarized by the following directed acyclic graph where doubled edges indicate a deterministic relationship:

$$\begin{array}{ccccccccc} \xi & \longrightarrow & \phi(\cdot) & \Longrightarrow & x\{\cdot,\phi(\cdot),\theta\} & \Longrightarrow & \eta(x) & \longrightarrow & y \\ & & & & \Uparrow & & & & \\ & & & & \theta & & & & \end{array}$$

We are interested in the conditional distribution of $\phi(\cdot)$, θ and ξ given all available observations y. In particular, for estimating a value $\phi(t)$ for a given t, we want to take also observations at times $t_i^{\mathrm{ob}} > t$ into account. In the state space model framework, this is called a 'smoothing' problem. Since this conditional distribution is not available in closed form, we use a Markov chain Monte Carlo algorithm with some special adaptations to the present situation (Buser, 2003; Elerian *et al.*, 2001).

The process $\phi(\cdot)$ is in principle an infinite dimensional object. Although the technical problems that are associated with this can be solved, we are satisfied here with the distribution of $\phi(\cdot)$ on a fine grid. Then we can assume that all random variables have densities (with respect to the Lebesgue measure). This grid will usually be chosen to be much finer than the grid of observation times. It may be convenient to choose a regular grid and to interpolate between grid points if the numerical variable-step algorithm that is used to solve the differential equation (2) requires intermediate values.

To draw a sample from $p\{\phi(\cdot),\theta,\xi|y\}$ with the Gibbs sampler, we would generate at the $(k+1)$th iteration the values

(a) θ^{k+1} according to $p\{\theta|\phi(\cdot)^k, \xi^k, y\}\,\mathrm{d}\theta = p\{\theta|\phi(\cdot)^k, y\}\,\mathrm{d}\theta \propto p(\theta)\, p\{y|\phi(\cdot)^k, \theta\}\,\mathrm{d}\theta$,
(b) ξ^{k+1} according to $p\{\xi|\phi(\cdot)^k, \theta^{k+1}, y\}\,\mathrm{d}\xi = p\{\xi|\phi(\cdot)^k\}\,\mathrm{d}\xi \propto p(\xi)\, p\{\phi(\cdot)^k|\xi\}\,\mathrm{d}\xi$ and
(c) $\phi(\cdot)^{k+1}$ according to $p\{\phi(\cdot)|\theta^{k+1}, \xi^{k+1}, y\}\,\mathrm{d}\phi(\cdot) \propto p\{\phi(\cdot)|\xi^{k+1}\}\, p\{y|\phi(\cdot), \theta^{k+1}\}\,\mathrm{d}\phi(\cdot)$.

Including the hyperparameters ξ of the Ornstein–Uhlenbeck process in the estimation procedure as described above leads, however, to problems of convergence if the process $\phi(\cdot)$ is considered on a fine grid. This phenomenon, which occurs for all diffusion models with unknown innovation variance, has been analysed in detail by Roberts and Stramer (2001). The reason is that, if we observe $\phi(\cdot)$ on a grid with step size T/m, then the innovation variance σ^2/τ can be estimated with an error of order $O(m^{-1/2})$ because the innovations are independent and identically distributed by the Markov property. In other words, the supports of the process $\phi(\cdot)$ on a fine grid are almost disjoint for two different values of the innovation variance σ^2/τ. Hence, if we generate ϕ on a fine grid in step (c) above, then the change in the ratio σ^2/τ in the next step (b) is small. To overcome this problem, we use a plug-in value for the hyperparameters τ and σ based on the cross-validation criterion (Gelfand and Dey, 1994). This is discussed in detail in Section 4.

In addition, depending on the prior, there is a danger that τ converges to 0 and σ to ∞, leaving complete freedom for $\phi(\cdot)$ and thus allowing $\eta\{x(t_i)\}$ to match the observations without error. Since we want to avoid such behaviour, we would presumably need a strong prior for ξ. Determining the hyperparameters σ and τ by cross-validation instead of including them in the Bayesian inference process avoids this problem (see Section 4 for more details).

In what follows, ξ is therefore considered to be fixed, and we iterate between steps (a) and (c) above. In both steps, we use the Metropolis–Hastings algorithm. The update in step (a) is done in a standard way with random-walk proposals and so its description is omitted here. The updates in step (c) are more difficult. Our algorithm uses ideas of Buser (2003) and Elerian *et al.* (2001), and we give more details next.

3.2. Updating $\phi(\cdot)$

Updating $\phi(\cdot)$ on the whole interval I in a single step with the Metropolis–Hastings algorithm is not practical. It seems impossible to find a proposal distribution that gives reasonable acceptance probabilities. We therefore partition I into smaller subintervals I_j, $j = 1, \ldots, N$, such that the end point of I_j is the starting point of I_{j+1} and we propose updates on one subinterval I_j while keeping the value of $\phi(\cdot)$ outside I_j fixed. The subintervals are visited in sequential order during one iteration, and the lengths of the subintervals are chosen randomly with equal average lengths. In this way, the boundaries of the intervals change from one iteration to the next.

When updating $\phi(\cdot)$ on I_j we use the prior distribution for our proposal. This means that we simulate a trajectory $\tilde{\phi}(\cdot)$ on I_j of the Ornstein–Uhlenbeck process conditional on $\phi^{k+1}(t_j^{\mathrm{start}})$ and $\phi^k(t_j^{\mathrm{end}})$ (see Section 2.2). We calculate the acceptance probability r (see below for more details) and accept or reject according to the Metropolis–Hastings recipe.

In other words, the stages of the procedure are as follows: for $j = 1, \ldots, N$, do the following processes.

(a) Draw $\tilde{\phi}(\cdot)$ on I_j according to a conditional Ornstein–Uhlenbeck process with $\tilde{\phi}(t_j^{\mathrm{start}}) = \phi^{k+1}(t_j^{\mathrm{start}})$ and $\tilde{\phi}(t_j^{\mathrm{end}}) = \phi^k(t_j^{\mathrm{end}})$. For $j = 1$, we only condition on the end value, and, if $j = N$, we only condition on the start value.
(b) Compute the acceptance probability r according to formula (10) given below.
(c) On I_j, set $\phi^{k+1}(\cdot) = \tilde{\phi}(\cdot)$ with probability r and $\phi^{k+1}(\cdot) = \phi^k(\cdot)$ with probability $1 - r$.

This constitutes the third step in the recursion that was described in Section 3.1.

It remains to discuss the computation of the acceptance probability r. For a function ϕ on I, denote its restriction to the intervals I_s with $s<j$, $s=j$ and $s>j$ by ϕ_1, ϕ_2 and ϕ_3 respectively. We thus propose a move from $\phi_2^k(\cdot)$ to $\tilde{\phi}_2(\cdot)$ while keeping $\phi_1(\cdot)=\phi_1^{k+1}(\cdot)$ and $\phi_3(\cdot)=\phi_3^k(\cdot)$ fixed. Then the ratio of the target densities is equal to the likelihood ratio

$$\frac{p\{y|\phi_1^{k+1}(\cdot),\tilde{\phi}_2(\cdot),\phi_3^k(\cdot),\theta^{k+1}\}}{p\{y|\phi_1^{k+1}(\cdot),\phi_2^k(\cdot),\phi_3^k(\cdot),\theta^{k+1}\}}$$

multiplied by the ratio of the conditional priors

$$\frac{p\{\tilde{\phi}_2(\cdot)|\phi_1^{k+1}(\cdot),\phi_3^k(\cdot)\}}{p\{\phi_2^k(\cdot)|\phi_1^{k+1}(\cdot),\phi_3^k(\cdot)\}}.$$

Now, by construction, the second ratio is the inverse of the ratio of the proposal densities, and thus it cancels when calculating the acceptance ratio. Hence we obtain

$$r=\min\left[1,\frac{p\{y|\phi_1^{k+1}(\cdot),\tilde{\phi}_2(\cdot),\phi_3^k(\cdot),\theta^{k+1}\}}{p\{y|\phi_1^{k+1}(\cdot),\phi_2^k(\cdot),\phi_3^k(\cdot),\theta^{k+1}\}}\right]. \tag{10}$$

Since the likelihood $p\{y|\phi(\cdot),\theta\}$ has the form $p(y|\eta[x\{\phi(\cdot),\theta\}])$, it means that for each subinterval I_j we must solve the differential equation (2) on the interval $[t_j^{\text{start}},T]$ which is time consuming in general.

4. Cross-validation

For a given choice of the hyperparameters $\xi=(\sigma,\tau)$ of the Ornstein–Uhlenbeck process, we assume that we can simulate from the posterior, i.e. the conditional distribution of $(\phi(\cdot),\theta)$ given y and ξ. The problem is how to choose the hyperparameters ξ without including them in the updates of the Markov chain Monte Carlo algorithm. We use a cross-validation criterion (Gelfand and Dey, 1994) which is similar to an empirical Bayes approach. Instead of the marginal likelihood we maximize the so-called pseudolikelihood

$$\text{psl}(\xi):=\sum_{i=1}^{p}\log\{p(y_i|y_{-i},\xi)\} \tag{11}$$

with respect to the hyperparameters ξ. Here, y_{-i} is a notation for the set of all observations except y_i, i.e. $y_{-i}=\{y_1,\ldots,y_{i-1},y_{i+1},\ldots,y_p\}$.

Calculating the predictive densities $p(y_i|y_{-i},\xi)$ in expression (11) in a naive way is computationally very expensive. The following lemma shows that we can estimate $p(y_i|y_{-i},\xi)$ for all i from a sample of the full posterior density $p\{\phi(\cdot),\theta|y,\xi\}$. To make the procedure feasible, we maximize psl(ξ) only over a relatively rough grid of values of ξ.

To simplify the notation in the lemma, we ignore that $\phi(\cdot)$ is a function and we use the notation $\psi=(\phi,\theta)$.

Lemma 1.

$$p(y_i|y_{-i})=\frac{1}{\displaystyle\int p(y_i|y_{-i},\psi)^{-1}p(\psi|y)\,\mathrm{d}\psi}. \tag{12}$$

Proof. By the definition of the conditional densities

$$p(y_i|y_{-i})=p(y)/p(y_{-i}),$$
$$p(y_i|y_{-i},\psi)=p(y|\psi)/p(y_{-i}|\psi).$$

By the law of total probability

$$p(y_{-i}) = \int p(y_{-i}|\psi)\, p(\psi)\, \mathrm{d}\psi = \int \frac{p(y|\psi)}{p(y_i|y_{-i},\psi)}\, p(\psi)\, \mathrm{d}\psi.$$

Using $p(y|\psi)\, p(\psi) = p(\psi|y)\, p(y)$, we obtain

$$p(y_{-i}) = p(y) \int \frac{1}{p(y_i|y_{-i},\psi)}\, p(\psi|y)\, \mathrm{d}\psi.$$

From this, lemma 1 follows.

