

# A Smoothing Algorithm for Estimating Stochastic, Continuous-Time Model Parameters and an Application to a Simple Climate Model

Lorenzo Tomassini

*Eawag: Swiss Federal Institute of Aquatic Science and Technology  
CH-8600 Dübendorf  
Switzerland*

Peter Reichert

*Eawag: Swiss Federal Institute of Aquatic Science and Technology  
CH-8600 Dübendorf  
Switzerland*

Hans R. Künsch

*Seminar for Statistics  
ETH Zurich  
CH-8092 Zürich  
Switzerland*

Christoph Buser

*Seminar for Statistics  
ETH Zurich  
CH-8092 Zürich  
Switzerland*

Mark E. Borsuk

*Dartmouth College  
Hanover, NH 03755  
USA*

**Summary.** To address systematic discrepancies between model simulations and measured data, we propose to make selected parameters in the model time-variant by modeling them as continuous-time stochastic processes. We present an algorithm for Bayesian estimation of such parameters that includes some special adaptations of the Markov chain Monte Carlo method. The algorithm consists of splitting the overall time period into subintervals, over each of which we use a conditional Ornstein-Uhlenbeck process with fixed endpoints as the proposal distribution in a Metropolis-Hastings algorithm. The hyperparameters of the stochastic process are then selected using a cross-validation criterion which maximizes a pseudo-likelihood value, for which we have derived a computationally efficient estimator. We tested our algorithm using a simple climate model in which an additional stochastic forcing component is introduced. The results show that the algorithm behaves well, is computationally tractable, and improves model fit to the data. The additional forcing term was found to strongly alter the posterior density of

*Address for correspondence:* Lorenzo Tomassini, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, CH-8600 Dübendorf, Switzerland.  
Email: [lorenzo.tomassini@eawag.ch](mailto:lorenzo.tomassini@eawag.ch)

the other, time-constant parameters. By a careful analysis of the results, we are able to identify a structural deficit of the simple climate model as the cause of this unexpected behavior.

*Keywords: state-space models, smoothing, stochastic process, cross-validation, climate modelling.*

## 1. Introduction

With the development of new monitoring technologies, environmental data are characterized by ever-increasing temporal resolution combined with decreasing measurement noise. This typically makes systematic discrepancies between the output of a deterministic model and the measurements obvious, as model structures are rarely good enough to reproduce the behaviour of the environmental system at the level of precision of modern measurement devices. In particular, the assumption of independent and identically distributed measurement errors which justifies the use of nonlinear least squares methods for parameter identification is often invalidated. This makes it difficult to quantify the uncertainty of estimated parameters and predictions.

The above situation leads to a demand for statistical techniques that can account for systematic deviations and suggest model adaptations that would lead to improved model fit. We suggest that this can be accomplished by considering some of the parameters or inputs of the model to be time varying and modeling them as stochastic processes. Bayesian techniques then allow us to estimate the values of both the fixed and time-dependent parameters, as well as the parameters of the stochastic process itself. An analysis of the identified temporal behaviour in parameters can then support the search for causes of this behaviour. If these causes can be modelled deterministically, then the model structure can be extended to represent the newly identified cause-effect relationships. On the other hand, if these causes vary randomly, or if it seems more realistic to describe the variation by a random process, then this random process can be included in the model, thus giving more realistic estimates of uncertainty in model predictions.

The approach described above has been applied for more than 20 years to discrete-time systems using the extended Kalman filtering technique for statistical inference, see e.g. Beck (1983) or Kristensen et al. (2004). In this paper, we: (i) extend the approach from discrete to continuous systems, and (ii) develop a numerical scheme for Bayesian implementation applicable to nonlinear models without relying on linearization techniques. Because we concentrate on application to environmental systems, we apply the stochastic approach to time-dependent model parameters and inputs only rather than to modeled quantities themselves. This is because, in contrast to other approaches (e.g. 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 true violations of fundamental laws. Therefore, models should describe these uncertainties explicitly and not make the conservation equations themselves stochastic.

We demonstrate our approach by application to a simple global climate model that has shown systematic discrepancies in comparisons against data.

## 2. Governing Equations

Let  $x : \mathbb{R} \rightarrow \mathbb{R}^d$  be the model function. In our case, it is assumed to be the solution of a differential equation of the form

$$\frac{dx(t)}{dt} = f(x(t), \phi_t, \theta), \quad x(0) = x_0, \quad t \in [0, T] =: I, \quad (1)$$

where  $t$  denotes time,  $\theta \in \mathbb{R}^n$  is a vector of constant parameters and  $\phi_t \in \mathbb{R}$  is a time varying parameter. 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. Note that (1) implies that  $x$  depends deterministically on  $t, \phi_t$  and  $\theta$ .

In the following  $x$  will be considered either as a function of  $t$ , or as a function of  $t$  and all model parameters  $\phi_t$  and  $\theta$ . It should be clear from the context which viewpoint was adopted.

Let  $y \in \mathbb{R}^p$  denote the vector of observations, and  $t^{\text{ob}} \in \mathbb{R}^p$  the vector of times at which the observations occurred. We assume that, conditional on the model output  $X$ , the observations are jointly Gaussian with mean vector  $x_{t^{\text{ob}}} = (x_{t_1^{\text{ob}}}, \dots, x_{t_p^{\text{ob}}})$  and a known covariance matrix  $R$ :

$$Y \mid X(t_1^{\text{ob}}) = x_{t_1^{\text{ob}}}, \dots, X(t_p^{\text{ob}}) = x_{t_p^{\text{ob}}} \sim \mathcal{N}(x_{t^{\text{ob}}}, R). \quad (2)$$

The covariance matrix  $R$  takes into account both measurement errors and short time fluctuations in  $X$  that are not represented in the model (1).

Our approach is Bayesian: we consider  $(\phi_t)_{t \in I}$  and  $\theta$  as realizations of random vectors  $(\Phi_t)_{t \in I}$  and  $\Theta$ . Our goal is to estimate the posterior probability densities of their distributions, given the observations  $Y = y$ .

### 2.1. Mean-reverting Ornstein-Uhlenbeck process

In our development, we will assume that the time dependent parameter  $\Phi_t$  follows a mean-reverting Ornstein-Uhlenbeck process (see for example Kloeden and Platen (1995)). This is a continuous stochastic process which is the solution of the Itô stochastic differential equation

$$d\Phi_t(\omega) = -\gamma(\Phi_t(\omega) - \bar{\Phi}) dt + \sigma_W \cdot dW_t(\omega) \quad (3)$$

The first term on the right hand side of the equation causes a drift of  $\Phi_t$  towards the mean  $\bar{\Phi}$ , the second term describes the random fluctuations induced by the increments of a Brownian motion  $W_t(\omega)$ .

The solution of (3) is available explicitly, namely for  $t > s$

$$\Phi_t = \bar{\Phi} + e^{-\gamma(t-s)}(\Phi_s - \bar{\Phi}) + \sigma_W \int_s^t e^{-\gamma(t-u)} dW_u.$$

This implies the Markov property (only the present state is relevant for the future development). Moreover, because the increments  $dW_t$  are normal with mean zero and covariance  $E(dW_t dW_s) = \delta_{ts} dt$ , it follows that  $\int_s^t e^{-\gamma(t-u)} dW_u$  is also normal with mean zero and variance  $\int_s^t e^{-2\gamma(t-u)} du$ . In other words

$$\Phi_t \mid \Phi_s = \phi_s \sim \mathcal{N}\left(\bar{\Phi} + e^{-\gamma(t-s)}(\phi_s - \bar{\Phi}), \frac{\sigma_W^2}{2\gamma} (1 - e^{-2\gamma(t-s)})\right) \quad (4)$$

In particular, in the limit  $t \rightarrow \infty$ , we obtain the invariant distribution  $\mathcal{N}(\bar{\Phi}, \frac{\sigma_W^2}{2\gamma})$ . If we choose  $\Phi_{t_0}$  according to this invariant distribution, then the process is Gaussian and strictly stationary.

For further use, we define the characteristic time of the process and the unconditional variance of  $\Phi_t$  as:

$$\tau := \frac{1}{\gamma}, \quad \sigma^2 := \lim_{s \rightarrow -\infty} \text{Var}(\Phi_t | \Phi_s = \phi_s) = \frac{\sigma_W^2}{2\gamma}. \quad (5)$$

We will also need the distribution of a mean-reverting Ornstein-Uhlenbeck process conditional on start and end values. In order to derive these, we consider for  $s < t < u$  the following formula for the conditional density

$$p(\phi_t | \Phi_s = \phi_s, \Phi_u = \phi_u) = \frac{p(\phi_t, \phi_u | \Phi_s = \phi_s)}{p(\phi_u | \Phi_s = \phi_s)} = \frac{p(\phi_t | \phi_s)p(\phi_u | \phi_t)}{p(\phi_u | \phi_s)}.$$

Since all these conditional densities on the right are normal with means and variances given by (4), we obtain by straightforward algebraic manipulations that  $\Phi_t$  given  $\Phi_s = \phi_s$  and  $\Phi_u = \phi_u$  is normal with

$$\mathbb{E}[\Phi_t | \Phi_s = \phi_s, \Phi_u = \phi_u] = \bar{\Phi} + \frac{e^{-\gamma(t-s)}(1 - e^{-2\gamma(u-t)})}{1 - e^{-2\gamma(u-s)}}(\phi_s - \bar{\Phi}) + \frac{e^{-\gamma(u-t)}(1 - e^{-2\gamma(t-s)})}{1 - e^{-2\gamma(u-s)}}(\phi_u - \bar{\Phi}) \quad (6)$$

and

$$\text{Var}(\Phi_t | \Phi_s = \phi_s, \Phi_u = \phi_u) = \frac{\sigma_W^2}{2\gamma} \frac{(1 - e^{-2\gamma(t-s)})(1 - e^{-2\gamma(u-t)})}{1 - e^{-2\gamma(u-s)}}. \quad (7)$$

Essentially the same argument shows that, conditional on the start and the end value,  $(\Phi_t)$  is still a Markov process. Namely, for  $t_0 < t_1 < \dots < t_n$ , the joint conditional density given  $\Phi_{t_0} = \phi_{t_0}$  and  $\Phi_{t_n} = \phi_{t_n}$  can be written as follows:

$$\begin{aligned} p(\phi_{t_1}, \dots, \phi_{t_{n-1}} | \Phi_{t_0} = \phi_{t_0}, \Phi_{t_n} = \phi_{t_n}) &= \frac{p(\phi_{t_1}, \dots, \phi_{t_n} | \Phi_{t_0} = \phi_{t_0})}{p(\phi_{t_n} | \Phi_{t_0} = \phi_{t_0})} \\ &= \prod_{i=1}^{n-1} \frac{p(\phi_{t_i} | \phi_{t_{i-1}})p(\phi_{t_n} | \phi_{t_i})}{p(\phi_{t_n} | \phi_{t_{i-1}})}. \end{aligned}$$

Note that all terms  $p(\phi_{t_n} | \phi_{t_i})$  except the first and the last one cancel on the right-hand-side. Hence, the last equality follows from the Markov property. This shows that the joint density is the product of the transition densities, and this is precisely the Markov property. In particular, simulation of  $(\Phi_t)$  on a fine grid given start and end values is straightforward.

