Taking uncertainty in ordinary differential equation models into account

Hans R. Künsch

Seminar für Statistik, ETH Zürich

Singapore, January 2008

Contents

Introduction

In environmental sciences, deterministic models (ODE's, PDE's) are predominant. Different sources of uncertainty are

- Measurement errors,
- Uncertain initial or boundary conditions,
- Unknown parameters and inputs,
- Model uncertainties (unresolved processes, discretization error)

Challenges:

- Estimate unknown quantities from data,
- Make predictions,
- Quantify uncertainties of estimates and predictions,
- Suggest ways to improve the models.

Notation

Let x be the state vector of an environmental system. Evolution given by the ODE

$$\frac{dx(t)}{dt} = g(x(t); \theta, \phi(t)); \quad x(0) = x_0.$$

θ denotes parameters, $\phi(.)$ inputs or boundary conditions.

$x(t)$ is usually not observed. Observations are given by

$$y_i = h(x(t_i)) + \varepsilon_i \quad (i = 1, 2, \ldots, n)$$

where ε_i is measurement error with density f_ε (often assumed to be normal).

Least squares estimation

If only parameters are unknown, standard procedure is weighted least squares (maximum likelihood for normal measurement error):

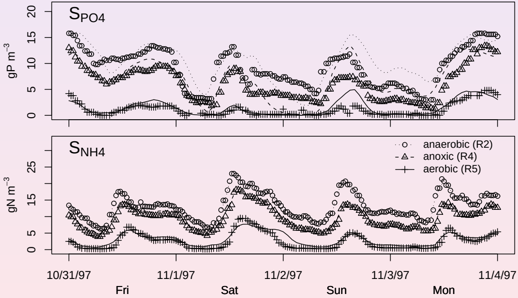
$$\widehat{\theta} = \arg\min \sum_i (y_i - h(x(t_i; \theta)))^2.$$

(can use weights to account for different precisions). Problems: Complicated shape of sum of squares as a function of θ (the log likelihood), leading to local minima and collinearity problems. Computation of the solution of the ODE as limiting factor. Similar remarks apply for Bayesian methods.

For non-chaotic systems, can add unknown initial conditions to the unknown parameters.

Model deficits

In many applications systematic errors often dominate random errors. This invalidates uncertainty estimates of unknown parameters.
Next slide: Comparison of observations and fitted model for a sewage plant The variables are phosphate and ammonia in 3 reactors.


S_{PO4}
gP m^{-3}
0
5
10
15
20
S_{NH4}
gN m^{-3}
0
5
15
25
anaerobic (R2)
anoxic (R4)
aerobic (R5)
10/31/97
11/1/97
11/2/97
11/3/97
11/4/97
Fri
Sat
Sun
Mon

Unknown inputs

If unknown input $\phi(.)$ is a function of time (or space), need to regularize maximum likelihood, e.g. by putting a prior on $\phi(.)$. Simplest priors are Gaussian (if necessary after some transformation).

A popular choice for the prior mean is the best available guess $\bar{\phi}(t)$ and for the prior covariance

$$\mathrm{Cov}(\phi(t), \phi(s)) = C_0 \exp(-\gamma|t-s|^{\alpha}).$$

One then needs an algorithm to sample from the posterior for $(\theta, \phi(.))$ and a way for choosing the hyperparameters C_0, γ and α.

Sampling from the posterior

Use Metropolis-Hastings. Updates of $\phi(.)$ are proposed by modifying the current value locally. For $\alpha = 1$, the prior is Markovian. Thus we can propose $\phi(.)$ on (u, v) from the prior given $\phi(u)$ and $\phi(v)$.

Drawbacks: One has to solve the ODE each time a new $\phi(.)$ is proposed. Would prefer to take the observations into account already in the proposal. Can we construct simple approximate solutions ?

Selecting hyperparameters

Putting a prior on the hyperparameters and updating them together with θ and $\phi(.)$ is a bad idea. The priors for different values of some combinations of hyperparameters are mutually singular. This invalidates most MCMC algorithms.

In practice, one can consider $\phi(.)$ only on a fine grid. Then the priors are only close to singularity. MCMC converges, but very slowly.

In the example below, we determine the hyperparameters by Bayesian crossvalidation, i.e. by maximizing

$$\sum_i \log p(y_i | y_j; j \neq i).$$

The terms on the right can be estimated from the MCMC output.