High likelihood values can be achieved by the model matching the observations very closely. Besides adjustments of the time constant parameters, this can be achieved by adjusting the time course of the time-dependent parameter ϕ. For this reason, higher likelihood values can be achieved if the hyperparameters allow ϕ to vary quickly. This is why inclusion of the hyperparameters in the Bayesian inference process leads to a tendency for τ to converge to 0 and σ to ∞ unless this is prevented by a strong prior. Because

$$p(y_i|y_{-i}) = \int p(y_i|y_{-i},\psi)\, p(\psi|y_{-i})\, \mathrm{d}\psi,$$

we expect to avoid this problem by determining these hyperparameters by cross-validation. For any given i, hyperparameters that lead to a high flexibility of ϕ will allow for realizations of ϕ with a large value of $p(y_i|y_{-i},\psi)$ (note that ϕ is part of ψ). However, as this expression is integrated with respect to the density $p(\psi|y_{-i})$ which will also have increasing mass for realizations of ϕ that lead to a poor fit of y_i, continued increase of the flexibility of ϕ will finally reduce the pseudolikelihood.

Lemma 1 allows us now to derive an estimator of the pseudolikelihood value from a Markov chain sample. If we have a sample $(\psi^k; k = 1, \ldots, M)$ from the full posterior density $p(\psi|y)$, we can approximate $p(y_i|y_{-i})$ by the harmonic mean of the $p(y_i|y_{-i},\psi^k)$s. In our situation, $p(y|\psi)$ must be replaced by $p[y|x\{\phi(\cdot),\theta\}]$. From equation (4) this is a Gaussian density with mean vector $\eta\{x(t^{\mathrm{ob}})\}$ and covariance matrix R. By standard results concerning normal distributions, the conditional density $p(y_i|y_{-i})$ is then again Gaussian with variance $1/\Lambda_{ii}$ and mean

$$\eta\{x(t_i^{\mathrm{ob}})\} - \sum_{j\neq i} \frac{\Lambda_{ij}}{\Lambda_{ii}}[y_j - \eta\{x(t_j^{\mathrm{ob}})\}]$$

where Λ is the inverse of the covariance matrix: $\Lambda = R^{-1}$. Hence, in our case, the Monte Carlo estimate $\hat{p}(y_i|y_{-i})$ of $p(y_i|y_{-i})$ is given by

$$\frac{1}{\hat{p}(y_i|y_{-i})} = \frac{(2\pi)^{1/2}}{M} \sum_{k=1}^{M} (\Lambda_{ii})^{-1/2} \exp\left\{ \frac{\Lambda_{ii}}{2} \left(y_i - \eta\{x^k(t_i^{\mathrm{ob}})\} + \sum_{j\neq i} \frac{\Lambda_{ij}}{\Lambda_{ii}}[y_j - \eta\{x^k(t_j^{\mathrm{ob}})\}] \right)^2 \right\}. \tag{13}$$

Computing the right-hand side is straightforward since we need to invert R only once and we shall have saved the values $x^k(t_j^{\mathrm{ob}})$. Substituting this into equation (11), we obtain our estimator of the pseudolikelihood:

$$\widehat{\mathrm{psl}} := \sum_{i=1}^{p} \log\{\hat{p}(y_i|y_{-i})\}. \tag{14}$$

5. Application to a simple climate model

We apply the techniques described to a hemispherically averaged, upwelling diffusion, energy balance model of global climate (Wigley and Raper, 1987, 1992). Similar climate models of

reduced complexity are an accepted and indispensable part of the model hierarchy in climate modelling (Intergovernmental Panel on Climate Change (2001), chapter 8). This is because more complex general circulation models are computationally too expensive for probabilistically estimating climate system properties and future projected global mean temperatures.

5.1. The climate model

The climate model consists of two land and two ocean boxes (representing the northern and southern hemispheres). The two ocean boxes are each vertically divided into 40 layers. The two top layers, which are assumed to represent the ocean mixed layers, absorb the energy of solar radiation (Fig. 1). It is assumed that no energy is absorbed above land.

At timescales that are relevant to climate change, the atmosphere may be assumed to be in equilibrium with the oceanic mixed layer. This leads to the following differential equation for ΔT_0, the difference in the temperature of the oceanic mixed layer from its equilibrium value for preindustrial times:

$$C_{\mathrm{ml}} \frac{\mathrm{d}\Delta T_0(t)}{\mathrm{d}t} = \Delta F(t) - Q(t) - \beta\, \Delta T_0(t) \tag{15}$$

where C_{ml} is effective heat capacity. The three terms on the right-hand side represent the causes for a change in temperature: ΔF is the variation in radiative forcing, i.e. the change (relative to the equilibrium state) in incoming power per unit area due to direct and indirect radiation, Q is the net heat flux into the deeper ocean and the third term is a sum of feedback mechanisms pulling the temperature back to equilibrium ($\Delta T_0 = 0$). The proportionality constant β determines the new equilibrium value of the temperature if the forcing is changed by a constant amount. This can be seen by considering the case in which the carbon dioxide (CO_2) concentrations in the atmosphere are doubled. Then the forcing will increase by an amount $\Delta F = F^{\mathrm{double}}$, the actual value being $F^{\mathrm{double}} = 3.71\,\mathrm{Wm}^{-2}$ (Myhre *et al.*, 1998). If there are no other changes in the forcing,

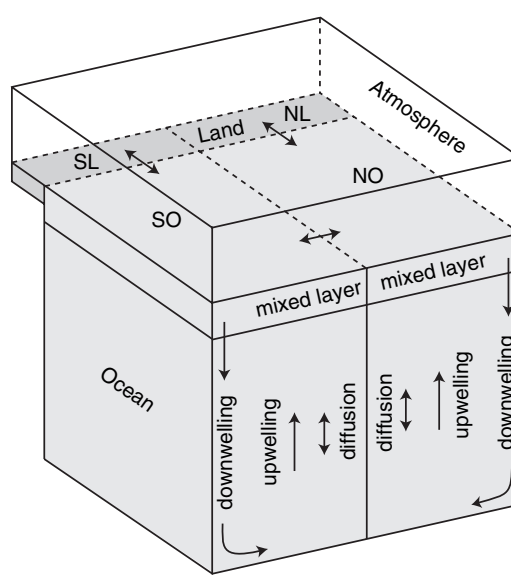

Fig. 1. Schematic figure of the simple climate model (Wigley and Raper, 1987, 1992) that is used to exemplify the smoothing algorithm

the atmosphere–ocean system will eventually reach a new equilibrium in which ΔT_0 is constant and $Q = 0$. Hence, from equation(15), we obtain that

$$0 = F^{\text{double}} - \beta\,\Delta T_0 \Leftrightarrow \Delta T_0 = F^{\text{double}}/\beta.$$

This equilibrium temperature increase due to a doubling of CO_2 concentrations is called *climate sensitivity* and is denoted by S. Hence we have $\beta = F^{\text{double}}/S$. Climate sensitivity is the most important uncertain quantity of the climate system (and climate models) when it comes to future climate projections on timescales of 100 years or more. It determines the long-term temperature response of the climate system to increased radiative forcing from rising levels of CO_2 in the atmosphere.

In our climate model, there is heat exchange between the ocean mixed layer and the land and between the two boxes of the mixed layer. This exchange of heat is determined by constant exchange rates. Moreover, heat is transported in the ocean by diffusion and by a simplified representation of the global, temperature-dependent ocean circulation (thermohaline circulation). The temperature of the ocean is therefore described by another differential equation:

$$\frac{\partial \Delta T(t,z)}{\partial t} = K\frac{\partial^2 \Delta T(t,z)}{\partial z^2} - w\frac{\partial \Delta T(t,z)}{\partial z}. \tag{16}$$

Here z denotes the vertical depth co-ordinate of the ocean, K is the constant rate of diffusion called *vertical ocean diffusivity* and w is the upwelling rate which is linear in ΔT_0. The diffusivity K describes how fast heat is transported in the ocean by diffusion. The temperature of the mixed layer ΔT_0 is the boundary condition at depth $z = 0$, and the net flux Q of equation (15) is proportional to the partial derivative $\partial \Delta T/\partial z$ at $z = 0$. Finally ocean depth is discretized into 40 layers, with partial derivatives replaced by finite differences.

The outputs of the climate model that we relate to observations are the global mean surface temperature $\Delta T^{\text{surface}}$ and the heat content of the ocean down to 700 m depth, $\Delta H^{700\text{m}}$. These are weighted averages of temperatures of the land and the oceanic mixed layers and of the first seven layers of the ocean respectively. For the complete set of model equations, we refer the reader to Wigley and Raper (1987, 1992) or Tomassini (2007).

In summary, the climate model that is used in the present analysis consists of a set of coupled differential equations that are integrated numerically by using a second-order explicit Euler scheme with variable step size. It produces two time series of yearly values for global mean surface temperature and heat content of the world ocean down to 700 m depth that will be compared with observations. 500 years of simulated time (which correspond to one model run in the present study) take about 4 s realtime on a personal computer.

5.2. The error covariance

The observations that we use consist of time series (yearly values) of global annual mean surface temperature data (Jones and Moberg, 2003) from the years 1861 to 2003 and annual mean change in world ocean heat content down to 700 m depth (Levitus *et al.*, 2005) from the years 1955 to 2003. Both data sets are publicly available (see Jones and Moberg (2003) and Levitus *et al.* (2005)). In addition to the values also a standard deviation is provided.

For the likelihood function, we need to specify the error covariance matrix R which reflects the observational errors ε_i as well as the errors v_i due to climate variability that are not included in the dynamics of the simple energy balance climate model. We assume that there is no dependence between these two types of errors so that R is the sum of the two respective covariance matrices.