### 3. Estimation Techniques

Let  $\xi = (\sigma, \tau)$  denote the two-dimensional vector of hyperparameters of the Ornstein-Uhlenbeck process (3).  $\xi$  may be considered as a realization of a random vector  $\Xi$  and included in the estimation procedure. We then would like to estimate the densities of  $(\Phi_t)_{t \in I}$ ,  $\Theta$  and  $\Xi$ , taking the full available data set into account. In the state space model framework, this technique is called ‘‘smoothing’’. To this end, we use a Markov chain Monte Carlo (MCMC) algorithm with some special adaptations to the present situation (Buser (2003)). We briefly review the basic MCMC recipe using the notation introduced above. We assume that all probability densities exist.

### 3.1. Markov chain Monte Carlo

We would like to estimate the density  $p(\phi_t, \theta, \xi | y)$ , which is proportional to  $p(\phi_t, \theta, \xi, y)$ .

According to the Gibbs sampler MCMC algorithm, to draw a sample from  $p(\phi_t, \theta, \xi, y)$  we generate recursively,

- (a)  $\theta^{k+1}$  according to  $p(\theta | \phi_t^k, \xi^k, y)d\theta = p(\theta | \phi_t^k, y)d\theta \propto p(\theta)p(y | \phi_t^k, \theta)d\theta$
- (b)  $\xi^{k+1}$  according to  $p(\xi | \phi_t^k, \theta^{k+1}, y)d\xi = p(\xi | \phi_t^k)d\xi \propto p(\xi)p(\phi_t^k | \xi)d\xi$
- (c)  $\phi_t^{k+1}$  according to  $p(\phi_t | \theta^{k+1}, \xi^{k+1}, y)d\phi_t \propto p(\phi_t | \xi^{k+1})p(y | \phi_t, \theta^{k+1})d\phi_t$

To generate these samples we use the Metropolis-Hastings recipe. We elaborate more on the details of this procedure in Section 3.2 and describe our adaptations of the MCMC algorithm which are appropriate for the present context.

Although the hyperparameters  $\xi$  of the Ornstein-Uhlenbeck process can in principle be included in the estimation procedure as described above, this generally leads to slow convergence and identifiability problems for  $\xi$ . To select the values for the hyperparameters  $\xi$  we will therefore resort to a cross-validation criterion (Gelfand and Dey (1994)). This is discussed in detail in Section 4. In the following,  $\xi$  is therefore considered to be fixed.

### 3.2. Adaptations

#### 3.2.1. Interval splitting and discretization

According to the Metropolis-Hastings algorithm, we have to first choose proposal distributions to use in drawing samples from the distributions described in Section 3.1. Samples from these proposal distributions are then accepted or rejected with an acceptance probability  $r$  that depends on the distribution from which one ultimately wants to sample. More precisely, suppose we want to generate a sample of a random variable  $X$  with density  $p(x)$ . We choose a proposal density  $q$ . Assuming the  $k$ -th sample  $x^k$  is given, the following steps are carried out repeatedly:

- (a) generate a draw  $\tilde{x}^k$  of  $\tilde{X}_k \sim q(\tilde{x}_k | x^k)d\tilde{x}_k$  and independently a draw  $u_k$  of  $U \sim \text{Uniform}(0, 1)$ ,
- (b) compute the acceptance probability  $r = \min\left(1, \frac{p(\tilde{x}^k)q(x^k|\tilde{x}^k)}{p(x^k)q(\tilde{x}^k|x^k)}\right)$ ,
- (c) if  $u_k \leq r$  set  $x^{k+1} = \tilde{x}^k$ , otherwise  $x^{k+1} = x^k$

Whereas this method can be used to update the constant parameters  $\theta$  (step (a) in Section 3.1), proposing a new value of  $\Phi_t$  on the whole interval  $I$  where the model  $x$  is defined (compare with (1)) is problematic. It is difficult to find a proposal distribution that leads to a reasonable acceptance probability. One possible solution is to split up the interval  $I$  into  $N$  subintervals,  $I_l, l = 1, \dots, N$ , and propose  $\Phi_{t,j} := (\Phi_t)_{t \in I_j}$  separately. The procedure is described in more detail in the next subsection.

To perform computations and simulate from an Ornstein-Uhlenbeck process for  $\Phi_t$  on the time interval  $I$ , we discretize time and split  $I$  in subintervals of equal lengths. Let  $\Delta t^{OU}$  be a time increment, supposed to be small in comparison with the difference between time points at which observations occurred. Suppose  $I = [0, T]$  and  $\Delta t^{OU} = \frac{T}{A}$ ,  $A \in \mathbb{N}$ , and set  $t_n^{OU} = n \cdot \Delta t^{OU}$ ,  $n = 0, \dots, A$ . Assume  $\Phi_{t_k^{OU}} = \phi_{t_k^{OU}}$  is given, and we want to generate a value for  $\Phi_{t_{k+1}^{OU}}$  according to the mean reverting Ornstein-Uhlenbeck process (3). To this end we draw a sample from a normal distribution with expectation according to (6) and

variance according to (7). For times  $t \in [t_k^{OU}, t_{k+1}^{OU}]$  the value of  $\Phi_t$  is determined by linear interpolation.

In the example in Section 5, the described approximation is tacitly assumed, but since the techniques presented in this paper do not depend on a specific numerical scheme to simulate from the Ornstein-Uhlenbeck process, in the following sections we will not reflect the chosen discretization scheme in the notation.

### 3.2.2. Proposal densities

Suppose  $\phi_t^k$  is given. We will not propose  $\phi_t^{k+1}$  in one step (step (c) in Section 3.1), because the acceptance probability  $r$  would be small and convergence of the algorithm slow in this case.

Instead we divide the interval  $I$  randomly into  $N$  subintervals  $I_l$ ,  $l = 1, \dots, N$ , with equal average lengths (Buser (2003)). Suppose  $I_l = [t_l^{\text{start}}, t_l^{\text{end}}]$ ,  $l = 1, \dots, N$ . In consecutive steps we propose  $\phi_{t,j}^{k+1}$  while the value of  $(\Phi_t)_{t \in I \setminus I_j}$  remains fixed. On  $I_j$  we use the distribution of an Ornstein-Uhlenbeck process conditional on  $\phi_{t_j^{\text{start}}}^k$  and  $\phi_{t_j^{\text{end}}}^k$  (see Section 2.1) as the proposal distribution for  $\tilde{\phi}_{t,j}^k$ . We calculate the acceptance probability  $r$  (see Subsection 3.2.3 for more details) and accept or reject according to the Metropolis-Hastings recipe. In this way, we go through all the subintervals  $I_j$  from  $j = 1$  to  $j = N$ . We then set  $\phi_t^{k+1} = (\phi_{t,1}^{k+1}, \dots, \phi_{t,N}^{k+1})$ .

In other words, the stages of the procedure are as follows: for  $j = 1$  to  $N$

- (a) set  $\tilde{\phi}_t^k = \phi_t^k$  for  $t \notin I_j$ ,
- (b) for  $t \in I_j$  draw  $\tilde{\phi}_t^k$  according to a conditional Ornstein-Uhlenbeck process with  $\tilde{\phi}_{t_j^{\text{start}}}^k = \phi_{t_j^{\text{start}}}^k$  and  $\tilde{\phi}_{t_j^{\text{end}}}^k = \phi_{t_j^{\text{end}}}^k$ ,
- (c) accept or reject the proposal  $\tilde{\phi}_t^k$  according to the acceptance probability  $r$ .

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

In order to draw samples from  $p(\theta)p(y | \phi_t^{k+1}, \theta)$  (step (a) in Section 3.1) according to the Metropolis-Hastings recipe, we choose a symmetric random walk to propose  $\theta^{k+1}$ . As already mentioned,  $\xi$  is considered to be fixed at a given value, and step (b) in Section 3.1 is therefore omitted.

### 3.2.3. Acceptance probabilities

The acceptance probability for sampling from  $p(\theta)p(y | \phi_t^{k+1}, \theta)$  is as described in Section 3.2.1.

