Tellus (1997), 49B, 123–135
Printed in UK – all rights reserved

TELLUS
ISSN 0280–6495

Bayesian inversion for reconciling uncertainties in global mass balances

By MILIND KANDLIKAR, *Department of Engineering and Public Policy, 129 Baker Hall, Carnegie Mellon University, Pittsburgh, PA, 15213, USA*

(Manuscript received 22 November 1995; in final form 28 October 1996)

ABSTRACT

This paper presents a Bayesian method for reconciling uncertainties in individual sources and sinks in global mass balance models and applies it to global cycle of methane. Ranges of values derived from in-situ flux measurements are used to define prior probability distributions for the individual sources and sinks. Atmospheric concentrations of carbon isotopes (^{14}C, ^{13}C and ^{12}C) and ice core measurements (^{12}C) provide additional information regarding the sources and sinks. Bayes Monte Carlo simulation is used to derive a posterior range for sources that combines data from field measurements, atmospheric observations and ice core data. It is shown that careful interpretation and analysis of available data can result in better resolution of source uncertainties. Emissions of methane from rice paddies and wetlands may be smaller than assessed in the past.

1. Introduction

Atmospheric concentrations of methane have more than doubled since pre-industrial times, and continue to increase steadily. Agricultural activity and land use change have, to a large extent, contributed to this increase (Shearer and Khalil, 1993). Methane plays an important role in tropospheric chemistry and is involved in the series of chemical reactions that regulate concentrations of the OH radical, which is the primary oxidizing agent in the atmosphere (Thompson and Cicerone, 1986; Logan et al., 1981). Methane also contributes about 15% of current excess radiative forcing (IPCC, 1992). Hence, abatement of methane emissions is an important policy response for combating global climate change.

Because of the diffuse nature of the sources of methane, extrapolations from field measurements to a global value for each representative source type, such as wetlands or agricultural sources have wide uncertainty bands. Uncertainties in the methane budget can significantly affect model predictions of the concentrations of methane (Guthrie and Yarwood, 1991), as reflected in a recent debate on slowing down of the atmospheric increase (Dlugocenky et al., 1993). Additionally, some sources of methane, such as landfill emissions have low to negligible costs of abatement relative to carbon dioxide (EPA, 1990; Rubin et al., 1992). This has led some analysts to suggest that methane sources provide an attractive short term means for greenhouse gas emissions reduction (Hogan et al., 1991). Explicit characterization and reduction of the uncertainty is, therefore, important because it can affect future methane concentrations as well as policies for limiting greenhouse warming.

Two types of methods, "forward" modeling/synthesis (Tans et al., 1990; Fung et al., 1991) and "inverse" modeling/synthesis (Enting et al., 1995), have been proposed to get a better understanding of the source and sink inventories of greenhouse gases, including methane. In the forward modeling approach sources and sink combinations are prescribed as inputs. These input combinations are

used to model global and/or regional concentration profiles. Model outputs are qualitatively compared with measured concentration data and the source/sink combination that provides the closest match is considered the "best guess" scenario. In contrast, inverse modeling and synthesis approaches are more rigorous. Here the model is mathematically inverted to determine the source and sink terms that provide a best fit to observational data. These models typically use linear inversion methods (Enting and Newsam, 1990; Enting et al., 1995) and optimization approaches (Enting and Pearman, 1987; Kandlikar and McRae, 1995) to minimize the error between model predictions and observational measurements.

Bayes Monte Carlo simulation is a generalized approach to inverse modeling in the presence of uncertainty and provides a flexible formalism within which to perform synthesis inversions of atmospheric models. In this paper we describe the application of Bayes Monte Carlo simulation to atmospheric inverse problems and apply it to the global cycle of methane. The remainder of the paper is structured as follows. Bayes Monte Carlo approach (Section 2) and its application to the global methane cycle (Section 3) is described. The results of the analysis and implications of this work are presented (Section 4).