Tracer transport in porous media

Tracer concentration C in a flow can be described by the conservation equation with inputs u and q:

$$\frac{\partial C}{\partial t} + \sum_i \frac{\partial (u_i C)}{\partial x_i} = q$$

q is a known source term. $u = (u_1, u_2, u_3)$ is the velocity of the flow. u depends on the unknown permeability and the pressure equation. A Gaussian prior is used for log permeability.

In fluid dynamics particle methods can approximate the corresponding pior distribution of C that are more efficient than Monte Carlo.

Predictive and filter distribution

Initial condition $x(0)$ uncertain, has distribution ν_0. If system is chaotic, uncertainty increases as time increases. Use data up to time t_i to reduce uncertainty about $x(t_i)$. In geophysical sciences, this is called data assimilation.

Notation:
Predictive distribution $\mu_i =$ law of $x(t_i)$ given $y_1, \ldots y_{i-1}$.
Filter distribution $\nu_i =$ law of $x(t_i)$ given $y_1, \ldots y_i$.
ν_i is used as initial condition for integrating from time t_i to t_{i+1}.

Basic recursions

In principle, we can compute predictive and filter distributions recursively in two steps.
Propagation according to the model

$$\mu_i(A) = \nu_{i-1}(\{x; R(t_i - t_{i-1}, x) \in A\})$$

where $R(t, x)$ is the solution of the ODE with initial condition x at time t.
Update by Bayes formula

$$\nu_i(dx) = \frac{\mu_i(dx) f_\varepsilon(y_i - h(x))}{\int \mu_i(dx) f_\varepsilon(y_i - h(x))}.$$

Implementation of recursions

Typically, dimension of x is huge (up to 10^9). Analytical or numerical computations rarely feasible.

Exception: Linear ODE, linear measurement h, Gaussian noise ε and Gaussian initial distribution ν_0, leads to the Kalman filter. Variational methods or Monte Carlo methods are used for an approximate solution. Monte Carlo methods to be discussed later.

Same structure if ODE is replaced by an SDE.

Accounting for the bias

Add an explicit model bias term to the observation equation

$$y_i = h(x(t_i)) + b(t_i) + \varepsilon_i,$$

where b has a Gaussian prior. (If $E[b(t)] = 0$ for all t, this means that we assume correlated errors.) Then estimate the posterior of $b(.)$ and θ by MCMC.

Allows to judge adequacy of the model for the data and to give more realistic uncertainty estimates for θ or for model predictions.
More in the lecture by David Higdon (presumably).

Time-varying parameters

Possible reason for systematic errors of deterministic models: Some parameters are not constant, but vary in time. Strategy to deal with this:

- Replace a selected component of θ by an unknown function of time and put a (Gaussian) prior on it.
- Sample the posterior of the time-varying and the constant components of θ.
- Look how posteriors of constant parameters change and how much the systematic errors are reduced.
- Analyze the posterior mean of the time-varying component for correlations with external or internal variables.

From ODE's to SDE's

Introduce a white noise term into the ODE as an additional forcing term:

$$\frac{dx}{dt} = g(x(t)) + \sigma(x(t))N(t).$$

But white noise in continuous time is pathological. Integrated white noise

$$B(t) = \int_0^t N(s)ds$$

(Brownian motion, Wiener process) is continuous, but not differentiable. The above equation can be treated rigorously as an integral equation. To keep this in mind, write

$$dx(t) = g(x(t))dt + \sigma(x(t))dB(t).$$

Posterior mode estimation of an SDE

If we consider an SDE model as a prior for $x(.)$, then the mode of the posterior (MAP) is the minimizer of

$$\frac{1}{\sigma_\varepsilon^2}\sum_i (y_i - h(x(t_i)))^2 + \int ||\sigma(x(t))^{-1}\left(\frac{dx(t)}{dt} - g(x(t),\theta)\right)||^2 dt$$

with respect to θ and $x(.)$.

Can approximate this numerically by choosing a set of basis functions (Ramsay).

Time varying inputs

Time-varying inputs: A simple global climate model

Joint work with L. Tomassini, P. Reichert, M. Borsuk and Ch. Buser.

Model of Wigley and Raper (1987, 1992) for deviations of surface and ocean temperature from equilibirum.