Here, we elaborate more on the formula for the acceptance probability of  $\tilde{\phi}_{t,j}^k$  on the subinterval  $I_j$ . We have

$$r = \min \left( 1, \frac{p(\tilde{\phi}_{t,j}^k | \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi) p(y | \tilde{\phi}_{t,j}^k, (\phi_t^{\text{current}})_{t \in I \setminus I_j}, \theta^k) q(\phi_{t,j}^k | \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi)}{p(\phi_{t,j}^k | \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi) p(y | \phi_t^{\text{current}}, \theta^k) q(\tilde{\phi}_{t,j}^k | \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi)} \right) \quad (8)$$

where  $\phi_t^{\text{current}} = \phi_t^k$  for  $t \in I_s$  with  $s > j$ , and  $\phi_t^{\text{current}} = \phi_t^{k+1}$  for  $t \in I_s$  with  $s < j$ . But recall that our proposal distribution  $q(\tilde{\phi}_{t,j}^k | \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi)$  is the distribution of a

conditional Ornstein-Uhlenbeck process on  $I_j$ . Thus  $q(\tilde{\phi}_{t,j}^k \mid \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi) = p(\tilde{\phi}_{t,j}^k \mid \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi)$  and  $q(\phi_{t,j}^k \mid \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi) = p(\phi_{t,j}^k \mid \phi_{t_j^{\text{start}}}^k, \phi_{t_j^{\text{end}}}^k, \xi)$ . Therefore

$$r = \min \left( 1, \frac{p(y \mid \tilde{\phi}_{t,j}^k, (\phi_t^{\text{current}})_{t \in I \setminus I_j}, \theta^k)}{p(y \mid \phi_t^{\text{current}}, \theta^k)} \right). \quad (9)$$

#### 4. Cross-Validation

For a given choice of the hyperparameters  $\gamma$  and  $\sigma$  of the Ornstein-Uhlenbeck process, we assume that we can simulate from the posterior, i.e. the conditional distribution of  $\Phi = (\Phi_t)$  given  $Y = (Y_1, \dots, Y_p)$ . For this, we have to discretize time, but the discretization can be much finer than the sequence of observation times  $t_i^{\text{ob}}$  (see section 3.2). We abbreviate  $x_{t_i} := x(t_i^{\text{ob}})$  and denote by  $x(\phi)$  the vector  $(x_{t_1}(\phi), \dots, x_{t_p}(\phi))$ . For the conditional distribution of  $Y$  given  $\Phi = \phi$ , we only need  $x(\phi)$ , that is

$$p(y \mid \phi) = p(y \mid x(\phi)) = f_\epsilon(y_1 - x_{t_1}(\phi), \dots, y_p - x_{t_p}(\phi)),$$

where  $f_\epsilon$  denotes the joint density of the observation errors (see (2)).

The problem is how to choose the hyperparameters  $\gamma$  and  $\sigma$ . Although they can in principle be included in the estimation procedure, as mentioned in Section 3.1, this generally leads to slow convergence and identifiability problems for  $\xi$ . Therefore, we introduce a cross-validation criterion (Gelfand and Dey (1994)). That is, we determine the hyperparameters such that the so-called pseudo likelihood

$$psl := \sum_{i=1}^p \log p(y_i \mid y_j, j \neq i) \quad (10)$$

is maximal.

Calculating  $p$  posterior distributions of the form  $p(y_i \mid y_j, j \neq i)$  as in (10) is computationally very expensive. The following lemma allows for estimating  $p(y_i \mid y_j, j \neq i)$  for all  $i$  from a sample of the full posterior distribution  $p(x \mid y)dx$ .

With the definition  $\{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_p\} := y_{-i}$  we have:

**Lemma.**

$$p(y_i \mid y_{-i}) = \frac{1}{\int p(y_i \mid y_{-i}, x)^{-1} p(x \mid y) dx} \quad (11)$$

*Proof.* By the definition of the conditional density

$$p(y_i \mid y_{-i}) = \frac{p(y)}{p(y_{-i})}.$$

By the law of total probability

$$p(y_{-i}) = \int p(y_{-i} \mid x(\phi)) p(\phi) d\phi = \int \frac{p(y_{-i} \mid x(\phi))}{p(y \mid x(\phi))} p(y \mid x(\phi)) p(\phi) d\phi.$$

By the definition of the conditional density, we have

$$p(y \mid x(\phi)) p(\phi) = p(\phi \mid y) p(y),$$

and

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

The proof is completed by combining all these identities.  $\square$

With the help of the above lemma, we can introduce an estimator for  $p(y_i | y_{-i})$  by averaging  $p(y_i | y_{-i}, x(\phi)^{(m)})^{-1}$  where  $\{x(\phi)^{(m)}\}_{m=1, \dots, M}$  now denotes a sample generated by a simulation from the full posterior distribution  $p(x(\phi) | y)dx$ . From (2) we have:

$$p(y | x) = (2\pi)^{p/2} \det(R^{-1})^{1/2} \exp\left(-\frac{1}{2} \sum_{j,k=1}^p (y_j - x_{t_j})^\top R_{jk}^{-1} (y_k - x_{t_k})\right).$$

To obtain  $p(y_i | y_{-i}, x)$ , we only need to collect all terms that contain  $y_i$  and then to find the proper normalization by completing the square. This shows that  $p(y_i | y_{-i}, x(\phi))$  is a normal density with variance  $1/R_{ii}^{-1}$  and mean

$$x_{t_i} - \sum_{j \neq i} \frac{R_{ij}^{-1}}{R_{ii}^{-1}} (y_j - x_{t_j}).$$

Hence in this case, the Monte Carlo estimate  $\hat{p}(y_i | y_{-i})$  of  $p(y_i | y_{-i})$  is

$$\hat{p}(y_i | y_{-i}) = \frac{M(R_{ii}^{-1})^{1/2}}{(2\pi)^{1/2} \sum_{k=1}^M \exp\left(\frac{R_{ii}^{-1}}{2} \left(y_i - x_{t_i}^{(k)} + \sum_{j \neq i} \frac{R_{ij}^{-1}}{R_{ii}^{-1}} (y_j - x_{t_j}^{(k)})\right)^2\right)}. \quad (12)$$

Our estimator of the pseudo-likelihood therefore becomes

$$\widehat{psl} := \sum_{i=1}^p \log(\hat{p}(y_i | y_{-i})). \quad (13)$$

## 5. Application to a Simple Climate Model

We exemplify our techniques by analyzing a hemispherically-averaged, upwelling-diffusion, energy balance model of global climate. The model has previously been used to study the relationship between greenhouse gas forcing, global mean temperature change, and sea-level rise (Wigley and Raper (1987), Wigley and Raper (1992), Raper et al. (1996)).

### 5.1. The climate model

The climate model consists of two land and two ocean boxes (representing the Northern and Southern Hemisphere). The two ocean boxes are each divided into 40 vertical layers. The two top layers, assumed to represent the ocean mixed layers, absorb the energy of solar radiation (see Figure 1). It is assumed that no energy is absorbed above land. All variables in the model represent deviations from an equilibrium state corresponding to preindustrial radiative forcing.

On time scales relevant to climate change, the atmosphere may be assumed to be in equilibrium with the underlying oceanic mixed layer. This leads to the differential equation

$$C_{ml} \frac{d\Delta T_0}{dt} = \Delta Q - \frac{F^{\text{double}}}{S} \cdot \Delta T_0 - \Delta F \quad (14)$$

where  $\Delta T_0$  is the temperature of the oceanic mixed layer and  $C_{\text{ml}}$  its effective heat capacity.  $\Delta Q$  denotes the radiative forcing,  $\Delta F$  the net heat flux into the ocean,  $S$  represents climate sensitivity (the equilibrium increase in global mean surface temperature for a doubling of the atmospheric  $\text{CO}_2$ -concentrations), and  $F^{\text{double}}$  the forcing that corresponds to a doubling of  $\text{CO}_2$ -concentrations with respect to preindustrial times.  $F^{\text{double}}$  is assumed to be  $3.71 \frac{\text{W}}{\text{m}^2}$  (Myhre et al., 1998). The four complete equations for the ocean mixed layers and the two land boxes are:

$$\begin{aligned} & \rho \cdot C \cdot h \frac{d\Delta T_0^{\text{NO}}}{dt} - \Delta Q + \lambda_{\text{O}} \cdot \Delta T_0^{\text{NO}} + \Delta F^{\text{NO}} \\ &= \frac{k_{\text{LO}}}{f_{\text{NO}}} (\Delta T_{\text{NL}} - \Delta T_0^{\text{NO}}) + \frac{k_{\text{NS}}}{f_{\text{NO}}} (\Delta T_0^{\text{SO}} - \Delta T_0^{\text{NO}}) \end{aligned} \quad (15)$$

$$\begin{aligned} & \rho \cdot C \cdot h \frac{d\Delta T_0^{\text{SO}}}{dt} - \Delta Q + \lambda_{\text{O}} \cdot \Delta T_0^{\text{SO}} + \Delta F^{\text{SO}} \\ &= \frac{k_{\text{LO}}}{f_{\text{SO}}} (\Delta T_{\text{SL}} - \Delta T_0^{\text{SO}}) + \frac{k_{\text{NS}}}{f_{\text{SO}}} (\Delta T_0^{\text{NO}} - \Delta T_0^{\text{SO}}) \end{aligned} \quad (16)$$

$$C_1 \frac{d\Delta T_{\text{NL}}}{dt} - \Delta Q + \lambda_{\text{L}} \cdot \Delta T_{\text{NL}} = \frac{k_{\text{LO}}}{f_{\text{NL}}} (\Delta T_0^{\text{NO}} - \Delta T_{\text{NL}}) \quad (17)$$

$$C_1 \frac{d\Delta T_{\text{SL}}}{dt} - \Delta Q + \lambda_{\text{L}} \cdot \Delta T_{\text{SL}} = \frac{k_{\text{LO}}}{f_{\text{SL}}} (\Delta T_0^{\text{SO}} - \Delta T_{\text{SL}}) \quad (18)$$

For a description of the parameters see Table 1.  $\Delta F^{\text{NO}}$  is the heat flux into the northern ocean (and analogously  $\Delta F^{\text{SO}}$  the heat flux into the southern ocean). The function  $\Delta F^{\text{NO}}$  (and analogously  $\Delta F^{\text{SO}}$ ) can be formulated as:

$$\Delta F^{\text{NO}} = \frac{\rho \cdot C \cdot K (\Delta T_0^{\text{NO}} - \Delta T_1^{\text{NO}})}{0.5d} - \rho \cdot C \cdot w (\Delta T_1^{\text{NO}} - \pi \cdot \Delta T_0^{\text{NO}}) \quad (19)$$

See again Table 1 for explanations on the parameters.  $K$  is the vertical ocean diffusivity and the function  $w$  denotes the ocean surface temperature-dependent upwelling rate given by

$$w = w_0 \left( 1 - \frac{\Delta T_0}{\Delta T^+} \right) \quad (20)$$

where  $w_0$  (the initial upwelling rate) and  $\Delta T^+$  are parameters (see Table 1).