2. Bayesian approaches

Bayesian approaches have been used by a number of workers to perform inversions of global biogeochemical cycles, particularly the carbon cycle (Enting 1985; Enting and Pearman, 1987; Hartley and Prinnn, 1993; Hein and Heimann, 1994; Enting et al., 1995). In the Bayesian inversion approach the modeler specifies independent prior distributions (priors) for a number of model inputs and parameters. Priors are updated using observational data and posterior distributions for model inputs derived. There are two advantages to using Bayesian approaches. First, specifying prior distributions for sources and model parameters is helpful in reducing the level of ill-conditioning in inversion (Enting and Pearman, 1987). Second, Bayesian methods can be used to determine the extent to which observational data helps in the reduction of prior uncertainties. The following discussion serves to clarify these points.

Consider the linear form given by:

$$Ax = y + e, \tag{1}$$

where y is a vector of observations, x is a vector of properties that can be observed only indirectly (source/sink terms), the matrix A represents a predictive model. Individual elements of x and y are denoted x_j and y_j, respectively. When x is known, the solution to eq. (1), is called the "forward" problem. When y is known, the solution to (1) is called an inverse problem. In many cases, inverse problems can be ill-posed (i.e., the matrix A is singular or nearly singular) or underdetermined (i.e., the number of variables in x exceeds the number of variables in y). The solution to the inverse problem is given by $\hat{x} = \mathrm{Inv}(A)y$, where $\mathrm{Inv}(A)$ is an operator chosen to optimise some properties of $\hat{x}$. For example, the unbiased minimum variance estimate of $\hat{x}$ is given by:

$$\hat{x} = (A^{\mathrm{T}} W^{-1} A)^{-1} A^{\mathrm{T}} W^{-1} y, \tag{2}$$

where W is the covariance matrix of the observation errors ε (Gelb, 1988). If A is ill-conditioned, small errors in the elements of y can lead to very large errors in the inferred values of the elements of x. Typically, the effects of ill-conditioning of A can be reduced or avoided by incorporating additional information regarding x in the solution of (2) (Enting and Newsam, 1990). Accordingly, if a set of initial estimates $\hat{x}$ with covariance matrix V are defined for x, the minimum variance solution is given by:

$$\hat{x} = x_0 + VA^{\mathrm{T}}(AVA^{\mathrm{T}} + W)^{-1}(y - Ax_0). \tag{3}$$

The error covariance matrices of the estimators given in eqs. (2) and (3) are given by $(A^{\mathrm{T}} W^{-1} A)^{-1}$ and $(V^{-1} + A^{\mathrm{T}} W^{-1} A)$, respectively. Thus, the two approaches give inverse error covariances that differ by the inverse covariance of the prior. The above maximum likelihood approach (Gelb, 1988) is not strictly Bayesian because variables x_j ($x_j \in x$) and y_0 ($y_0 \in y$) are still assumed to be deterministic variables with associated error; the elements of x are not random variables.

Eqs. (2) and (3) can also be derived alternatively using statistical arguments based on a strictly Bayesian approach. The Bayesian view of uncertainty requires modelers to assign independent (priors) to model input and parameters based on current scientific knowledge. Observational data serves to refine previous knowledge regarding

model inputs by narrowing their posterior probability distributions. Prior information about $\boldsymbol{x}$, is incorporated by treating each model input x_j as a random variable with a prior probability density function $p(x_j)$. This results in a prior probability density function $p(y_0)$ for the model output y_0. A measurement y_m is used to update these priors. It is critical to recognize that the measurement results in updated posteriors for model inputs and outputs. From Bayes rule the updated posterior density for y_0 is:

$$p\langle y_0 \mid y_m\rangle = \frac{p\langle y_m \mid y_0\rangle p(y_0)}{\int p\langle y_m \mid y_0\rangle p(y_0)\,\mathrm{d}y_0}. \tag{4}$$