The covariance matrix of the observational error is constructed on the basis of the standard deviations of the yearly data that are provided by Jones and Moberg (2003) and Levitus *et al.*

(2005) for the surface temperature data and the ocean heat content change data respectively. These standard deviations vary from year to year with a decreasing trend over the time intervals for which data are available. For the surface temperature data we assume independence of observational errors from year to year. Owing to methodological difficulties of measuring ocean heat content and changes in applied measurement techniques over time, it is unrealistic to assume independence of the observational errors of ocean heat content. This is supported by the following argument. On the basis of equation (3) and the independence between the two types of errors we obtain

$$\begin{aligned} E[(y_{i+1}-y_i)^2] &= E[(\eta_{i+1}-\eta_i+v_{i+1}-v_i)^2]+E[(\varepsilon_{i+1}-\varepsilon_i)^2] \\ &> E[(\varepsilon_{i+1}-\varepsilon_i)^2] = \mathrm{var}(\varepsilon_{i+1})-2\alpha_i\sqrt{\{\mathrm{var}(\varepsilon_{i+1})\,\mathrm{var}(\varepsilon_i)\}}+\mathrm{var}(\varepsilon_i) \end{aligned} \tag{17}$$

where α_i is the correlation coefficient between y_{i+1} and y_i. For the data set that was provided by Levitus *et al.* (2005), we obtain $\Sigma(y_{i+1}-y_i)^2 = 204.5(\times 10^{22}\mathrm{J})^2$ and $\Sigma\{\mathrm{var}(\varepsilon_{i+1})+\mathrm{var}(\varepsilon_i)\} = 437.4(\times 10^{22}\mathrm{J})^2$ which are clearly in disagreement with independent errors. Since it is difficult to obtain detailed information about the correlation structure, we assume an exponential decay

$$\mathrm{cov}(\varepsilon_i,\varepsilon_j) = \sqrt{\{\mathrm{var}(\varepsilon_i)\mathrm{var}(\varepsilon_j)\}}\alpha^{|i-j|}. \tag{18}$$

We used a value of α of 0.6. With this value, $\Sigma E[(\varepsilon_{i+1}-\varepsilon_i)^2]$ is equal to $176.9(\times 10^{22}\mathrm{J})^2$. It is thus a conservative estimate of α as significantly smaller values would not be compatible with the constraining equations whereas larger values are.

Short-term climate variability is the major source of randomness in the climate system that is not represented by the simple energy balance climate model. We cannot estimate natural variability from observed data alone because of the difficulty of separating natural variability from the underlying trend, and because the data time series is relatively short. As is common practice in climate change detection and attribution studies (compare for example Stott *et al.* (2001)), we therefore consider a control run of a complex climate model, in our case the Hadley Centre climate model HadCM3 (see Collins *et al.* (2001) for a detailed discussion of the internal variability of HadCM3), as a representation of climate variability. This control run contains processes such as short-term weather fluctuations, the north Atlantic oscillation and El Ñino southern oscillation related variability that are not included in the simple climate model that we are using.

The control run has a length of 900 years. It is based on constant radiative forcing and, being the result of a simulation program, no observation error. For these reasons, it can directly be used to analyse climate variability (terms v_i in equation (3)). On the basis of the Akaike information criterion, we identified univariate AR(3) models for both mean surface temperature and ocean heat content. An analysis of the cross-correlations between the innovations of these series indicated significant terms for lags -1, 0 and 1. We therefore identified a multivariate auto-regressive model (which resulted again to be of order 3) and calculated the correlation covariance matrix of this model.

Fig. 2 illustrates the auto- and cross-correlation structure of the final error model.

5.3. Radiative forcing and stochastic model term

A crucial input to our model is the radiative forcing ΔF in equation (15), which must be reconstructed from the past (see Crowley (2000) and Joos *et al.* (2001)). This is done by decomposing it into nine different components that are reconstructed individually: green-house gas forcing (the combined effect of CO_2, methane, nitrous oxide, sulphur hexafluoride and halocarbons), stratospheric ozone forcing, tropospheric ozone forcing, direct aerosol forcing, indirect aerosol

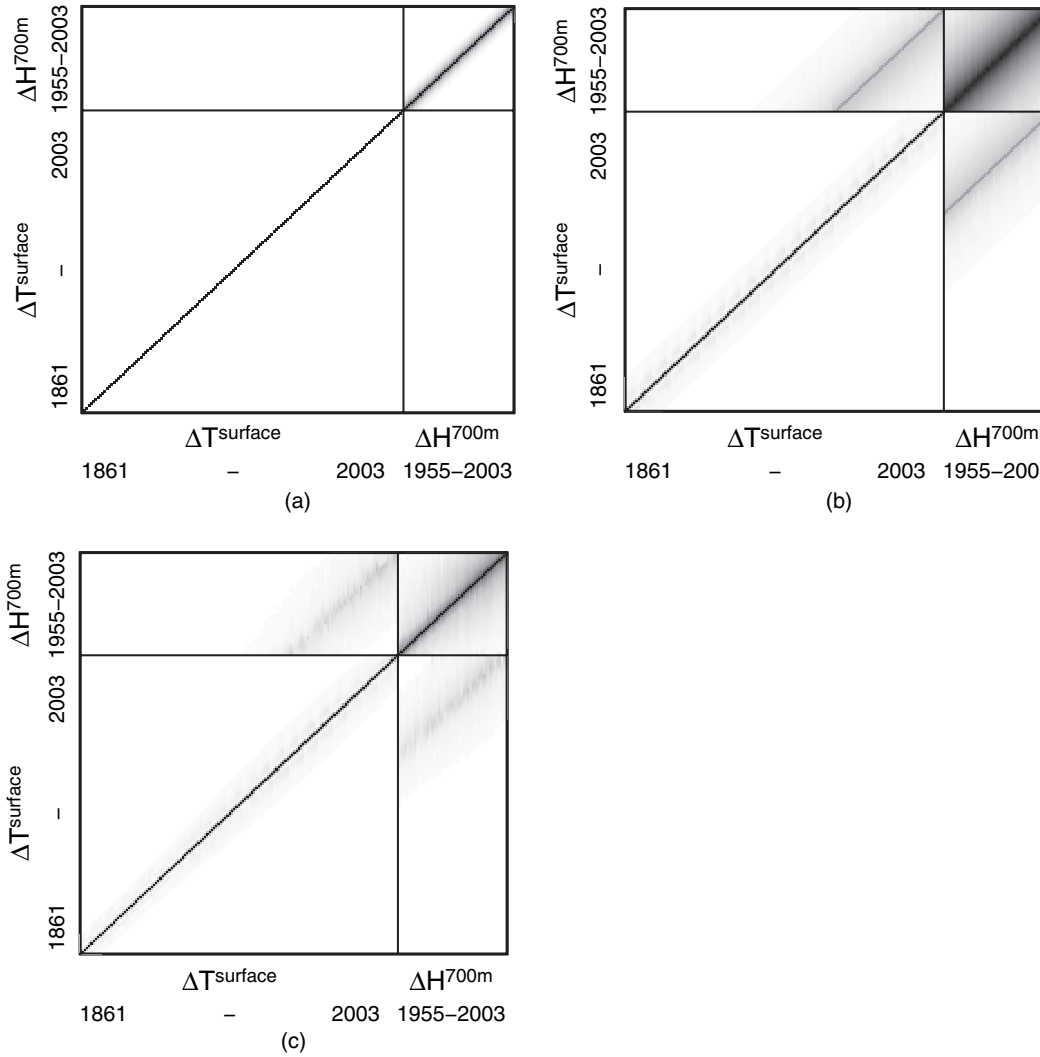

Fig. 2. Illustration of the auto- and cross-correlation structure of the normal distribution that was used to describe observation error and climate variability (grey levels indicate absolute values of correlation coefficients (0, white; 1, black); note that the available surface temperature time series (1861–2003) is longer than that of ocean heat content (1955–2003); see the text for details of the construction of the error model): (a) observation error only; (b) climate variability only; (c) correlation of total error

forcing, organic and black carbon forcing, stratospheric water forcing, volcanic forcing and solar forcing.

In the case of the volcanic forcing, the historic reconstructions actually relate to the optical depth of stratospheric volcanic aerosols (Ammann *et al.*, 2003), and not directly to the forcing that is caused by volcanic eruptions. The forcing is assumed to be a scalar multiple of the optical depth of stratospheric volcanic aerosols. This scaling factor, however, is uncertain. We therefore include a scaling parameter s_{volc} for the volcanic forcing in the Bayesian estimation. The parameter represents the relative deviation from the best estimate of Ammann *et al.* (2003). The standard deviation of the prior distribution is based on uncertainty estimates in Intergovernmental Panel on Climate Change (2001). Fig. 8(a) in Section 5.5.5 shows the total reconstructed

radiative forcing $\Delta F_{\text{recon}}(\cdot)$, i.e. the sum of all forcing components (with s_{volc} set to 1). The sharp negative peaks that can be seen in the forcing are due to volcanic eruptions.

Because there is considerable uncertainty in the reconstructed historic forcing (Intergovernmental Panel on Climate Change (2001), chapter 6) and, because this uncertainty can be expected to have a time-dependent structure, we added a stochastic and time varying component $\phi(\cdot)$:

$$\Delta F(t) = \Delta F_{\text{recon}}(t) + \phi(t). \tag{19}$$

Here, ΔF_{recon} is the forcing reconstructed from the past as described above, and the additional forcing $\phi(\cdot)$ is assumed to follow the Ornstein–Uhlenbeck process for time-dependent parameters that was described in Section 2.2 with $\bar{\phi} = 0$ and with the hyperparameters σ and τ.

5.4. Priors for constant parameters

To account for input uncertainty of volcanic forcing, we included a scaling factor of volcanic forcing as a parameter to be estimated from the data. In other studies (e.g. Knutti *et al.* (2002) and Forest *et al.* (2006)), similar factors were included for other forcing components as well. As these other forcing components vary at similar timescales to our time-dependent additional forcing component, we can omit these factors in the current study. In addition to the scaling factor of volcanic forcing, we included five constant parameters to be estimated together with the time varying component $\phi(\cdot)$. These parameters and their prior distributions are listed in Table 1 and visualized as broken curves in Fig. 3. The prior distributions of all parameters are assumed to be independent. For climate sensitivity S and vertical ocean diffusivity K (see Section 5.1) uniform priors over large ranges are used (see for example Knutti *et al.* (2002), Stainforth *et al.* (2005) and Forest *et al.* (2006) for reasonable values of climate sensitivity, and Raper *et al.* (2001) for a discussion of the vertical ocean diffusivity in the Wigley–Raper model). The depth of the ocean mixed layer h is a quantity that can actually be measured very well locally, but it varies over the globe between 40 m and 140 m. The global mean value that was used in our model is therefore considered to be uncertain with a mean of 90 m and a standard deviation of 10 m. The two parameters $\Delta T_{\text{ini}}^{\text{surface}}$ and $\Delta H_{\text{ini}}^{700\text{m}}$ are initial values for global mean surface temperature and annual mean change in ocean heat content down to 700 m depth respectively. These two parameters are of minor importance for our analysis since they are model dependent and should not be assigned a direct physical interpretation.