Essential features of the Wigley-Raper model

- Consists of 4 boxes, one each for land/ocean and northern/southern hemisphere.
- Assumes a tendency to return to equilibrium.
- Incoming radiation power absorbed over the ocean only.
- Heat exchanged between ocean surface and land in each hemisphere and between the two ocean surfaces.
- Heat transported in the ocean by diffusion and by a simplified thermohaline circulation.
- Observed variables are global mean surface temperature and heat uptake of the ocean down to 700 meters.

Illustration of the model

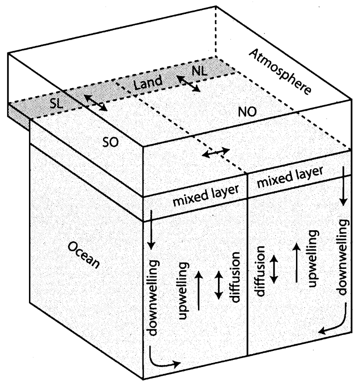


Fig. 1: Schematic figure of the simple climate model that is used to exemplify the smoothing algorithm.

Inputs and parameters

The model has radiative forcing F as input. This is the incoming power per unit area due to direct and indirect radiation. It is reconstructed from the past as a sum of 9 different components like greenhouse gas forcing, solar forcing, volcanic forcing etc.

The model has two main unknown parameters:
Climate sensitivity measures the effect of changes in radiative forcing on temperature. Defined as temperature increase in equilibrium due to doubling of the CO_2 content of the atmosphere.
Ocean diffusivity describes how fast heat is transported in the ocean due to diffusion.

Fitting the model with reconstructed input

We assume correlated observation errors. The covariance matrix of these errors is estimated from a control run of a complex climate model. It is assumed to be known in our analysis.

Estimating the unknown parameters gives reasonable posteriors, but the residuals show systematic patterns, and the posteriors for surface temperature and heat uptake are unrealistically narrow. We conjecture that uncertainty about climate sensitivity is too low because model deficits are not taken into account.

Posteriors for parameters

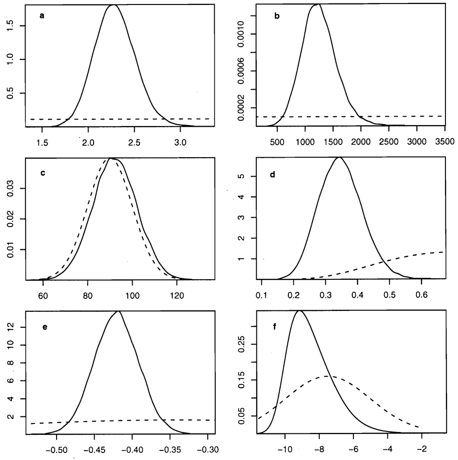


Fig. 2: Prior (dashed lines) and posterior (solid lines) distributions of constant parameters

Data and posterior for temperature

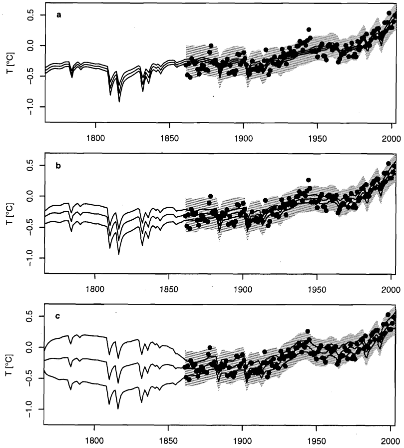
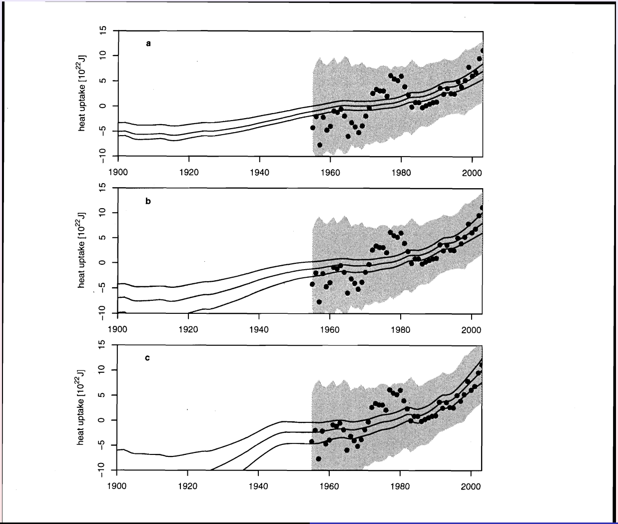