In eq. (4), the denominator is equal to $p(y_m)$; in practice, the only way to calculate $p(y_m)$ is by evaluating the integral in the denominator. The conditional density $p\langle y_m \mid y_0\rangle$, is also called the likelihood function $L\langle y_m \mid y_0\rangle$, because it represents the likelihood that a particular value y_m is observed if the known value is y_0. Similarly, the posterior density for model input x_j is given by:

$$p\langle x_j \mid y_m\rangle = \frac{p\langle y_m \mid x_j\rangle p(x_j)}{\int p\langle y_m \mid x_j\rangle p(x_j)\,\mathrm{d}x_j}. \tag{5}$$

In eq. (5) the posterior density $p\langle x_j \mid y_m\rangle$ incorporates prior information about x_j as well as observation y_m and its associated error structure. A number of different estimates $\hat{x}_j$ can now be derived from the posterior density $p\langle x_j \mid y_m\rangle$. For example, the mode of the posterior density provides the maximum likelihood estimate; alternatively the mean of x_j over its posterior density is the minimum variance estimate (Gelb, 1988). In addition, the posterior probability density $p\langle x_j \mid y_m\rangle$ provides a measure of the extent to which observational data can help reduce uncertainties in sources and sinks (given their prior distributions).

The previous discussion serves to illustrate the key features of the Bayesian approach. In Subsection 2.1, we describe a Bayesian method, Bayes Monte Carlo simulation (Patwardhan and Small, 1992) which operationalizes the application of Bayes rule to inverse problems. The Bayes Monte Carlo approach avoids the need to find an explicit solution of the (potentially ill-posed or underdetermined) inverse problem given in eq. (1). Instead the forward problem is repeatedly solved for many sets of values for the inputs. For example, if the system is underdetermined, posterior density $p\langle x_j \mid y_m\rangle$ can be narrower than the prior density $p(x_j)$ if the model equations constrain allowed values for x_j. Additionally if A is ill-conditioned but overdetermined the method finds a posterior density that explicitly reflects the underlying instability.

While there have been a number of inversion studies perfomed for the global carbon cycle there have been few inversion studies for atmospheric methane. Fung et al. (1991) performed a forward synthesis and determined a "best guess" scenario for global sources and sinks. However, they did not pay attention to source and sink uncertainties. Hein and Heimann (1994) determined source uncertainties in methane emissions by inverting atmospheric concentration data (^{12}C and ^{13}C) using a 3-D transport model for ^{12}C and ^{13}C methane. In addition to constraints for ^{12}C and ^{13}C methane, the present work incorporates constraints on the atmospheric methane derived from ice-core data, and constraints from ^{14}C mass balance. This work also explicitly accounts for uncertainties in the sinks of methane, this results in wider uncertainty bounds for methane sources. Moreover, large uncertainties in the methane budget suggest the need for performing analyses with multiple sets of prior distributions. Unlike other methods which require the use of normal distributions, we demonstrate that this approach is more flexible in the choice of prior source distributions and measurement errors. As a consequence, it is possible to derive multiple plausible budgets for sources of methane based on the different priors.

In previous work, we have used a chance constrained programming formalism to perform synthesis inversions of the global methane (Kandlikar and McRae, 1995). The chance constrained programming method attempts to find the set of normally distributed posteriors for the source terms that best fit the observational data, for a given set of normally distributed priors. Unfortunately, the requirement that model variables be normally distributed severely limits the range of allowed prior distribution and model forms; the approach cannot be used for many problems of interest. Specifically, like many other methods the approach requires the use of a linear model, normally distributed priors and a normal form for the likelihood function. As noted above, Bayes Monte Carlo simulation is flexible and can

be applied to inverse calculation using any model form, prior distributions and observational error structure.