The absorbed heat is transported within each ocean box by diffusion and upwelling:

$$\frac{d\Delta T_i^{\text{NO}}}{dt} = \frac{K}{d^2} (\Delta T_{i-1}^{\text{NO}} - \Delta T_i^{\text{NO}}) - \frac{K}{d^2} (\Delta T_i^{\text{NO}} - \Delta T_{i+1}^{\text{NO}}) + \frac{w}{d} (\Delta T_{i+1}^{\text{NO}} - \Delta T_i^{\text{NO}}) \quad (21)$$

for  $i = 1, \dots, 39$ . The equation for  $\Delta T_{39}^{\text{NO}}$  contains an additional term that represents downwelling from the ocean mixed layer to the bottom layer.

The outputs of the climate model are the surface temperature and the heat uptake down to 700 meters ocean depth which are calculated according to

$$\Delta T^{\text{surface}} = \frac{1}{2} \cdot (f^{\text{NO}} \Delta T_0^{\text{NO}} + f^{\text{SO}} \Delta T_0^{\text{SO}} + f^{\text{NL}} \Delta T_{\text{NL}} + f^{\text{SL}} \Delta T_{\text{SL}}) \quad (22)$$

and

$$\Delta H^{700\text{m}} = \text{vol}^d \cdot \rho \cdot C \left( \sum_{i=1}^7 \frac{1}{2} \cdot (\Delta T_i^{\text{NO}} + \Delta T_i^{\text{SO}}) \right) \quad (23)$$

Parameter	Description	Unit	Value
$f^{\text{NO}}$	ocean fraction in northern hemisphere		0.58
$f^{\text{SO}}$	ocean fraction in southern hemisphere		0.79
$f^{\text{NL}}$	land fraction in northern hemisphere		0.42
$f^{\text{SL}}$	land fraction in southern hemisphere		0.21
$\rho$	density of sea water	$\text{kg} \cdot \text{m}^{-3}$	1029
$C$	specific heat capacity of water	$\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$	3900
$\text{vol}^d$	ocean volume of height $d$	$\text{m}^3$	$3.368 \cdot 10^{16}$
$w_0$	initial upwelling rate	$\text{m} \cdot \text{a}^{-1}$	5
$\Delta T^+$	upwelling parameter	K	12
$\pi$	temperature change of downwelling region relative to global mean		0.2
$R$	ratio of equilibrium temperature change over land versus ocean		1.4
$d$	height of the ocean layers	m	100
$k_{\text{LO}}$	land-ocean exchange coefficient	$\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$	0.5
$k_{\text{NS}}$	north-south exchange coefficient	$\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$	0.5
$h$	depth of oceanic mixed layer	m	estimated
$S$	climate sensitivity	K	estimated
$K$	vertical ocean diffusivity	$\text{m}^2 \cdot \text{a}^{-1}$	estimated
$C_{\text{ml}}$	heat capacity of mixed layer	$\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$	$h \cdot \rho \cdot C$
$C_1$	heat capacity of land masses	$\text{J} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$	0
$\lambda_{\text{O}}$	feedback parameter above ocean	$\text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$	calculated from $S$ and $R$
$\lambda_{\text{L}}$	feedback parameter above land	$\text{W} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$	calculated from $S$ and $R$
$\Delta T_{\text{ini}}^{\text{surface}}$	initial value for surface temperature	K	estimated
$\Delta H_{\text{ini}}^{00\text{m}}$	initial value for ocean heat uptake	$10^{22} \text{J}$	estimated

Table 1: Parameters with descriptions and values. For a detailed discussion of parameter values see Raper et al. (2001).

### 5.2. The likelihood function

The likelihood function is obtained from assumption (2) which describes the deviations between the model outputs (22) and (23), and the corresponding observations:

$$\ell(y, \theta) = \exp\left(-\frac{1}{2}(y - x(\theta))^T R^{-1}(y - x(\theta))\right) \quad (24)$$

The observations that we use consist 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 meters depth (Levitus et al. (2005)) from the years 1955 to 2003. Both data sets are publicly available (see the URLs in the reference section).

The covariance matrix  $R$  describes the observational error as well as the climate variability that is not included in the dynamics of the climate model.  $R$  is thus the sum of the diagonal matrix  $D$  that contains the variances of the observational error and the covariance matrix that is made up of the estimated variance-covariance structure of climate variability.

We cannot estimate natural variability from data alone because of the difficulty in separating natural variability from the underlying trend, and because the data time series is relatively short. As is common practice in climate change attribution and detection studies (compare e.g. 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 and ENSO related variability that are not included in the simple climate model that we use.

The control run has a length of 341 years and was first detrended by a local polynomial fit. We then estimated the autocovariance in the HadCM3 control run time series of global annual mean surface temperature and annual mean change of world ocean heat content down to 700 meters. For the cross-correlation, we proceeded as follows: we first fitted autoregressive processes to both time series (the AIC criterion selected a process of order 3 for global annual mean temperature, and a process of order 19 for annual mean change of

world ocean heat content) and then estimated the cross-correlation of the residuals. From this, the cross-correlation of the original time series could be calculated. We found this cross-correlation to be insignificant. The cross-covariance coefficients in  $R$  between global average annual mean surface temperature and annual mean change of world ocean heat content were therefore set to zero.

We scaled the observational standard deviation of the ocean heat content data by a factor of 1.8. We believe this is justified because not all types of uncertainties are considered in the calculation of the observational error (Levitus et al. (2005)). The observations for heat content are sparse in some regions of the earth, especially in the Southern ocean, and there is considerable uncertainty in the choice of an interpolation scheme between data points (see Gregory et al. (2004) for a thorough discussion of this issue).

Moreover, based on the results of previous studies (Tomassini et al. (2006)), we scaled our estimate of the autocovariance in the climate variability for ocean heat content by a factor of 1.25. This was done to account for the fact that there is some variance in the control runs of HadCM3 for ocean heat content and that complex climate models tend to underestimate the climate variability of ocean heat content (see Collins et al. (2001); see also Gent and Danabasoglu (2004) for a detailed discussion of this point with respect to the Community Climate Model version 2).

For surface temperature data, no such scalings were used since the observations are more reliable (Jones et al. (2001); Jones and Moberg (2003)) and climate variability is believed to be well reproduced by HadCM3 (Collins et al. (2001)).

For a more detailed discussion of possible scalings of the likelihood function we refer the reader to Tomassini et al. (2006).

### 5.3. Radiative forcing, prior parameter distribution, and stochastic model term

A crucial input to our model is the radiative forcing  $Q$  in equation (14). It has to be reconstructed from the past (see Crowley (2000), Joos et al. (2001)) as a sum of forcings due to nine different components: Greenhouse gas forcing (the combined effect of  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{SF}_6$  and halocarbons), stratospheric  $\text{O}_3$  forcing, tropospheric  $\text{O}_3$  forcing, direct aerosol forcing, indirect aerosol forcing, organic and black carbon forcing, stratospheric  $\text{H}_2\text{O}$  forcing, volcanic forcing and solar forcing. Since there is also considerable uncertainty in these reconstructions, we introduced individual multiplicative scaling factors for these components as unknown, but time-constant parameters. The means of the marginal prior distributions for the forcing scale parameters were all set to 1.0 and the standard deviations were derived from the assumption that the uncertainties given by IPCC in the Third Assessment Report represent a range of plus or minus one standard deviation. A Gaussian prior distribution is assumed where the uncertainties are given in percent, and a lognormal distribution is used where the uncertainty is given as a factor. These are the same priors as used by Tomassini et al. (2006).

In addition, we estimated the five parameters of the climate model indicated in Table 1. We used the same uniform marginal prior for climate sensitivity as in Tomassini et al. (2006). The priors of the other four parameters were chosen to reflect subject-matter knowledge and uncertainty (see Table 2). All marginal priors are indicated by dashed lines in Figure 2. The probability density of the joint prior distribution was assumed to be equal to the product of the marginal densities.

Because there is considerable uncertainty in the reconstructed forcing, we added a com-

Parameter	Unit	Prior distribution
$S$	K	Uniform(1,10)
$h$	m	Normal(90, $10^2$ )
$K$	$\text{m}^2\text{a}^{-1}$	Uniform(100,10000)
$\Delta T_{\text{ini}}^{\text{surface}}$	K	Normal(-0.35, $0.25^2$ )
$\Delta H_{\text{ini}}^{700\text{m}}$	$10^{22}\text{J}$	Normal(-7.5, $2.5^2$ )

Table 2: Estimated parameters and their marginal prior distributions.

ponent to the forcing,  $dQ$ , that is stochastic and time-varying:

$$Q(t) = Q_{\text{recon}}(t) + dQ(t) \quad (25)$$

Here,  $Q_{\text{recon}}$  is the forcing reconstructed from the past as described above, and the additional forcing  $dQ_t$  is assumed to follow the Ornstein-Uhlenbeck process for time-dependent parameters described in section 2.1 with  $\overline{dQ} = \text{E}[dQ(t)] = 0$  and with the hyperparameters  $\sigma$  and  $\tau$ .

#### 5.4. Choice of hyperparameters

For the cross-validation procedure, the hyperparameters  $\sigma$  and  $\tau$  of the Ornstein-Uhlenbeck process were varied in the ranges  $(0.2\text{Wm}^{-2}, 1.5\text{Wm}^{-2})$  and  $(10\text{a}, 25\text{a})$ , respectively. These ranges were chosen according to the following considerations: If the correlation length  $\tau$  is allowed to be too small, then the procedure will map observation noise and fluctuations of the climate due to its chaotic nature into the time-dependent forcing. Additionally, the covariance matrix  $R$  for the ocean heat uptake fitted to a control run from the HadCM3 model has a correlation length of approximately 15a. Since it is difficult to predict what the relation between the fluctuations in the forcing and the fluctuations in temperature and heat uptake would be, we choose a comparable correlation length. If the stochastic forcing is to represent only the uncertainty in the reconstruction of the forcing, then the values at the upper end of the interval of the standard deviation  $\sigma$  are presumably not plausible. However, since the stochastic forcing can also represent other model deficits, we allowed for these higher values.