5.5. Results and discussion

In all our analyses, the Markov chain Mante Carlo sample size was 60 000. Every second point was disregarded to save computer storage space and to facilitate post-processing. In addition,

Table 1. Parameters to be estimated and their marginal prior distributions

Parameter	*Units*	*Prior distribution*
s_{volc}		log-normal(1.0,0.379)
S	K	uniform(1,10)
K	m^2a^{-1}	uniform(100,10000)
h	m	normal(90,10^2)
$\Delta T_{\text{ini}}^{\text{surface}}$	K	normal(-0.35,0.25^2)
$\Delta H_{\text{ini}}^{700\text{m}}$	10^{22} J	normal(-7.5,2.5^2)

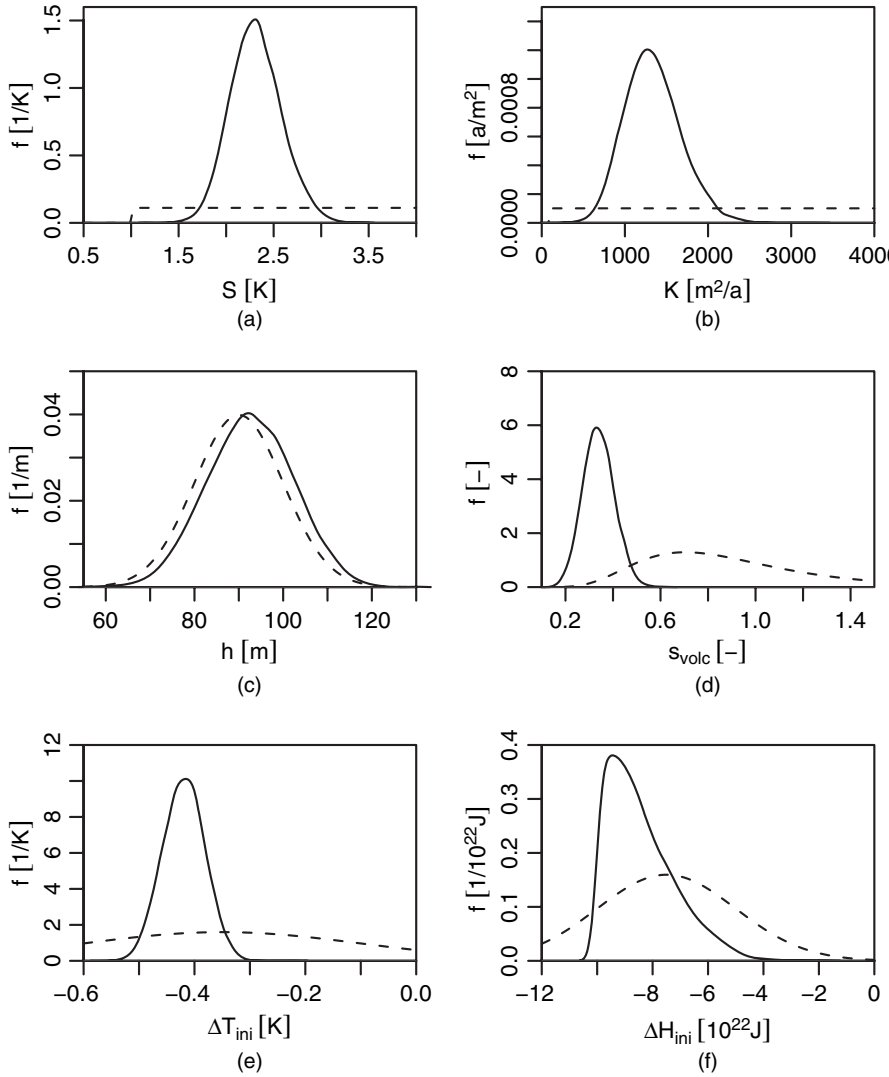

Fig. 3. Prior (– – –) and posterior (———) marginal distributions of constant parameters without inclusion of time-dependent parameters: (a) S; (b) K; (c) h; (d) s_{volc}; (e) ΔT_{ini}; (f) ΔH_{ini}

the first sixth of the chain was disregarded to overcome burn-in effects. An iterative process of improving the jump distribution of the Metropolis algorithm was run before finally calculating the chains of full length. The Heidelberger and Welch stationarity and half-width diagnostics (Heidelberger and Welch, 1983; Cowles and Carlin, 1996) as implemented in the R library `boa` were run with both parameters, and the accuracy of posterior estimates and the confidence level of the sample mean set to 0.1. All chains for which we show posterior marginals passed both of these tests. However, a small fraction of the chains for which we calculated psl-values (the small dots in Fig. 4) did not pass both tests.

5.5.1. Results without time-dependent parameters

First, we performed an analysis without time-dependent parameters. As there is no model inter-

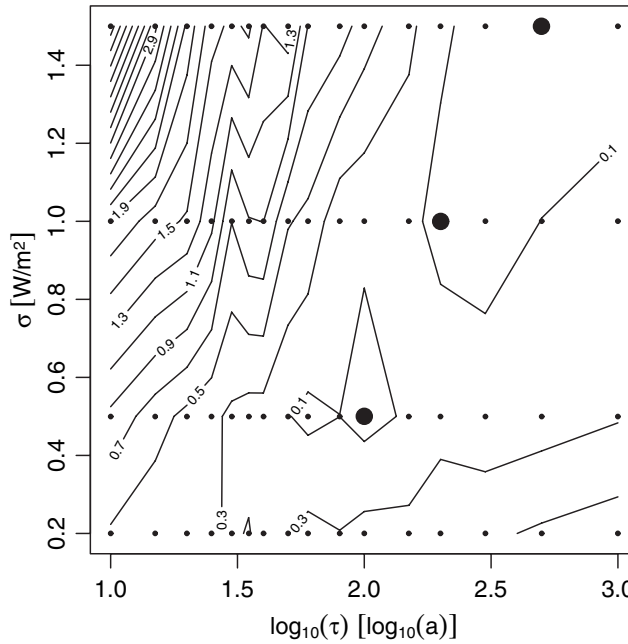

Fig. 4. Contour lines of pseudolikelihood values max(psl) − psl according to equation (14): the maximum is taken over (σ, τ) grid points at which estimated pseudolikelihood values were calculated (•): ●, positions of solutions that are discussed in more detail in the text

nal bias correction in this model, an additive bias correction to the output would be required. As this is not done in most applications, as a reference case we applied Bayesian inference to this model without explicitly considering model bias.

Fig. 3 shows the marginals of prior and posterior parameter distributions resulting for this reference case. According to this analysis, all parameters with the exception of the depth of the mixed ocean layer, h, are identifiable. The scaling factor for the volcanic forcing component, s_{volc}, obviously requires a reduction from its default value of 1. However, the results that are shown in Fig. 3 are not reliable, as a comparison of model results with observations shows a significant bias. (By combining a frequentist likelihood function with Bayesian priors we can do frequentist residual analyses at the maximum of the posterior.) This means that we must consider bias of this simple model either by considering an additive bias term or by trying to compensate for the bias by making one or more parameters time dependent. As input uncertainty is considerable, we first try to improve the model by adding a time-dependent input correction as described by the model equation (2) and further details given in Section 2. We shall then analyse whether this represents a reasonable input error or whether it indicates structural deficiencies of the model.

5.5.2. Choice of hyperparameters of stochastic forcing

Inclusion of the hyperparameters σ and τ of the Ornstein–Uhlenbeck process in the Bayesian estimation process as described in Section 3 turned out to be feasible from a technical point of view. However, unless the correlation time τ was bounded away from 0 with an informative prior, the posterior had the tendency to reduce τ to gain more freedom in adjusting the solution to the data. For this reason, we applied the cross-validation procedure that was described in Section 4 to constrain the hyperparameters of stochastic forcing.

Fig. 4 shows contour lines of the resulting pseudolikelihood surface calculated by using equation (14). Fig. 4 clearly demonstrates that the cross-validation procedure is successful in excluding small values of τ at which the predictability becomes worse. However, there are no constraints with respect to large values of τ. The results that are shown in Fig. 4 indicate that the higher stiffness of input correction resulting from large values of τ can be compensated by increasing σ as well. As we are interested in small input corrections, we choose the solution with $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a for further analysis but we shall use the solutions with $\sigma = 1.0\,\mathrm{Wm}^{-2}$ and $\tau = 200$ a and with $\sigma = 1.5\,\mathrm{Wm}^{-2}$ and $\tau = 500$ a for comparative purposes. These three pairs of values of σ and τ are marked by bold points in Fig. 4.

5.5.3. *Model output*

Figs 5 and 6 show the observations of mean surface temperature and ocean heat content respectively, together with the posterior quantiles of the corresponding model output $\eta\{x(t)\}$. To assess the differences between observations and model outputs, we also give predictive intervals for a hypothetical independent observation at the same time point. In both Fig. 5 and Fig. 6 the results are shown for simulations without time-dependent parameters ($\sigma = 0$) and for the three pairs of hyperparameter values that are marked by big dots in Fig. 4. Without stochastic forcing, the 90% credible intervals for the model output appear too narrow and the model shows deficiencies especially in the mean surface temperature of the 1940s and 1950s and in the ocean heat content at the end of the 20th century. If a stochastic forcing is added, the 90% credible intervals are more realistic and the model deficiencies that were mentioned above are mostly corrected. This is true for any choice of values for σ and τ from among the three pairs identified in Fig. 4. For the surface temperature, in particular, the model residuals (which are defined as the observations minus the posterior median of the model output) are now much closer to having only a small correlation structure as described in Section 5.2. However, ocean heat data still show a strong temporal variation that is not compatible with our model assumptions. This decadal variability of the observed ocean heat content is not reproduced even by the most comprehensive models (Gregory *et al.*, 2004). Either all models underestimate decadal variability in ocean temperature, or the observations are biased. There is recent evidence supporting the latter assumption (Domingues *et al.*, 2008). A similar *caveat* applies to the surface temperature cooling in 1945 which is likely to be an artefact of ship-measured sea surface temperatures (Thompson *et al.*, 2008).