2.1. *Bayes Monte Carlo simulation*

In a Monte Carlo simulation model parameters are randomly sampled according to their prior probability density functions. Probability density functions are approximated by discrete probability mass functions. A mass balance model is evaluated for each sample (i.e., each combination of source and sink terms) resulting in a simulated probability mass function for the model output (i.e., the simulated observation)*. Each simulated replication is then compared to a "real world" observation of the same quantity. The model output and the observation are reconciled using Bayes rule to determine a posterior mass function for the model output. Similarly, posterior probability mass functions can be determined for the model inputs. The algorithm proceeds as follows (Kandlikar, 1994; Patwardhan & Small, 1992):

(1) Define a prior density $p(x_j)$ on all model parameters ($x_j \in \boldsymbol{x}$)

(2) Sample from this distribution n times using a Monte Carlo scheme and generate sample inputs $x_{i,j}$ ($i = 1, n$) for each ($x_j \in \boldsymbol{x}$)

(3) Run the model for each set of input samples to determine the sampled output $y_{i,0}$ ($i = 1, n$) for model output of interest y_0 ($y_0 \in \boldsymbol{y}$)

(4) Evaluate the likelihood function $L\langle y_m \mid y_{i,0}\rangle$ for each sample $y_{i,0}$ of model output y_0 using the error structure of the observational data/measurement

(5) Reconcile the model output and observation using Bayes Rule; obtain posterior mass function for inputs and outputs.

An illustration of the Bayes Monte Carlo method is provided in Fig. 1. Using Monte Carlo simulation a representative sample $x_{i,j}$ ($i = 1, n$) for each input x_j is generated from the prior density $p(x_j)$. A probability mass function $p(x_{i,j}) = 1/n$ is associated with ith sample $x_{i,j}$ ($i = 1, n$) for each input x_j. The model is then run iteratively n times for each vector of sampled inputs. This results in n sample values, $y_{i,0}$ ($i = 1, n$) for model ouput y_0 each with probability mass function equal to $1/n$. If the error (ε) on y_m is normally distributed and mean zero and variance σ_{ym}, then the likelihood function $L\langle y_m \mid y_{i,0}\rangle$ is:

$$L\langle y_m \mid y_{i,0}\rangle = \frac{1}{\sqrt{2\pi}\sigma_{ym}} \exp - \frac{1}{2\sigma_{ym}^2}(y_m - y_{i,0})^2. \tag{6}$$

The posterior mass function for each sample $y_{i,0}$ is determined from Bayes rule as

$$P\langle y_{i,0} \mid y_m\rangle = \frac{p(y_{i,0}) * L\langle y_m \mid y_{i,0}\rangle}{\sum_{i=1}^{n} p(y_{i,0}) * L\langle y_m \mid y_{i,0}\rangle}. \tag{7}$$

Since $p(y_{i,0}) = 1/n$, eq. (6) readily reduces to:

$$P\langle y_{i,0} \mid y_m\rangle = \frac{L\langle y_m \mid y_{i,0}\rangle}{\sum_{j=1}^{M} L\langle y_m \mid y_{i,0}\rangle}. \tag{8}$$

It is critical to recognize that for each simulated replication, the same posterior probability mass function is associated with inputs *and* outputs. Hence, the posterior probability mass function for input samples $x_{i,j}$ is also given by:

$$P\langle x_{i,j} \mid y_m\rangle = \frac{L\langle y_m \mid y_{i,0}\rangle}{\sum_{j=1}^{M} L\langle y_m \mid y_{i,0}\rangle}. \tag{9}$$

The sample values $x_{i,j}$ and the associated posterior probability mass functions $P\langle x_{i,j} \mid y_m\rangle$ characterize the posterior density function for model input x_j.

2.2. *Selection of priors and likelihood function*

The selection of prior distributions for model parameters and inputs is typically based on independent values presented in the literature. If a best estimate and corresponding range for all or a subset of model parameters is available, they are used to inform the construction of prior distributions. Often such information may not be available. In such cases, the modeler's subjective distribution gleaned from a careful reading of the literature is used. In many cases, limited information might force the modeler to assume uniform "non-informative" prior distributions.

When uncertainties are poorly understood, it is important that the analysis be done for several different assumptions regarding prior distributions. If the posteriors are similar for a range of

* For an application of Monte Carlo sampling to a Carbon cycle model see Gardner and Trabalka (1984).