#### 5.5. Results and discussion

We concentrate our presentation on the following two analyses: (i) no time-varying random forcing ( $dQ(t) = 0$ ) and forcing scale parameters estimated as described above, (ii) time-varying random forcing  $dQ(t)$  with Ornstein-Uhlenbeck process and forcing scale parameters fixed at the maximum values of the posterior density from analysis (i). The reason for this was to avoid potential identifiability problems between the forcing scale parameters and the estimated time dependent forcing.

The MCMC sample size was 30'000 unless mentioned otherwise.

##### 5.5.1. Posteriors without time-dependent random forcing.

Figure 2 shows the marginals of prior and posterior parameter distributions without time-dependent random forcing ( $dQ(t) = 0$ ). The results show that we cannot learn much relative to the priors about the forcing parameters with the exception of indirect aerosol forcing,

$\sigma$ ( $\text{Wm}^{-2}$ )	$\tau$ (a)	$\widehat{psl}$
0.2	10	5.18
0.2	18	5.32
0.2	25	5.48
0.5	10	6.35
0.5	18	6.80
0.5	25	6.12
1.0	10	5.98
1.0	18	6.83
1.0	25	6.74
1.5	10	3.44
1.5	18	5.01
1.5	25	6.15

Table 3: Estimated cross-validation criterion  $\widehat{psl}$  according to equation (13) for different combinations of hyperparameters.

volcanic forcing and solar forcing. These three forcing factors all tend towards values smaller than unity. This agrees qualitatively with the results of a similar analysis with the more complex Bern2.5D climate model. The posterior for climate sensitivity has a similar shape, but is somewhat narrower than was the case for the Bern2.5D climate model (Tomassini et al. (2006)).

The results of the cross-validation procedure for the hyperparameters of the Ornstein-Uhlenbeck process are given in Table 3. The combinations  $(\sigma = 0.5 \text{ Wm}^{-2}, \tau = 18\text{a})$ ,  $(\sigma = 1.0 \text{ Wm}^{-2}, \tau = 18\text{a})$  and  $(\sigma = 1.0 \text{ Wm}^{-2}, \tau = 25\text{a})$  led to the best results. To check the variability of the estimated pseudo-likelihood  $\widehat{psl}$ , we repeated the computations with  $(\sigma = 1.0 \text{ Wm}^{-2}, \tau = 18\text{a})$  twice (with the same sample size) and obtained the values 6.50 and 6.40. Thus, the cross-validation procedure excludes the values at the boundary of the range considered, but cannot distinguish between several possible values of the hyperparameters. We focus here on the results for  $\sigma = 1.0 \text{ Wm}^{-2}$  and  $\tau = 18\text{a}$ .

### 5.5.2. Time-dependent random forcing.

Figure 3a shows the reconstructed forcing  $Q_{\text{recon}}(t)$  as used for the smoothing simulations. Figure 3b depicts the 90% uncertainty band that arises from the prior uncertainty of the scaling factors. The posterior quantiles of the additional term  $dQ(t)$  for  $\sigma = 1.0 \text{ Wm}^{-2}$ ,  $\tau = 18\text{a}$  are shown in Figure 3c, and the posterior median of  $dQ(t)$  for three values of  $\sigma$ , namely  $\sigma = 0.2, 0.35$  and  $1.0$ , in Figure 3d. The figure shows that with increasing value of  $\sigma$ , the time-dependent forcing has a higher flexibility for local adaptations. However, besides such local adaptations, higher values of  $\sigma$  also allow for a trend to higher forcing over the last 40 years.

Figures 4 and 5 show the model output and the observations for the mean surface temperature and the ocean heat uptake, respectively.  $\tau$  is fixed at 18a and  $\sigma = 0 \text{ Wm}^{-2}$ ,  $0.2 \text{ Wm}^{-2}$ ,  $1.0 \text{ Wm}^{-2}$ . Without stochastic forcing, the 90% confidence-bands are too narrow and the model shows deficiencies especially in the mean surface temperature of the 1940s and 1950s and in the ocean heat uptake at the end of the twentieth century. This is so despite the inclusion of unknown forcing scale parameters. If a stochastic forcing is added, the 90% confidence-bands are more realistic and the model deficits mentioned above are corrected. For the surface temperature, in particular, the model residuals (defined as the observations minus the posterior median predictions) are now much closer to being independent and identically distributed. It should also be recalled that the covariance matrix  $R$ , estimated

from the variability of the HadCM3 control run, includes an AR process of order 3 which cannot be depicted in the model predictions shown in figure 4.

The residuals for the ocean heat uptake still contain a clear structure which cannot be removed by a correction of the forcing. However, this structure is also partially taken into account by our covariance matrix  $R$ , which, for the ocean heat uptake, includes an AR process of order 19 (see section 5.2).

It should be noted that the decadal variability of the observed ocean heat uptake 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.

### 5.5.3. *Posterior distribution of the constant parameters with time-dependent random forcing.*

Figures 6 and 7 show the posterior densities of the two most interesting parameters, the climate sensitivity  $S$  and the vertical ocean diffusivity  $K$  for various combinations of the hyperparameters  $\sigma$  and  $\tau$ .

The most striking feature is the strong dependence of the posterior of climate sensitivity on  $\sigma$ . As  $\sigma$  increases, the posterior gets narrower and shifts to values that are smaller than what more complex climate models predict. In contrast, the value of  $\tau$  seems much less crucial. At first thought, this is counter-intuitive: When there is more uncertainty about forcing, we would expect the uncertainty about the climate sensitivity to increase instead of decrease. This apparent contradiction is discussed in more detail below.

In contrast to this, the behaviour of the ocean diffusivity  $K$  is as expected. The inclusion of the time dependent forcing leads to a wider posterior distribution of  $K$  and probably a more realistic uncertainty estimate (Raper et al. (2001)). The posterior distribution of  $K$  does not seem to be affected by the choice of  $\sigma$ , whereas the width decreases somewhat with decreasing value of  $\tau$ .

### 5.5.4. *Interpretation*

To understand the negative dependence between the prior standard deviation  $\sigma$  of  $dQ(t)$  and the climate sensitivity  $S$ , it is useful to look at the distribution of the terms  $1/p(y_i | y_{-i}, x^{(k)})$  whose averages over  $k$  are the basis of the cross-validation criterion formulated in equation (13). High values of this quantity indicate poor model fit to the data.

Figure 8 shows the scatter plots of these quantities for  $y_i$  corresponding to surface temperature data between 1940 and 2003. The correlation length  $\tau$  is in all cases equal to 18a. The standard deviation  $\sigma$  is equal to  $1.0\text{Wm}^{-2}$  in the left column and equal to  $0.2\text{Wm}^{-2}$  in the right column. In the first row  $S$  is included in the estimation, whereas in the second row  $S$  was fixed at the value of  $3^\circ\text{C}$  which is rather implausible a posteriori, especially for  $\sigma = 1.0\text{Wm}^{-2}$ . In addition, the reconstructed forcing is shown. One can see that in all cases the model has difficulties 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 deficit is stronger for large values of  $\sigma$  and also when we fix  $S$  at  $3^\circ\text{C}$  (notice the different scales in the two rows). Hence it seems that for large values of climate sensitivity the model is unable to predict the temperature observations at the times of volcanic eruptions. Modifying the

forcing randomly does not improve the situations: with a larger prior standard deviation of  $dQ(t)$ , the fit gets even worse.

Figure 8 seems to contradict our result from Table 3 that  $\sigma = 1\text{Wm}^{-2}$ ,  $\tau = 18\text{a}$  give a higher pseudo-likelihood than  $\sigma = 0.2\text{Wm}^{-2}$ ,  $\tau = 18\text{a}$ . However, the hyperparameters  $\sigma = 1\text{Wm}^{-2}$ ,  $\tau = 18\text{a}$  perform better for the prediction of the ocean heat uptake.

Additional insight in the differences between the analyses with and without stochastic forcing can be gained by looking at the estimated random forcing term  $dQ(t)$  itself. Figure 3 shows that with increasing values of  $\sigma$ , the estimated  $dQ(t)$  increases at the end of the observation period. This increase in forcing allows the same temperature increase to be achieved with a lower climate sensitivity.

To conclude, the reduced upper tail of the posterior of climate sensitivity associated with higher  $\sigma$  is caused by the balance of the following two situations: (i) The observed temperature increase at the end of the simulation period supports larger values of climate sensitivity. (ii) However, when climate sensitivity is large, the model reacts quickly to fast changes in forcing (e.g. during volcanic eruptions) but the measured temperature data only vary smoothly. Therefore, there is a tendency of the inference procedure to reduce climate sensitivity.

Under these circumstances, when our time-dependent random forcing can be used to produce a significant trend in forcing at the end of the measurement period, the algorithm can lead to misleading results. Instead of explaining temperature increase with a reasonable value of climate sensitivity, the algorithm keeps climate sensitivity low and “misuses” the freedom of random forcing to increase the trend in forcing, thereby increasing temperature.

The estimated small values of climate sensitivity in the case of large  $\sigma$  should therefore not be interpreted as suggesting that the climate sensitivity of the Earth is actually around  $1^\circ\text{C}$ . Rather, the results point to a structural deficiency of the simple climate model and the fact that energy balance climate models react differently to abrupt large volcanic forcings than more complex models (Yokohata et al (2005)).

### 5.5.5. Other analyses

One might conjecture that some of the discrepancies between the cases without and with stochastic forcing are due to the fact that the forcing scales were in one case estimated and in the other held fixed. To check this, we repeated part of the analyses by estimating both the forcing scales and the additional stochastic forcing  $dQ(t)$ . However, this did not lead to substantial changes.

To further evaluate our technique, we also conducted an analysis with  $dQ_t = 0$  fixed and with a time-dependent vertical ocean diffusivity  $K(t)$  instead. We chose  $\bar{K} = \text{E}[K(t)] = 1800\text{m}^2\text{a}^{-1}$ ,  $\sigma = 500\text{Wm}^{-2}$ ,  $\tau = 18\text{a}$ . In this case, we did not include the stochastic forcing term  $dQ(t)$  but did estimate the forcing scaling factors. The results were not much different from the case in which  $K$  was assumed to be constant. We conclude that the parameter  $K$  does not have a strong enough effect on the model output to correct the model deficiencies by making it time-varying.