5.5.4. *Posterior distributions of constant parameters with stochastic forcing*

Fig. 7 shows the posterior marginals of the two most interesting parameters, climate sensitivity S and vertical ocean diffusivity K, without time-dependent parameter and with time-dependent forcing correction for the combinations of hyperparameter values σ and τ marked by big dots in Fig. 4. We would expect an increase in posterior uncertainty of these parameters when introducing a time-dependent parameter because conditioning reduces the variance on average. However, if the model with time constant parameters is misspecified, the opposite behaviour is also possible. The expected widening of the posterior marginal by introducing the time-dependent parameter is clearly visible for vertical ocean diffusivity K. There is a similar increase in posterior uncertainty for all three choices of hyperparameter values. The most striking feature shown in Fig. 7, however, is the strong dependence of the posterior of climate sensitivity on the standard deviation of the time-dependent parameter σ. We see the expected widening of the distribution for $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a but, as σ increases further, the posterior becomes narrower and shifts to values that are smaller than what more complex climate models predict (Knutti *et al.*, 2002; Forest *et al.*, 2002; Stainforth *et al.*, 2005).

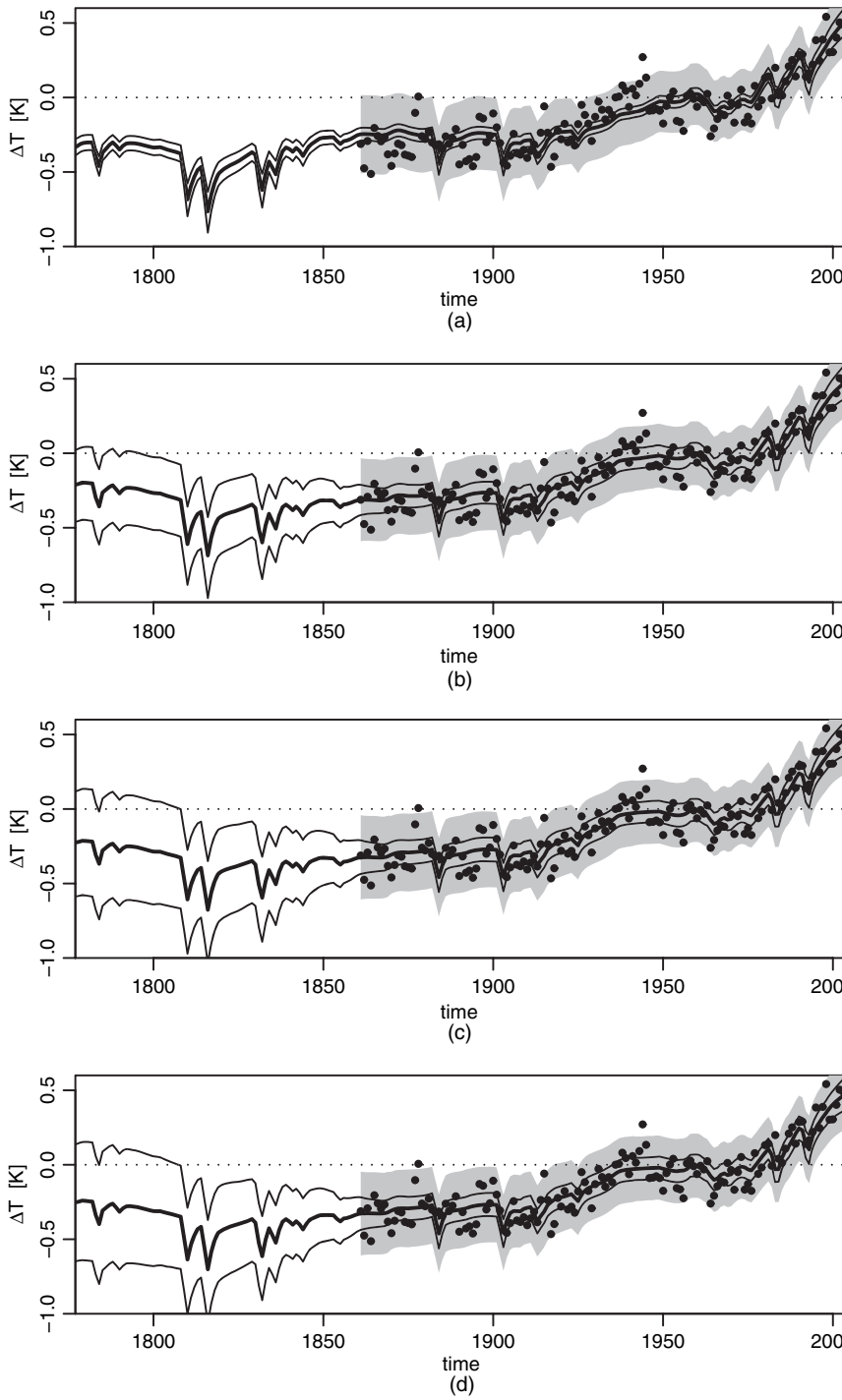

Fig. 5. Observations and model output for global surface temperature for various choices of the hyper-parameters of time-dependent forcing (———, 5%, 50% and 95% posterior quantiles for model output; ■, estimate of the 90% predictive interval for a hypothetical independent observation): (a) $\sigma = 0$; (b) $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a; (c) $\sigma = 1\,\mathrm{Wm}^{-2}$ and $\tau = 200$ a; (d) $\sigma = 1.5\,\mathrm{Wm}^{-2}$ and $\tau = 500$ a

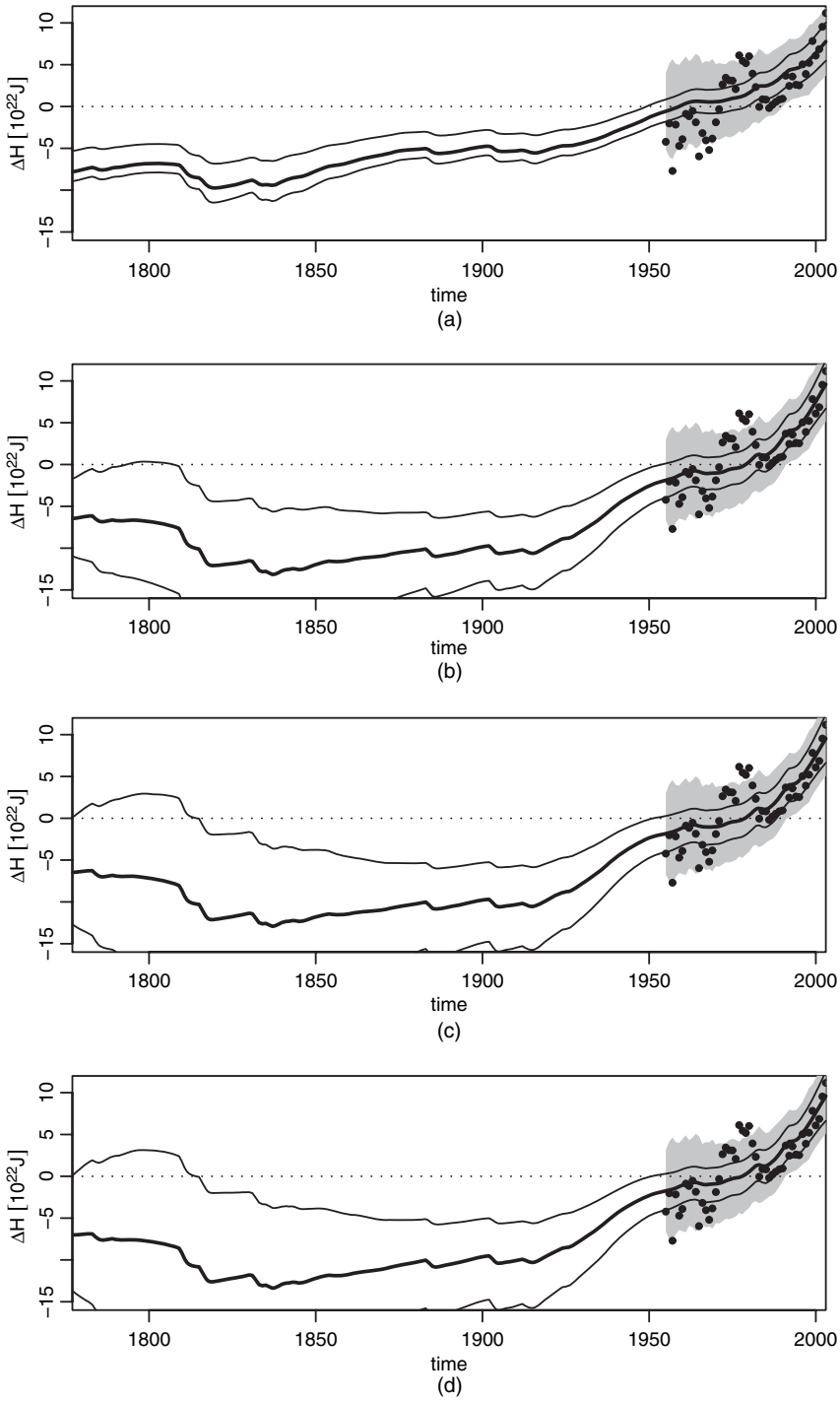

Fig. 6. Observations and model output for ocean heat uptake down to 700 m depth for various choices of the hyperparameters of time-dependent forcing (———, 5%, 50% and 95% posterior quantiles for model output; ■, estimate of the 90% predictive interval for a hypothetical independent observation): (a) $\sigma = 0$; (b) $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a; (c) $\sigma = 1\,\mathrm{Wm}^{-2}$ and $\tau = 200$ a; (d) $\sigma = 1.5\,\mathrm{Wm}^{-2}$ and $\tau = 500$ a