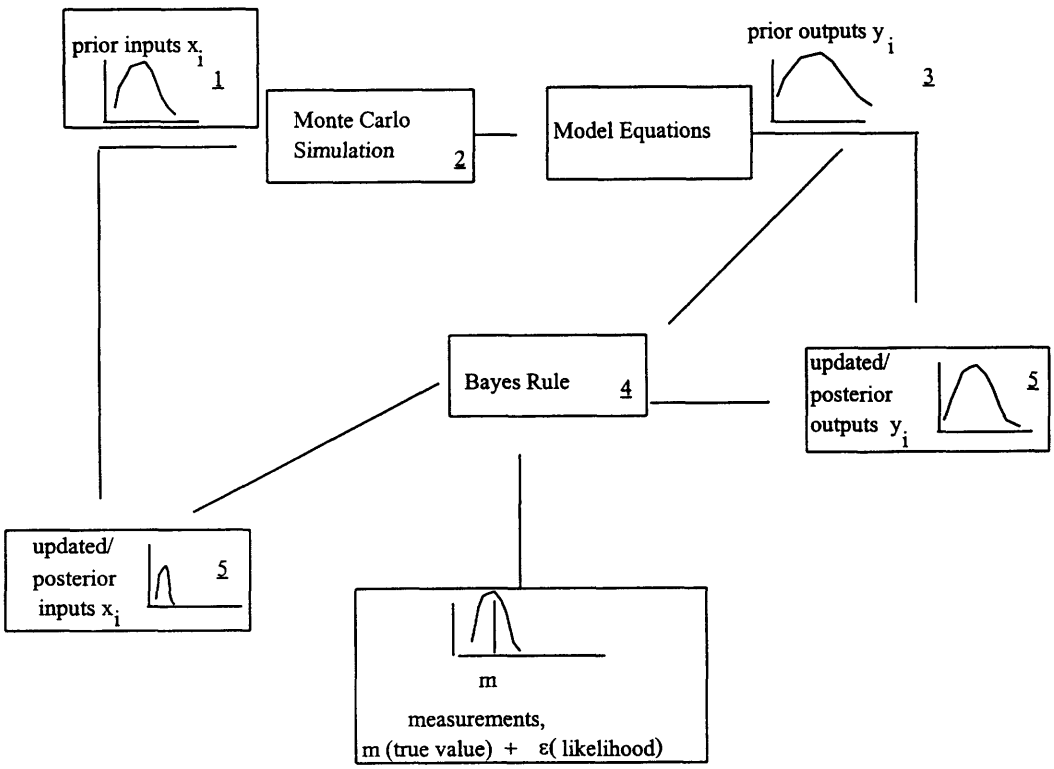

Fig. 1. A schematic for Bayes Monte Carlo Simulation. Underlined numbers correspond to steps in the algorith described in Section 2.

different assumptions on the priors then inferences regarding the distribution of inputs are more robust. The selection of the likelihood function $L\langle y_m | y_0 \rangle$ depends on the physical relationship between the model output y_0 and the measurement y_m. If the model output y_0 is a directly measured quantity then the likelihood function is determined using the error structure of the measurement process. However, depending on the level of aggregation in the model, the measured quantity may be a function of the model output of interest. For example, the mean annual increase in global atmospheric methane is not directly observed; in its place we use the mean annual increase in the global observational record of methane as the "observed" quantity. The mean annual increase in the global observational record of methane is derived from measurements of atmospheric methane made at several sites over the period of an year. The likelihood function for such aggregated quantities can be synthesized from the likelihood function for their components.

3. Application to methane cycle

In this section, we provide a brief description of the global methane cycle and illustrate the applicability of the Bayes Monte Carlo simulation to the model of mass balance. The equation

$$\frac{dC}{dt} = \sum_p \text{sources} - \sum_q \text{sinks}$$

represents the mass balance equation for one species in a single, well-mixed reservoir, with p sources and q sinks. If the model has more than one chemical species then $C(t)$ is a vector. A Monte Carlo simulation for the model output $C(t)$ can be performed for given initial conditions $C(0)$ and prior distributions for the source and sink

terms. Observations of the concentration $C(t)$ at time t can be used to update the priors and determine posteriors for the source and sink terms. Alternatively, if the change in concentration dC/dt is the observed quantity, then it is possible to directly update the source and sink terms using the "snap shot" mass balance model. In such a formulation source and sink variables for methane correspond to the model inputs x_j in Fig. 1. The observed change in annual concentration (dC/dt) corresponds to the model output y_0.