### 5.5.6. Discussion

We can summarize the findings of the previous subsections as follows. The concept of using time-dependent parameters to identify model deficits was partially successful in this case study. When no time-dependent forcing is included in the estimation, then the dynamic

behavior of the ocean heat uptake, in particular the increase at the end of the series, and the reaction of the temperature to abrupt changes in the forcing in case of volcanic eruptions are not correctly represented by the model. Additionally, the trend in the temperature and ocean heat uptake data in the last 40 years of the twentieth century forces the model to admit large (probably realistic) values of climate sensitivity to account for this trend.

While the use of time-dependent ocean diffusivity did not significantly improve the simulation, the addition of a stochastic time-dependent forcing term did. This was most clearly visible for the ocean heat uptake at the end of the 20th century, and for the temperature at periods where the forcing changes suddenly. However, the last improvement is achieved indirectly, by decreasing climate sensitivity to unrealistically small values and explaining the steady temperature increase over the past 40 years by an increased forcing.

## 6. Conclusions and outlook

We present a Markov chain Monte Carlo algorithm with some special adaptations as a technique to estimate continuous-time stochastic parameters. The time-dependent parameter is assumed to follow a mean-reverting Ornstein-Uhlenbeck process. The main idea consists of splitting the time interval into subintervals which reduce the rejection rate in the Metropolis-Hastings algorithm and accelerates convergence of the Markov chain. The conditional Ornstein-Uhlenbeck process with fixed endpoints is used as 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, but this generally leads to slow convergence and identifiability problems. The model tends to overfit the data and it is typically not possible to constrain the hyperparameters by means of the data. As a consequence the estimates for the hyperparameters may become unphysical.

The algorithm was tested with a simple climate model. An additional stochastic forcing is estimated using the described algorithm. The results are compared with the case when the forcing uncertainty is taken into account by constant parameters which consist of scaling factors of the reconstructed forcing.

The results show that the proposed smoothing algorithm converges well and is able to estimate the additional time-dependent forcing and its uncertainty effectively and adequately. 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 as well performed very satisfactorily. It is able to produce a rough constraint on the hyperparameters of the Ornstein-Uhlenbeck process and at the same time selects values for the hyperparameters which lead to an improved model in the sense that the residuals now better satisfy the statistical assumptions formulated in the likelihood function.

However, as our example shows, the technique must be applied carefully. Time dependent parameters have 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 have seen that it can create a trend for the time-dependent parameter that allows a constant parameter to better adapt to a different feature in the data, in our case to the sudden changes in forcing due to volcanic eruptions. By a careful analysis of the different behaviour with and without time-varying parameters, we can obtain insight into the weak points of the model.

When used thoughtfully, the techniques presented in this paper are expedient and powerful. They can effectively help to estimate time-dependent stochastic parameters, to detect

model deficiencies if present, to help improve the structure of the model, and to estimate uncertainties of stochastic influence factors of models.

There are many ways to refine our approach. We could use more complicated stochastic process models or different hyperparameter values for the Ornstein-Uhlenbeck process in different time periods. Moreover, there are other possibilities to deal with time-varying parameters, for instance maximum likelihood methods combined with some spline-type regularizations. We believe however, that the particular approach taken is not of crucial importance. It is much more important to try to identify model deficits, and time-varying parameters are a useful tool for this task.

**Acknowledgments.** We thank Reto Knutti for providing helpful comments on an earlier version of this paper.

## References

- Beck M.B., Uncertainty, System Identification and the Prediction of Water Quality, in: M.B. Beck and G. van Straten (eds), *Uncertainty and Forecasting of Water Quality*, Springer, Berlin 1983.
- Brun R., *Learning from Data: Parameter Identification in the Context of Large Environmental Simulation Models*, ETH Diss. No. 14575, 2002.
- Buser Ch., *Differentialgleichungen mit zufällig zeitvariierenden Parametern*, ETH Diploma thesis, 2003.
- Collins M., S.F.B. Tett, and C. Cooper, 2001: The internal climate variability of HadCM3, a version of the Hadley Centre coupled model without flux adjustments, *Climate Dynamics*, **17**, 61-81.
- Crowley T.J., 2000: Causes of climate change over the past 1000 years, *Science*, **289**, 270-277.
- Elerian O., S. Chib, and N. Shephard, 2001: Likelihood inference for discretely observed nonlinear diffusions, *Econometrica*, **69**, 959-993.
- Gelfand A.E., and D.K. Dey, 1994: Bayesian model choice: asymptotics and exact solutions, *J. R. Statist. Soc. B*, **56**, 501-514.
- Gent P.R., and G. Danabasoglu, 2004: Heat uptake and the thermohaline circulation in the community climate system model, version 2, *Journal of Climate*, **17**, 4058-4069.
- Gregory J.M., H.T. Banks, P.A. Stott, J.A. Lowe, and M.D. Palmer, 2004: Simulated and observed decadal variability in ocean heat content, *Geophys. Res. Lett.*, **31**, L15312.
- Jones P.D., and A. Moberg, 2003: Hemispheric and large-scale air temperature variations: An extensive revision and an update to 2001, *Journal of Climate*, **16**, 206-223. See also [http://www.met-office.gov.uk/research/hadleycentre/CR\\_data/Annual/land+sst.web.txt](http://www.met-office.gov.uk/research/hadleycentre/CR_data/Annual/land+sst.web.txt).
- Jones P.D., T.J. Osborne, K.R. Briffa, C.K. Folland, E.B. Horton, L.V. Alexander, D.E. Parker, and N.A. Rayner, 2001: Adjusting for sampling density in grid box land and ocean surface temperature time series, *J. Geophys. Res.*, **106**, 3371-3380.
- Joos F., I.C. Prentice, S. Sitch, R. Meyer, G. Hooss, G.-K. Plattner, S. Gerber, and K. Hasselmann, 2001: Global warming feedbacks on terrestrial carbon uptake under the IPCC emission scenarios, *Global Biogeochem. Cyc.*, **15**, 891-907.
- Kristensen N.R., Madsen H. and Jørgensen S. B., 2004: A method for systematic improvement of stochastic grey-box models, *Computers and Chem. Eng.*, **28**, 1431-1449.
- Kloeden P.E., and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin 1995.
- Kuczera G., D. Kavetski, S. Franks, and M. Thyer, 2006: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, in press.

- Levitus S., J. Antonov, and T. Boyer, 2005: Warming of the world ocean, 1955-2003, *Geophys. Res. Lett.*, **32**, L02604. See also [http://www.nodc.noaa.gov/DATA\\_ANALYSIS/temp/basin/hc1yr-wO-700m.dat](http://www.nodc.noaa.gov/DATA_ANALYSIS/temp/basin/hc1yr-wO-700m.dat).
- Myhre G., E.J. Highwood, K.P. Shine, and F. Stordal, 1998: New estimates of radiative forcing due to well mixed greenhouse gases, *Geophys. Res. Lett.*, **25**, 2715-2718.
- Raper S.C.B., T.M.L. Wigley, and R.A. Warrick, 1996: Global sea-level rise: past and future. In: Milliman J.D., Haq B.U. (eds), *Sea-level rise and coastal subsidence: causes, consequences and strategies*, Kluwer Academic, Dordrecht 1996.
- Raper S.C.B., J.M. Gregory, and T.J. Osborn, 2001: Use of an upwelling-diffusion energy balance climate model to simulate and diagnose A/OGCM results, *Climate Dynamics*, **17**, 601-613.
- Stott P.A., S.F.B. Tett, G.S. Jones, M.R. Allen, W.J. Ingram, and J.F.B. Mitchell, 2001: Attribution of twentieth century temperature change to natural and anthropogenic causes, *Climate Dynamics*, **17**, 1-21.
- Tomassini L., P. Reichert, R. Knutti, Th.F. Stocker, and M.E. Borsuk, 2006: Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods, *Journal of Climate*, accepted for publication.
- Vrugt J.A., C.G.H. Diks, H.V. Gupta, W. Bouten, and J.M. Verstraten, 2005: Improved Treatment of Uncertainty in Hydrological Modeling: Combining the Strength of Global Optimization and Data Assimilation, *Water Resources Research*, **41**, doi:10.1029/2004WR003414.
- Wigley T.M.L., S.C.B. Raper, 1987: Thermal expansion of sea water associated with global warming, *Nature*, **330**, 127-131.
- Wigley T.M.L., S.C.B. Raper, 1992: Implications for climate and sea level of revised IPCC emissions scenarios, *Nature*, **357**, 293-300.
- Yokohata T., S. Emori, T. Nozawa, T. Ogura, and M. Kimoto, 2005: Climate response to volcanic forcing: Validation of climate sensitivity of a coupled atmosphere-ocean general circulation model, *Geophys. Res. Lett.*, **32**.

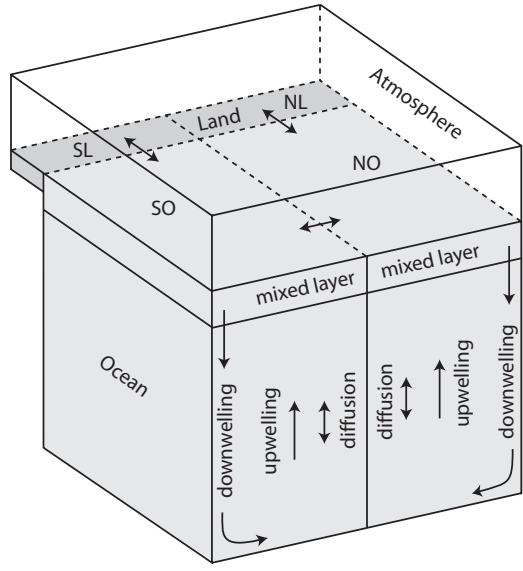


Fig. 1: Schematic figure of the simple climate model that is used to exemplify the smoothing algorithm.