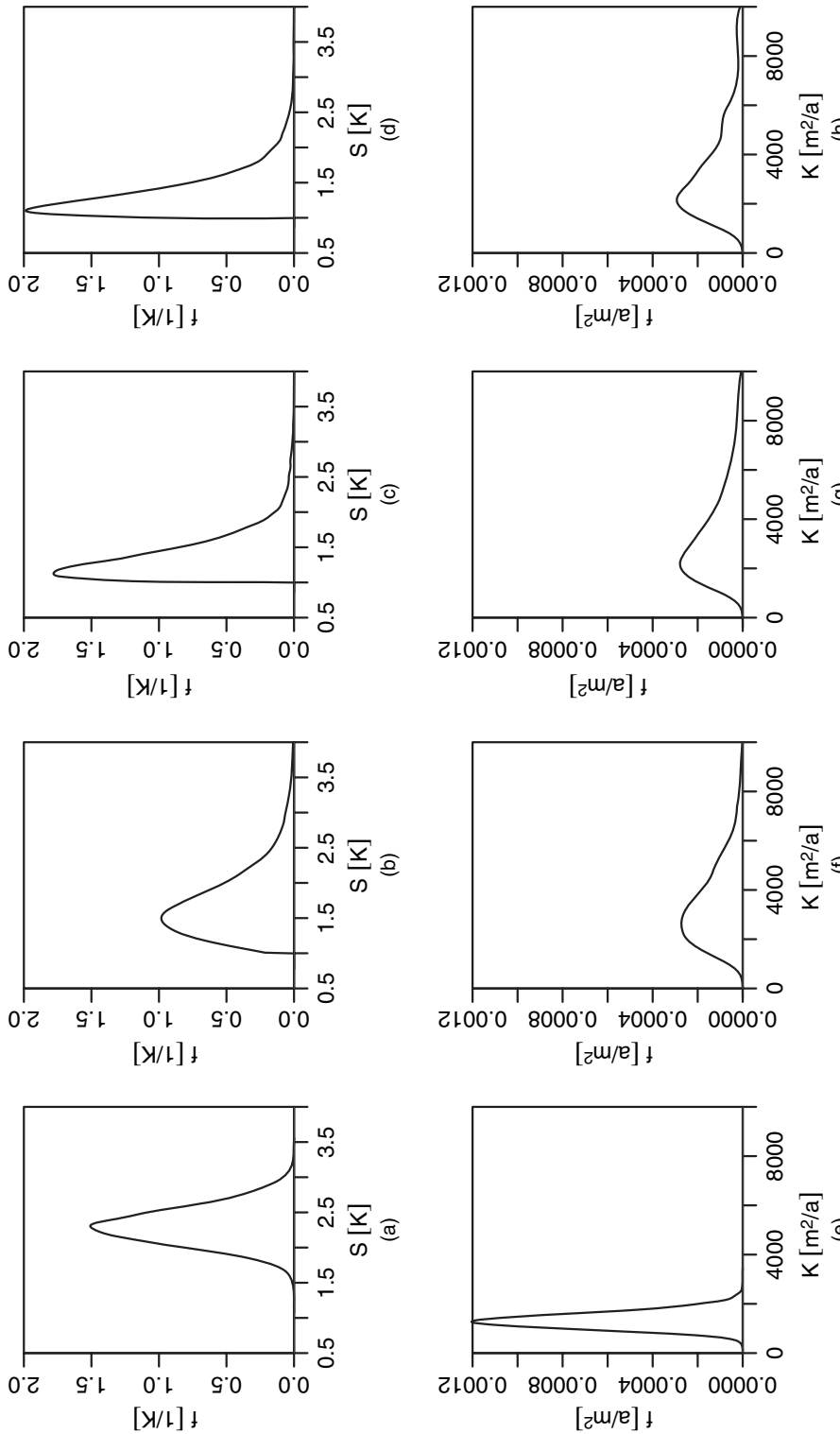

Fig. 7. Posterior distribution of (a)–(d) climate sensitivity S and (e)–(h) vertical ocean diffusivity for various values for the hyperparameters σ and τ: (a), (e) $\sigma = 0$; (b), (f) $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a; (c), (g) $\sigma = 1\,\mathrm{Wm}^{-2}$ and $\tau = 200$ a; (d), (h) $\sigma = 1.5\,\mathrm{Wm}^{-2}$ and $\tau = 500$ a

The shift to smaller values of climate sensitivity with increasing standard deviation of the time-dependent parameter could have been caused by the time-dependent parameter raising the input for modelling the temperature increase during the last 40 years or by avoiding a too high sensitivity to the sharp peaks of volcanic forcing. To analyse whether these effects are dominant causes of this behaviour, we redid all analyses by either

(a) keeping additional forcing at zero during the last 40 years of simulation,
(b) setting the factor on volcanic forcing to zero or
(c) combining (a) and (b).

The results showed that keeping additional forcing at zero during the last 40 years reduced but did not eliminate the shift and narrowing of the posterior for climate sensitivity for both cases with and without including volcanic forcing. In contrast, excluding volcanic forcing by setting the corresponding forcing factor to 0 did not have a large effect.

5.5.5. *Estimated stochastic forcing*

Fig. 8(a) shows the reconstructed forcing $\Delta F_{\mathrm{recon}}(\cdot)$ with s_{volc} set to 1. The posterior quantiles of the stochastic forcing term $\phi(\cdot)$ for $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a are shown in Fig. 8(b). For comparative purposes, Fig. 8(c) shows the posterior medians of ϕ for all three pairs of hyperparameters that are marked in Fig. 4 ($\sigma = 0.5\,\mathrm{Wm}^{-2}$, $\tau = 100$ a; $\sigma = 1.0\,\mathrm{Wm}^{-2}$, $\tau = 200$ a; $\sigma = 1.5\,\mathrm{Wm}^{-2}$, $\tau = 500$ a). Fig. 8 shows that the smaller flexibility of the solution with increasing values of τ can to some degree be compensated by larger values of σ. However, higher values of σ also lead to higher trend estimates of the forcing over the last 40 years.

We believe that the amount of forcing correction that is shown in Fig. 8 is unrealistically large. This indicates that we are correcting model structural errors by adjusting the input. This implies that we would need to make other model parameters time dependent to analyse the cause of the model deficiencies further. As heat storage at decadal timescales must be related to ocean heat storage, it seems to be logical to make parameters time dependent that relate to ocean heat uptake. For this reason, we next try to make vertical ocean diffusivity K time dependent.

5.5.6. *Results for time-dependent vertical ocean diffusivity*

The results with stochastic, time-dependent vertical ocean diffusivity K were not much different from those with constant K (not shown). We conclude that the parameter K does not have a sufficiently strong effect on the model output to correct for model deficiencies with regard to the interdecadal variability of ocean heat content change. This is in agreement with results from Knutti and Tomassini (2008) where the Bern2.5D climate model, an earth system model of intermediate complexity, was used. The findings indicate that atmospheric parameters such as climate sensitivity have a more decisive influence on change in ocean heat content than the vertical ocean diffusivity.

From a physical point of view this is an interesting and important result of the present work: with respect to the change in ocean heat content and, to a lesser degree, also for mean surface temperature, the discrepancies between model and data cannot be explained either by a forcing error or by a time-dependent effective vertical ocean diffusivity. This implies that the model structure of this simple climate model has deficiencies that may not be easily corrected without including more detail of ocean circulation. Two-dimensional dynamical models (e.g. Knutti *et al.* (2002)) have some representation of the large-scale circulation, but still have deficiencies. Three-dimensional models simulate ocean heat uptake much more realistically (Barnett *et al.*, 2001; Gregory *et al.*, 2004) but are computationally too expensive for such a method.

Other discrepancies may arise from biases in observations (Domingues *et al.*, 2008; Thompson

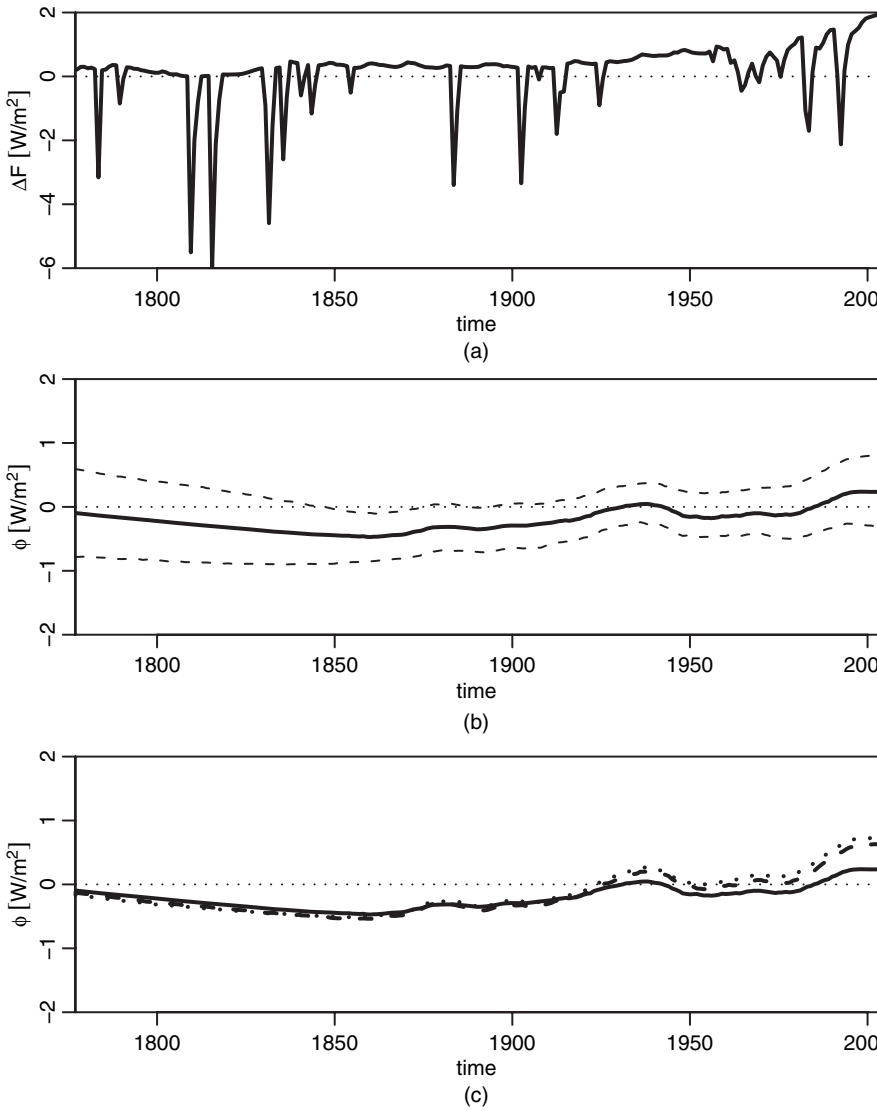

Fig. 8. (a) Time series of the reconstructed forcing $\Delta F_{\text{recon}}(t)$, (b) posterior 5%, 50% and 95% quantiles of the additional forcing term $\phi(t)$ for $\sigma = 0.5\ \text{Wm}^2$ and $\tau = 100$ a, and (c) posterior median of $\phi(t)$ for the three combinations of (σ, τ) of ($0.5\,\text{Wm}^{-2}$, 100 a) (———), ($1\,\text{Wm}^{-2}$, 200 a) (– – –) and ($1.5\,\text{Wm}^{-2}$, 500 a) (·····) that are marked in Fig. 4

et al., 2008). Also, the model obviously lacks internal climate variability. Although this is taken into account by the covariance matrix R of our error model, the internal variability component is estimated from a comprehensive model rather than observations, and may not be entirely adequate, particularly on long timescales and in the ocean (Gregory *et al.*, 2004). The use of a more complicated model would, however, make the problem computationally intractable.