There are four constraints on the sources and sinks of atmospheric methane; one from each of the mass balance condition ^{12}C, ^{13}C and ^{14}C carbon isotopes and one from ice core data. ^{13}C and ^{14}C mass balances make it possible to discriminate between sources because different sources have differing isotopic ratios. Ice core data provides estimates for pre-industrial concentrations of methane. Hence, ice core data can be used to constrain non-anthropogenic sources. Note that the problem of determining the individual sources of methane is an underdetermined one. In this analysis, we used 12 sources categories that represent the sources of methane ($\boldsymbol{x}$), the four constraints were used to determine the model output ($\boldsymbol{y}$).

3.1. *Priors for the sources of methane*

In this paper, no attempt is made to discuss in detail the uncertainties in the individual sources of methane. A more detailed discussion of individual source uncertainties is available in earlier work by the author (Kandlikar and McRae, 1995). The reader may also refer to the extensive literature on the sources and global mass balance of methane (Fung et al., 1991; Cicerone and Oremland, 1988; Khalil et al., 1993). Priors for the individual sources of methane were constructed using the IPCC 1992 budget, with some modifications to reflect new findings. Two different sets of prior distributions are specified: (i) an informative prior that uses the specified range as well as a best estimate; (ii) a non-informative prior that ignores information regarding the best estimate and specifies the source to be uniformly distributed in the range.

Table 1 shows the individual sources of methane, and the assumed prior distributions used in this analysis. In Table 2, the ^{13}C and ^{14}C isotopic ratios for the different source terms are provided.

3.2. *Model equations and likelihood functions*

If the dependence of OH on levels of the CH_4 in the atmosphere is ignored, a reasonable assumption for time scales of one year, the global mass balance for isotopes of CH_4 can be represented by a linear relationship. The mass balance for ^{12}C is written as:

$$\frac{dC}{dt} = \sum_i E_i - K^*OH^*C - S_1, \tag{10}$$

where E_i are individual source terms, K the rate constant for the reaction between CH_4 and OH, OH represents the atmospheric concentration of the OH radical, C is the amount of CH_4 present in the atmosphere, S_1 is the soil sink of CH_4. Differences in ^{13}C isotopic concentrations in different sources arise because anaerobic bacteria which produce CH_4 show a marked preference for ^{12}C over ^{13}C during CH_4 production (Stevens, 1987). The isotopic ratios of a sample, R, can be derived from reported δ values with respect to the Pee Dee Belemnite standard:

$$R = \frac{(^{13}C/^{12}C)_{\text{sample}}}{(^{13}C/^{12}C)^{\text{standard}}} = \frac{\delta^{13}C_{\text{sample}}}{1000} + 1.$$

The mass balance for ^{13}C is written as:

$$R_g \frac{dC}{dt} + C\frac{dR_g}{dt} = \sum_i R_i^* E_i - K^*K_j^*OH^*R_g^*C - K_s^*R_g^*S_1, \tag{11}$$

where R_g is the isotopic ratio for atmospheric CH_4, dC/dt is the annual increase in atmospheric CH_4, and dR_g/dt is the annual change in the ^{13}C isotopic ratio, henceforth referred to as R'_g. The kinetic ratios, K_j and K_s, are the ratio of reaction rates with the two isotopes (K_{13}/K_{12}) for the reaction of CH_4 with OH and soil, respectively.