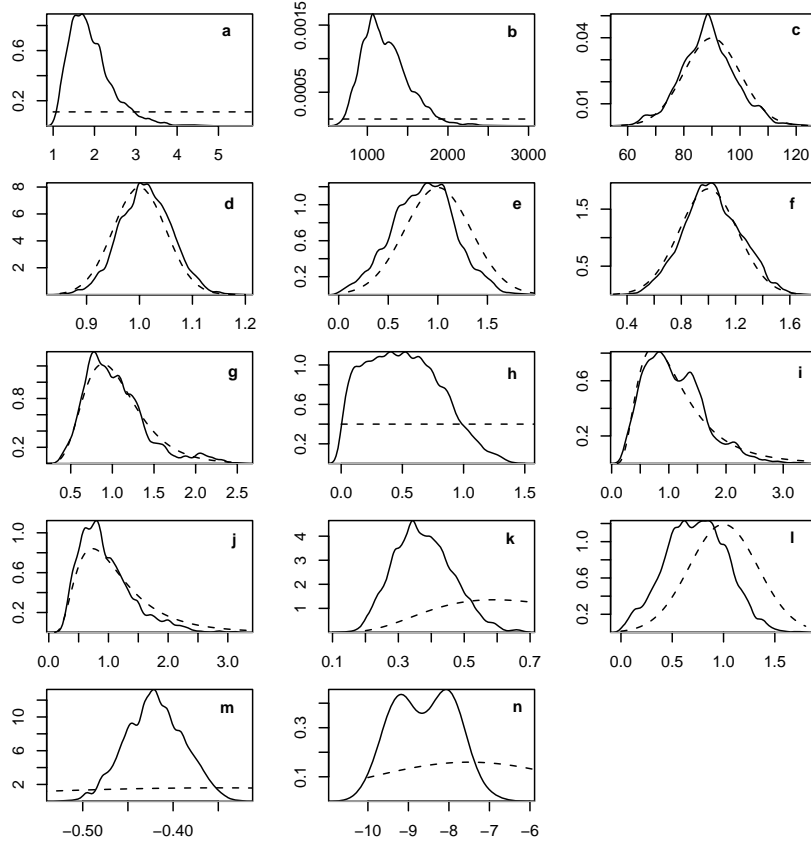


Fig. 2: Prior (dashed lines) and posterior (solid lines) distributions of constant parameters without inclusion of time dependent parameters ( $dQ_t = 0$ ). a. Climate sensitivity  $S$ ; b. Vertical ocean diffusivity  $K$ ; c. Mixed layer depth  $h$ ; d. Greenhouse gas forcing scale; e. Stratospheric  $O_3$  forcing scale; f. Tropospheric  $O_3$  forcing scale; g. Direct aerosol forcing scale; h. Indirect aerosol forcing scale; i. Organic and black carbon forcing scale; j. Stratospheric  $H_2O$  forcing scale; k. Volcanic forcing scale; l. Solar forcing scale; m. Initial condition for surface temperature  $\Delta T_{ini}^{surface}$ ; n. Initial condition for ocean heat uptake  $\Delta H_{ini}^{700m}$ . See Table 1 for units.

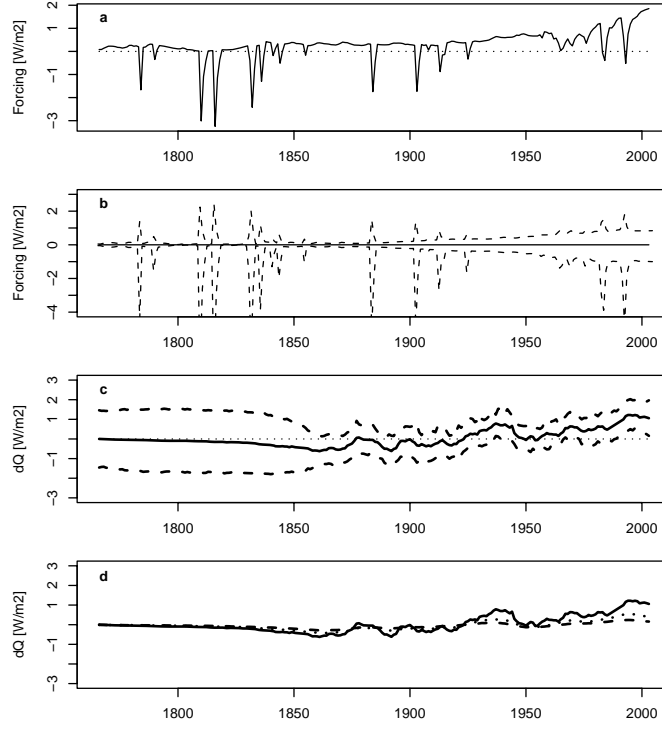


Fig. 3: a. Times series of the reconstructed forcing  $Q_{\text{recon}}(t)$ . b. 90% uncertainty band in forcing that arises from the prior uncertainty of the scaling factors. The median has been subtracted. c. The posterior 5%, 50% and 95% quantiles of the additional forcing term  $dQ(t)$  for  $\sigma = 1.0 \text{ Wm}^{-2}$ ,  $\tau = 18a$ . d. Posterior median of  $dQ(t)$  for three values of  $\sigma$ , namely  $\sigma = 0.2$  (dashed),  $0.35$  (dotted) and  $1.0$  (solid). In all cases,  $\tau = 18a$ .

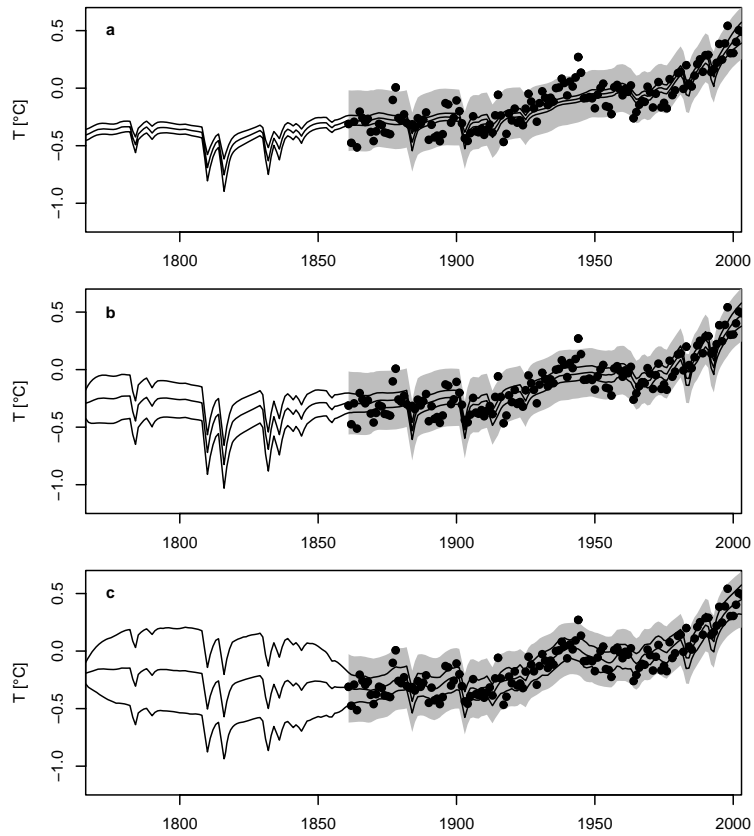


Fig. 4: Model output for global surface temperature. The solid lines show the 5%, 50%, and 95% posterior quantiles. The shaded area shows an estimate of the 90% predictive interval. a. Without stochastic forcing ( $dQ_t = 0$ ), but including the forcing scale parameters; b. Stochastic forcing included.  $\sigma = 0.2, \tau = 18$ ; c. Stochastic forcing included.  $\sigma = 1.0, \tau = 18$ .

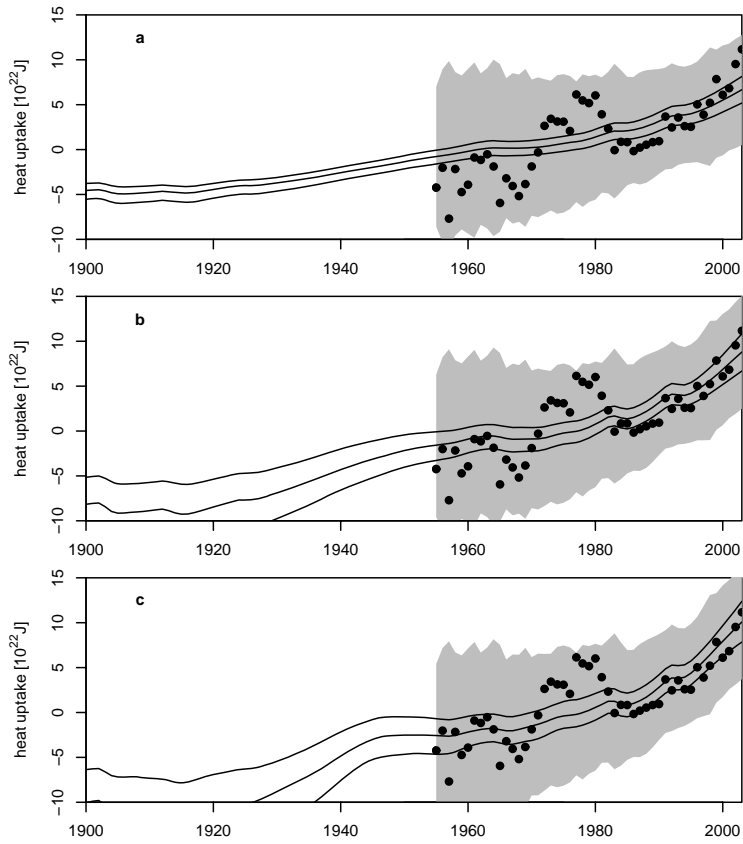


Fig. 5: Model output for ocean heat uptake down to 700 meters depth. The solid lines show the 5%, 50%, and 95% posterior quantiles. The shaded area shows an estimate of the 90% predictive interval. a. Without stochastic forcing ( $dQ_t = 0$ ), but including the forcing scale parameters. ; b. Stochastic forcing included.  $\sigma = 0.2, \tau = 18$ ; c. Stochastic forcing included.  $\sigma = 1.0, \tau = 18$ .