5.5.7. Discussion

Introducing additional stochastic forcing led to significantly reduced systematic deviations of

model results from data, but making ocean diffusivity stochastic did not substantially improve model performance. This demonstrates that poor parameterization of ocean diffusivity cannot be the dominant model deficiency, whereas forcing errors could be one reason for observed systematic deviations. However, the amount of forcing correction that is needed to improve the model results is unrealistically large. This implies more severe model deficiencies, probably related to large-scale ocean circulation.

The major problems of the application of our approach to the climate model are the sensitivity of the marginal posterior of climate sensitivity to the variance of additional forcing (see Fig. 7) and the increasing positive trend of additional forcing during the last 40 years with increasing variance of additional forcing (see Fig. 8). These two effects are related to one another: increasing additional forcing allows the model to reproduce observed temperature increase at the end of the simulation period with smaller climate sensitivity. Further, as evidenced below, the model reacts too quickly to forcing changes, a tendency that is increased by larger values of climate sensitivity. Therefore, the algorithm keeps climate sensitivity small and 'misuses' the freedom of random forcing to increase forcing at the end of the simulation period.

The high sensitivity of the model response to short-term forcing changes can best be illustrated by its response to volcanic forcing. Several negative peaks in Fig. 5 demonstrate that the model reacts too strongly to the (negative) forcing peaks of volcano outbreaks. This occurs despite the fact that the inference process reduced volcanic forcing by more than 50%. As a further indication of this sensitivity, Fig. 9 shows scatter plots of $1/p(y_i|y_{-i}, \psi^k)$ for y_i corresponding to surface temperature data. The three pairs of hyperparameter values that were identified in Fig. 4 are considered. In Figs 9(a)–9(c), S is included in the Bayesian inference process, whereas in Figs 9(d)–9(f) S was fixed at the value of 3 K. In addition, the reconstructed forcing is shown. We can see that in all cases the model has difficulties in predicting some of the data points which are either outliers (such as the year 1878 in the surface temperature data) or correspond to years where strong troughs in the forcing occur due to volcanic eruptions. This model deficiency is stronger for large values of σ and also when we fix S at 3 K (note the different scales in Figs 9(a)–9(c) and 9(d)–9(f)).

Hence it seems that for large values of climate sensitivity the model cannot predict the temperature observations at the times of volcanic eruptions. This is because it shows a stronger response to volcanic forcing than the data.

6. Conclusions and outlook

We present a Bayesian technique for estimating time-dependent parameters of dynamic models. Results of this technique can be used for identifying and analysing model structure deficiencies and input errors of deterministic, dynamic simulation models. This extends earlier work based on discrete time parameters (Beck and Young, 1976; Beck, 1983; Young, 2001; Young *et al.*, 2001; Kristensen *et al.*, 2004) and earlier approaches of using continuous time parameters (Brun, 2002; Buser, 2003).

The technique is implemented by using a carefully designed Markov chain Monte Carlo algorithm as a technique to estimate continuous time stochastic parameters. The main idea consists of splitting the time interval into subintervals which reduce the rejection rate in the Metropolis–Hastings algorithm and accelerate convergence of the Markov chain (Buser, 2003). A conditional Ornstein–Uhlenbeck process with fixed end points is used as a proposal distribution for the time-dependent parameter on the different subintervals. The hyperparameters of the Ornstein–Uhlenbeck process are selected by a cross-validation criterion. In principle it is possible to include the hyperparameters in the estimation procedure, but this led to a tendency

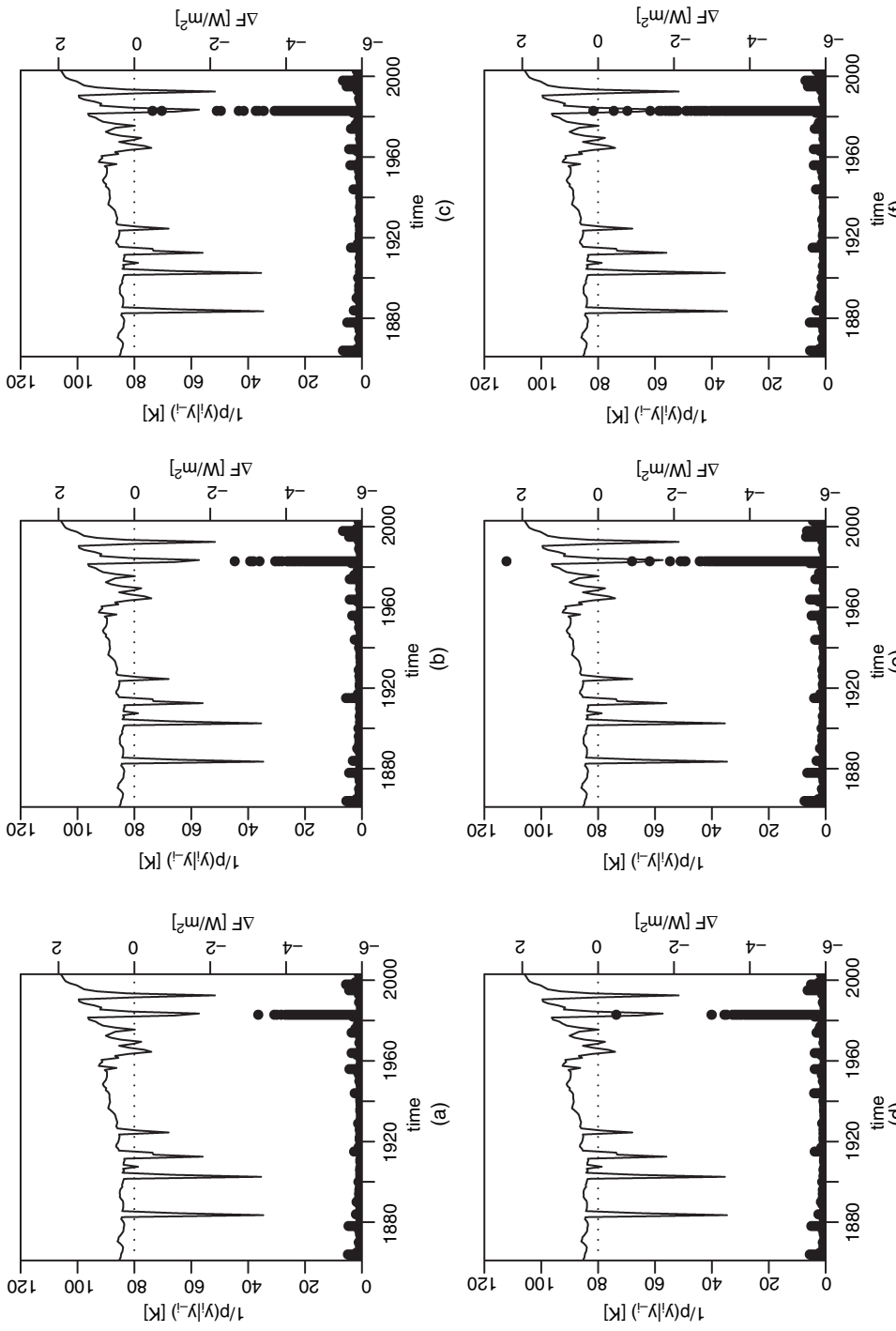

Fig. 9. (a)–(c) Scatter plots of the quantities $1/p(y_i|y_{-i}, \psi^k)$ for surface temperature with various choices of the hyperparameters of the time-dependent parameters with all constant parameters estimated and (d)–(f) analogous scatter plots with climate sensitivity kept at 3 K instead of including it in the Bayesian inference process (——, forcing time series): (a), (d) $\sigma = 0.5\,\mathrm{Wm}^{-2}$ and $\tau = 100$ a; (b), (e) $\sigma = 1\,\mathrm{Wm}^{-2}$ and $\tau = 200$ a; (c), (f) $\sigma = 1.5\,\mathrm{Wm}^{-2}$ and $\tau = 500$ a

to overfit the data unless a strong prior avoids this problem. As a consequence the estimates for the hyperparameters may become physically unrealistic.

We tested our algorithm by using a simple climate model. The algorithm was applied to estimate an additional stochastic radiative forcing contribution and vertical ocean diffusivity. Results show that the smoothing algorithm proposed converges well and can estimate the time-dependent parameters and their uncertainty effectively. The technique is well suited to detect and correct model deficiencies which can lead to an improved understanding of the physical context. The cross-validation scheme could constrain the correlation time but did not lead to a unique selection. Application of the additional forcing component led to a significant improvement of the fit, whereas making vertical ocean diffusivity time dependent did not.

As the climate example shows, our technique of using time-dependent parameters must be applied carefully. It adds many degrees of freedom, and it may be difficult to understand how the model uses this freedom to improve the fit. In our example, we found that it can create an unrealistically large, time-dependent correction to radiative forcing. In the current application, we thus could not profit substantially from application of the technique but rather must conclude that the model has severe deficiencies that cannot be corrected by realistic time-dependent variation of any of the parameters to which we applied the technique. Nevertheless, this is an important insight gained through the analysis. The climate model application should thus be interpreted as an idealized example to illustrate the method rather than an attempt to constrain climate sensitivity. The simplicity of the model and the added freedom by the additional forcing term lead to a distribution that is too narrow and biased towards values that are unrealistically low, in particular when the additional forcing is allowed to take large values. The range of climate sensitivities disagrees with ranges that were gained by using more comprehensive models (e.g. Forest *et al.* (2002), Knutti *et al.* (2002) and Stainforth *et al.* (2005)) and should therefore not be interpreted as being realistic for the real climate system.