Measurements of ^{14}C content of CH_4 constrain the fraction that is depleted in ^{14}C, a measure of the fossil sources. Natural gas, coal mining, geological and industrial emissions are the known sources of CH_4 depleted in ^{14}C (Lacroix, 1993). The isotopic ratio of a sample, Q, is defined with respect to an oxalic acid standard as

$$Q = \frac{(^{14}C/^{12C})_{\text{sample}}}{(^{14}C/^{12}C)_{\text{standard}}} \quad \text{in terms of}$$

$$pMC, \quad Q = \frac{pMC}{100}.$$

Table 1. *Prior distributions for sources of methane based on a modified version of the IPCC(1992) budget*

Sources (Tg/yr)	Set 1 Informative prior	Set 2 non-informative prior
wetlands	triangular (100, 115, 200)	uniform (100, 200)
rice	triangular (25, 60, 150)	uniform (25, 150)
landfills/waste management	triangular (20, 30, 70)	uniform (20, 70)
biomass	triangular (20, 40, 80)	uniform (20, 80)
fossil sources¶	triangular (70, 100, 120) + triangular (5, 10, 15)	uniform (75, 135)
oceans	triangular (5, 10, 20)	uniform (5, 10)
ruminants	triangular (65, 80, 100)	uniform (65, 100)
hydrates†	triangular (5, 10, 20)	uniform (5, 20)
freshwater	triangular (1, 5, 25)	uniform (1, 25)
termites	triangular (10, 20, 50)	uniform (10, 50)
animal waste‡	triangular (20, 25, 30)	uniform (20, 30)
domestic sewage‡	triangular (10, 25, 50)	uniform (10, 50)

¶The first term includes coal mining and natural gas emissions (IPCC, 1992), the second term is to account for fossil based industrial emissions.
†Including other geological emissions.
‡See Kandlikar and McRae (1995).

Table 2. *Values of ^{13}C and ^{14}C ratios for the different sources of methane.*

Sources (Tg/yr)	^{13}C ratio R	^{14}C ratio Q
wetlands	0.94	1.12
rice	0.937	1.16
landfills	0.947	1.15
biomass	0.975	1.37
fossil sources	0.96	0
hydrates	0.935	0
oceans	0.96	1.16
ruminants	0.94	1.2
termites	0.935	1.24
freshwater	0.96	1.12
animal waste	0.94	1.2
domestic sewage	0.947	1.15

The mass balance equation for $^{14}CH_4$ is:

$$Q_g^* \frac{dC}{dt} + C^* \frac{dQ_g}{dt} = \sum_i Q_i^* E_i + \frac{E_{nuc}}{A_{abs}} - K^* K_j^{2*} Q_g^* OH^* C - K_s^{2*} Q_g^* S_1. \quad (12)$$

Q_g is the isotopic ratio for atmospheric CH_4, dQ_g/dt is the annual change in the ^{14}C isotopic ratio, henceforth referred to as Q_g'. A_{abs} is the $(^{14}C/^{12}C)$ ratio for the oxalic acid standard, and E_{nuc} is the ^{14}C source from nuclear reactors.

Ice core measurements provide strong evidence that concentrations of CH_4 in a pre-industrial atmosphere were stable at around 650 ppbv primarily reflecting natural emissions. After accounting for changes in the global sinks from pre-industrial times Khalil and Rasmussen (1990), conclude that the fraction of anthropogenic CH_4 emissions should be between 40%–70% of total emissions, i.e., a 30–60% range for natural emissions s_{nat}. Hence:

$$f_1 \leqslant \sum_i s_{nat} \Big/ \sum_i s_i \leqslant f_2 \quad (13)$$

where f_1 and f_2 are the minimum and maximum values of the natural fraction of CH_4 emissions and equal to 0.3 and 0.6, respectively.

The mean value and likelihood function for the annual increase in atmospheric concentration of ^{12}C, ^{13}C and ^{14}C isotopes were constructed from available data using the left hand side of eqs. (11), (12), and (13). Globally averaged values of variables in the equations and the corresponding error structures given in Table 3 were used to determine a measured value and corresponding likelihood function for each isotope. For example, in the case of the ^{12}C isotope the measurement is the mean value of measured annual increase in atmospheric methane, C, (32 Tg/yr) and the likelihood function is normally distributed with mean zero and standard deviation of 4 Tg/yr as prescribed in