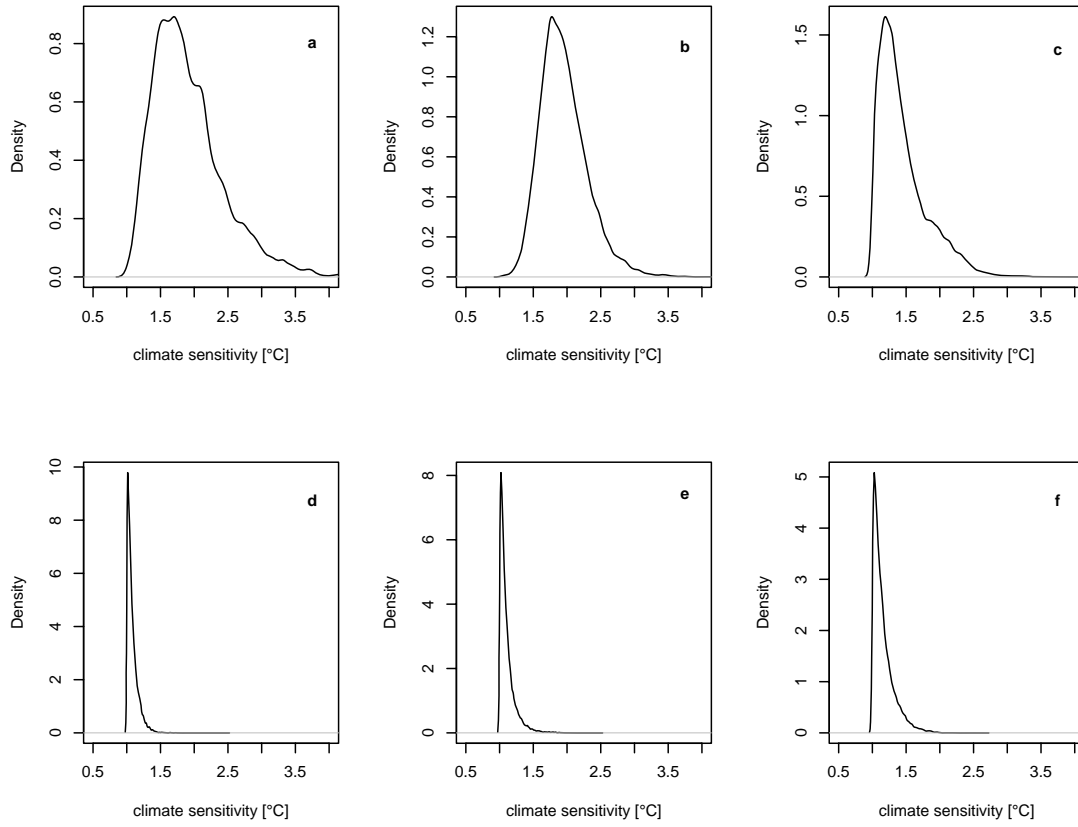


Fig. 6: Posterior distribution of climate sensitivity  $S$  with different values for the hyperparameters  $\tau$  and  $\sigma$ . a. No stochastic forcing ( $dQ_t = 0$ ), but with the forcing scale parameters included; b.  $\sigma = 0.2$ ,  $\tau = 18$ ; c.  $\sigma = 0.5$ ,  $\tau = 18$ ; d.  $\sigma = 1.0$ ,  $\tau = 10$ ; e.  $\sigma = 1.0$ ,  $\tau = 18$ ; f.  $\sigma = 1.0$ ,  $\tau = 25$ .

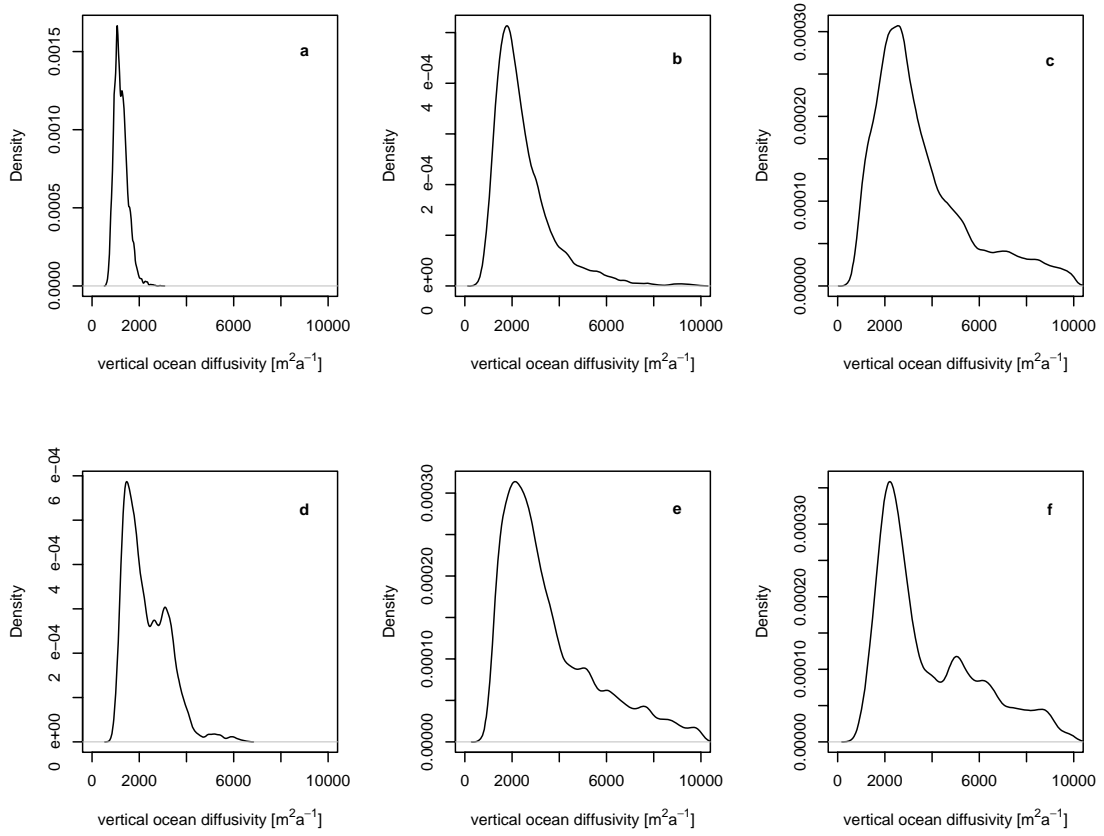


Fig. 7: Posterior distribution of vertical ocean diffusivity  $K$  with different values for the hyperparameters  $\tau$  and  $\sigma$ . a. No stochastic forcing ( $dQ_t = 0$ ), but with the forcing scale parameters included; b.  $\sigma = 0.2$ ,  $\tau = 18$ ; c.  $\sigma = 0.5$ ,  $\tau = 18$ ; d.  $\sigma = 1.0$ ,  $\tau = 10$ ; e.  $\sigma = 1.0$ ,  $\tau = 18$ ; f.  $\sigma = 1.0$ ,  $\tau = 25$ .

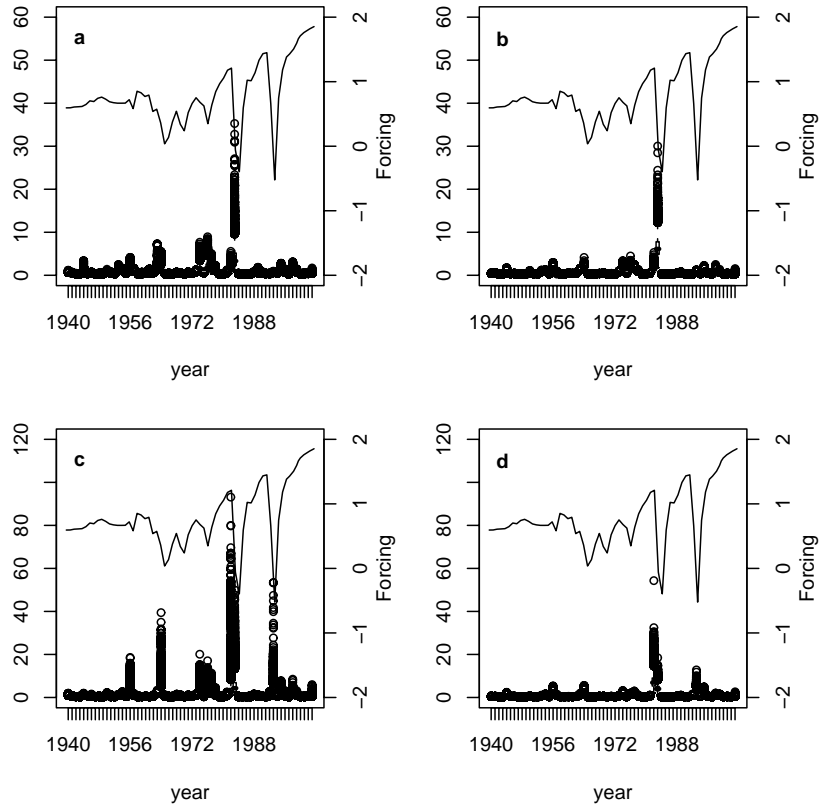


Fig. 8: Scatter plots of the quantities  $\frac{1}{p(y_i|y_{-i},x^{(k)})}$ ,  $k = 1, \dots, 10^4$  (for this figure only every third sample point was considered), for a.  $\sigma = 1.0, \tau = 18$ ; b.  $\sigma = 0.2, \tau = 18$ . The index  $i$  corresponds to the surface temperature data of the years 1940 to 2003. In the two upper panels climate sensitivity  $S$  is included in the estimation. The two lower panels show the same quantities, but with climate sensitivity fixed at  $S = 3.0$ . Again the hyperparameters are set at c.  $\sigma = 1.0, \tau = 18$ ; d.  $\sigma = 0.2, \tau = 18$ . In all four panels the forcing time series is also depicted as a solid line.