We expect that, in general, when used thoughtfully, the technique that is presented in this paper can be a powerful tool for analysing and correcting model input errors and structural deficiencies. It can effectively aid in estimating time-dependent stochastic parameters, detecting model deficiencies, helping to improve the structure of the model and estimating uncertainties of stochastic influence factors. There are many ways to refine this approach. We could use more complicated stochastic process models or different hyperparameter values in different time periods. Moreover, there are other possibilities for dealing with time varying parameters, for instance maximum likelihood methods combined with spline-type regularizations. We believe, however, that the particular approach that is taken is not as important as the attempt to identify model deficiencies. Time varying parameters are a useful tool for this task.

Acknowledgements

We thank two referees for careful reading and constructive comments on the first submission. LT was supported by the Swiss National Science Foundation and MEB was supported in part by a grant from the US Environmental Protection Agency. However, this paper has not been subjected to the Agency's peer and policy review. Therefore, it does not necessarily reflect the views of the Agency and no official endorsement should be inferred.

References

Ammann, C. M., Meehl, G. A., Washington, W. M. and Zender, Ch. S. (2003) A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.*, **30**, 1657, doi 10.1029/2003GL016875.

Barnett, T. P., Pierce, D. W. and Schnur, R. (2001) Detection of anthropogenic climate change in the world's oceans. *Science*, **292**, 270–274.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, Ch.-H. and Tu, J. (2007) A framework for validation of computer models. *Technometrics*, **49**, 138–154.

Beck, M. B. (1983) Uncertainty, system identification and the prediction of water quality. In *Uncertainty and Forecasting of Water Quality* (eds M. B. Beck and G. van Straten). Berlin: Springer.

Beck, M. B. and Young, P. C. (1976) Systematic identification of DO-BOD model structure. *J. Environ. Engng Divsn Am. Soc. Civ. Engrs*, **103**, 902–927.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006) Retrospective exact simulations of diffusion sample paths with applications. *Bernoulli*, **12**, 1077–1098.

Brun, R. (2002) Learning from data: parameter identification in the context of large environmental simulation models. *Dissertation 14575*. Eidgenössiche Technische Hochschule, Zurich.

Buser, Ch. (2003) Differentialgleichungen mit zufällig zeitvariierenden Parametern. *Diploma Thesis*. Eidgenössiche Technische Hochschule, Zurich.

Collins, M., Tett, S. F. B. and Cooper, C. (2001) The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dynam.*, **17**, 61–81.

Cowles, M. K. and Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Am. Statist. Ass.*, **91**, 883–904.

Craig, P. S., Goldstein, M., Rougier, J. C. and Seheult, A. H. (2001) Bayesian forecasting for complex systems using computer simulators. *J. Am. Statist. Ass.*, **97**, 717–729.

Craig, P. S., Goldstein, M., Seheult, A. H. and Smith, J. A. (1996) Bayes linear strategies for matching hydrocarbon reservoir history. In *Bayesian Statistics 5* (eds J. M. Bernardo, J. O. Berger and A. P. Dawid), pp. 69–95. Oxford: Clarendon.

Cramér, H. and Leadbetter, M. R. (1967) *Stationary and Related Processes*. New York: Wiley.

Crowley, T. J. (2000) Causes of climate change over the past 1000 years. *Science*, **289**, 270–277.

Currin, C., Mitchell, T., Morris, M. and Ylvisaker, D. (1992) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Am. Statist. Ass.*, **86**, 953–963.

Domingues, C. M., Church, J. A., White, N. J., Gleckler, P. J., Wijffels, S. E., Barker, P. M. and Dunn, J. R. (2008) Improved estimates of upper-ocean warming and multi-decadal sea-level rise. *Nature*, **453**, 1090–1093.

Edmonds, J. A., Wise, M. A. and MacCracken, C. N. (1994) Advanced energy technologies and climate change: an analysis using the Global Change Assessment Model (GCAM). *Report PNL-9798*. Pacific Northwest Laboratory, Richmond.

Elerian, O., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, **69**, 959–993.

Forest, C. E., Stone, P. and Sokolov, A. (2006) Estimated PDFs of climate system properties including natural and anthropogenic forcings. *Geophys. Res. Lett.*, **33**, doi 10.1029/2005GL023977.

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

Gelfand, A. E. and Dey, D. K. (1994) Bayesian model choice: asymptotics and exact calculations. *J. R. Statist. Soc.* B, **56**, 501–514.

Gent, P. R. and Danabasoglu, G. (2004) Heat uptake and the thermohaline circulation in the community climate system model, version 2. *J. Clim.*, **17**, 4058–4069.

Goldstein, M. and Rougier, J. (2004) Probabilistic formulations for transferring inferences from mathematical models to physical systems. *SIAM J. Sci. Comput.*, **26**, 467–487.

Gregory, J. M., Banks, H. T., Stott, P. A., Lowe, J. A. and Palmer, M. D. (2004) Simulated and observed decadal variability in ocean heat content. *Geophys. Res. Lett.*, **31**, L15312, doi 10.1029/2004GL020258.

Heidelberger, P. and Welch, P. D. (1983) Simulation run length control in the presence of an initial transient. *Ops Res.*, **31**, 1109–1144.

IMAGE-team (2001)*The IMAGE 2.2 Implementation of the SRES Scenarios: a Comprehensive Analysis of Emissions, Climate Change and Impacts in the 21st Century*. Bilthoven: National Institute for Public Health and the Environment.

Intergovernmental Panel on Climate Change (1995) Climate change 1995: the science of climate change. In *Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (eds J. T. Houghton, L. G. Meira Filho, B. A. Callender, N. Harris, A. Kattenberg and K. Maskell). Cambridge: Cambridge University Press.

Intergovernmental Panel on Climate Change (2001) Climate change 2000: the scientific basis. In *Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds J. T. Houghton, Y. Ding, D. O. Griggs, M. Noguer, P. J. van der Linden and D. Xiaosu). Cambridge: Cambridge University Press.

Jones, P. D. and Moberg, A. (2003) Hemispheric and large-scale air temperature variations: an extensive revision and an update to 2001. *J. Clim.*, **16**, 206–223. (Available from `http://www.met-office.gov.uk/research/hadleycentre/CR_data/Annual/land+sst_web.txt`.)

Jones, P. D., Osborne, T. J., Briffa, K. R., Folland, C. K., Horton, E. B., Alexander, L. V., Parker, D. E. and

Rayner, N. A. (2001) Adjusting for sampling density in grid box land and ocean surface temperature time series. *J. Geophys. Res.*, **106**, 3371–3380.

Joos, F., Prentice, I. C., Sitch, S., Meyer, R., Hooss, G., Plattner, G.-K., Gerber, S. and Hasselmann, K. (2001) Global warming feedbacks on terrestrial carbon uptake under the IPCC emission scenarios. *Globl Biogeochem. Cyc.*, **15**, 891–907.

Kennedy, M. C. and O'Hagan, A (2001) Bayesian calibration of computer models (with discussion). *J. R. Statist. Soc.* B, **63**, 425–464.

Kloeden, P. E. and Platen, E. (1995) *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.

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

Knutti, R. and Tomassini, L. (2008) Constraints on the transient climate response from observed global temperature and ocean heat uptake. *Geophys. Res. Lett.*, **35**, L09701, doi 10.1029/2007GL032904.

Kristensen, N. R., Madsen, H. and Jørgensen, S. B. (2004) A method for systematic improvement of stochastic grey-box models. *Comput. Chem. Engng*, **28**, 1431–1449.

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. (2006) Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters. *J. Hydrol.*, **331**, 161–177.

Levitus, S., Antonov, J. and Boyer, T. (2005) Warming of the world ocean, 1955-2003. *Geophys. Res. Lett.*, **32**, L02604. (Available from `http://www.nodc.noaa.gov/DATA_ANALYSIS/temp/basin/hc1yr-wO-700m.dat`.)

Meinshausen, M. (2005) On the risk of overshooting 2°C. In *Avoiding Dangerous Climate Change* (eds H. J. S. Schellnhuber, W. Cramer, N. Nakicenovic , T. Wigley and G. Yohe). Cambridge: Cambridge University Press.

Myhre, G., Highwood, E. J., Shine, K. P. and Stordal, F. (1998) New estimates of radiative forcing due to well mixed greenhouse gases. *Geophys. Res. Lett.*, **25**, 2715–2718.

Oakley, J. and O'Hagan, A. (2002) Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, **89**, 769–784.

O'Hagan, A. (1992) Some Bayesian numerical analysis. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 345–363. Oxford: Oxford University Press.

Osborne, T. J. and Wigley, T. M. L. (1994) A simple model for estimating methane concentration and lifetime variations. *Clim. Dynam.*, **9**, 181–193.

Raper, S. C. B., Gregory, J. M. and Osborne, T. J. (2001) Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results. *Clim. Dynam.*, **17**, 601–613.

Roberts, G. O. and Stramer, O. (2001) On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603–621.

Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A., Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J. and Allen, M. R. (2005) Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406.

Stott, P. A., Tett, S. F. B., Jones, G. S., Allen, M. R., Ingram, W. J. and Mitchell, J. F. B. (2001) Attribution of twentieth century temperature change to natural and anthropogenic causes. *Clim. Dynam.*, **17**, 1–21.

Thompson, D. W. J., Kennedy, J. J., Wallace, J. M. and Jones, P. D. (2008) A large discontinuity in the mid-twentieth century in observed global-mean surface temperatures. *Nature*, **453**, 646–648.

Tomassini, L. (2007) Case studies in estimation and representation of uncertainty in climate modelling. *Dissertation 17109*. Eidgenössiche Technische Hochschule, Zurich.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. and Verstraten, J. M. (2005) Improved treatment of uncertainty in hydrological modeling: combining the strength of global optimization and data assimilation. *Wat. Resour. Res.*, **41**, doi 10.1029/2004WR003414.

Wigley, T. M. L. and Raper, S. C. B. (1987) Thermal expansion of sea water associated with global warming. *Nature*, **330**, 127–131.

Wigley, T. M. L. and Raper, S. C. B. (1992) Implications for climate and sea level of revised IPCC emissions scenarios. *Nature*, **357**, 293–300.

Wigley, T. M. L. and Raper, S. C. B. (2001) Interpretation of high projections for global-mean warming. *Science*, **293**, 451–454.

Young, P. C. (2001) The identification and estimation of nonlinear stochastic systems. In *Nonlinear Dynamics and Statistics* (ed. A. I. Mees), pp. 127–166. Boston: Birkhäuser.

Young, P. C., McKenna, P. and Bruun, J. (2001) Identification of non-linear stochastic systems by state dependent parameter estimation. *Int. J. Control*, **74**, 1837–1857.