Fig. 4: Model output for global surface temperature. The solid lines show the 5%, 50%, and

Data and posterior for ocean heat uptake



Uncertainty of radiative forcing

The input F is uncertain since it is reconstructed from the past. Take this into account by adding an unknown disturbance $\Phi(.)$

$$F(t) = F_{\text{recon}}(t) + \Phi(t).$$

Prior for $\Phi(.)$: mean-reverting Ornstein-Uhlenbeck process (a continuous time AR(1)-model). MCMC is used to approximate the posterior distribution of $\Phi(.)$ and the unknown parameters.

Forcing: Reconstruction and posterior for disturbance

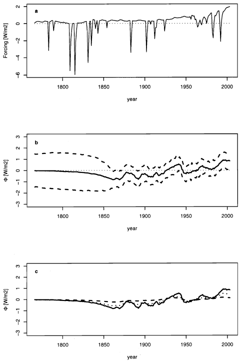


Fig. 3: a. Times series of the reconstructed forcing $\Delta F_{\text{recon}}(t)$. b. The posterior 5%, 50% and 95% quantiles of the additional forcing term $\Phi(t)$ for $\sigma = 1.0\text{Wm}^{-2}$, $\tau = 18\text{a}$. c. Posterior median of $\Phi(t)$ for three values of σ, namely $\sigma = 0.2$ (dashed), 0.5 (dotted) and 1.0 (solid). In all cases, $\tau = 18a$.

Posterior for diffusivity with different hyperparameters

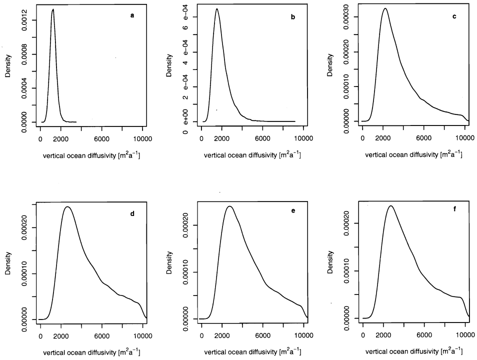


Fig. 7: Posterior distribution of vertical ocean diffusivity K with different values for the hyperparameters τ and σ. a. No stochastic forcing ($\Phi(.) = 0$); b. $\sigma = 0.2$, $\tau = 18$; c. $\sigma = 0.5$, $\tau = 18$; d. $\sigma = 1.0$, $\tau = 10$; e. $\sigma = 1.0$, $\tau = 18$; f. $\sigma =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

Posterior for climate sensitivity with different hyperparameters

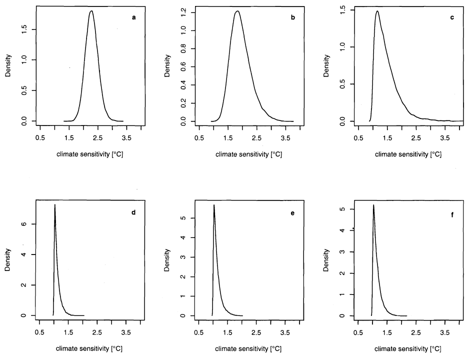


Fig. 6: Posterior distribution of climate sensitivity S with different values for the hyperparameters τ and σ. a. No stochastic forcing ($\Phi(t) = 0$); b. $\sigma = 0.2$, $\tau = 18$; c. $\sigma = 0.5$, $\tau = 18$; d. $\sigma = 1.0$, $\tau = 10$; e. $\sigma = 1.0$, $\tau = 18$; f. $\sigma = 1.0$, $\tau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

Effects of making forcing uncertain

Posterior for diffusivity becomes more spread out, but for climate sensitivity it is shifted to the left and less spread out.

Why ?

Small sensitivity is gives better fit for short-time changes in the forcing, e.g. volcanic eruptions. The temperature increase at the end can also be explained with larger forcing and smaller sensitivity.

Effects of making forcing uncertain

Posterior for diffusivity becomes more spread out, but for climate sensitivity it is shifted to the left and less spread out.

Why ?

Small sensitivity is gives better fit for short-time changes in the forcing, e.g. volcanic eruptions. The temperature increase at the end can also be explained with larger forcing and smaller sensitivity.