Table 3. *Prior distributions and values for model variables*

Quantity (units)	Definition	Ref.
OH	normal (8.1, 0.9) (10^5rad/cm^3)	Prinn et al. (1992)
C	normal (4860, 50) (Tg)	IPCC (1992)
dC/dt	normal (32, 4) (Tg/yr)	IPCC (1992)
K	$3.52 \cdot 10^{-15}$ (cm^3/rad/sec)	Vagjiani & Ravi Shankar (1992)
S_1	normal (30, 7.5)	IPCC (1992)
R_g	0.9646	Stevens (1993)
$R_{g'}$	$1.26\ 10^{-4}$ yr^{-1}	Stevens (1993)
Q_g	1.22	Quay et al. (1991)
$Q_{g'}$	1.4 PM/yr	Quay et al. (1991)
K_s	0.979	King et al. (1989)
K_j	0.993	Quay et al. (1991)
E_{nuc}	normal (8, 2.5) (10^{24} molecules)	Quay et al. (1991)
A_{abs}	$1.176 \cdot 10^{-12}$	Stuiver and Pollach (1977)

IPCC(1992). For ^{13}C and ^{14}C isotopes, the likelihood function was derived from the uncertainties in C, as well uncertainties in dR_g^2/dt^2.

More detailed discussion of the values of parameters in Table 3 can be found in Kandlikar and McRae (1995). The likelihood function for the ice core data is given by the fraction of natural emissions, assumed to be uniformly distributed between 0.3 and 0.6. The probability mass function for each sample i was determined from eq. (8), by replacing the likelihood function $L\langle y_m | y_{i,o}\rangle$ with a joint likelihood function

$$L_{\text{joint}} = \prod_{o=1}^{4} L\langle y_{m,o} | y_{i,o}\rangle, \tag{14}$$

where the subscript o ($o = 1, 2, 3, 4$) corresponds to the isotopes ^{12}C, ^{13}C and ^{14}C, and the ice core data. The posterior distributions for individual sources are determined by the probability mass functions and associated values for the variables $x_{i,j}$.

4. Results and discussion

The analysis was carried out using a stratified latin hypercube sampling approach with 1000 samples and posterior probability distributions were obtained for the sources of methane. Latin hypercube sampling (LHS) is a form of stratified sampling which yields more precise estimates of distribution functions than random Monte Carlo sampling (Iman et al., 1980). The posterior distributions can be used to calculate the range of values that capture the 10th and the 90th percentile confidence levels for each of the sources. The ranges given in Table 4 provide the upper and lower bounds for each source of methane. The ranges are provided for both informative and non-informative priors assumed for the source terms. We term the budget derived using informative priors is the best guess budget. Additionally, in Fig. 2 we show the prior and posterior (updated) cumulative distributions for four methane sources from wetland, rice paddies, fossil related emissions and termites using the informative priors. In Fig. 3 the same distributions are shown for non-informative priors. One general observation can be made from these results: the reduction in uncertainty is most significant when the prior distributions of sources have large uncertainties associated with them as in the case of rice paddies and wetlands. For other sources the reduction in uncertainty is only marginal, which implies that the observational data does little to further constrain the priors for source terms.

Table 4. *Posterior ranges for sources of methane, including a best guess budget; the number in brackets is the minimum variance Bayes estimate for each of the sources of methane*

Sources (Tg/yr)	Set 1 informative prior (best guess budget)	Set 2 non-informative prior
wetlands	110–155 (132)	110–180 (141)
rice	37–85 (57)	25–90 (52)
landfills/waste management	25–50 (37)	23–62 (42)
biomass	25–60 (30)	25–70 (32)
fossil sources	85–115 (104)	85–115 (100)
hydrates	7–15 (12)	6–17 (12)
oceans	8–15 (12)	7–18 (13)
ruminants	70–90 (80)	68–92 (85)
termites	16–40 (26)	14–45 (29)
fresh water	5–18 (11)	4–23 (13)
animal waste	22–27 (25)	21–29 (25)
domestic sewage	16–35 (27)	12–40 (26)

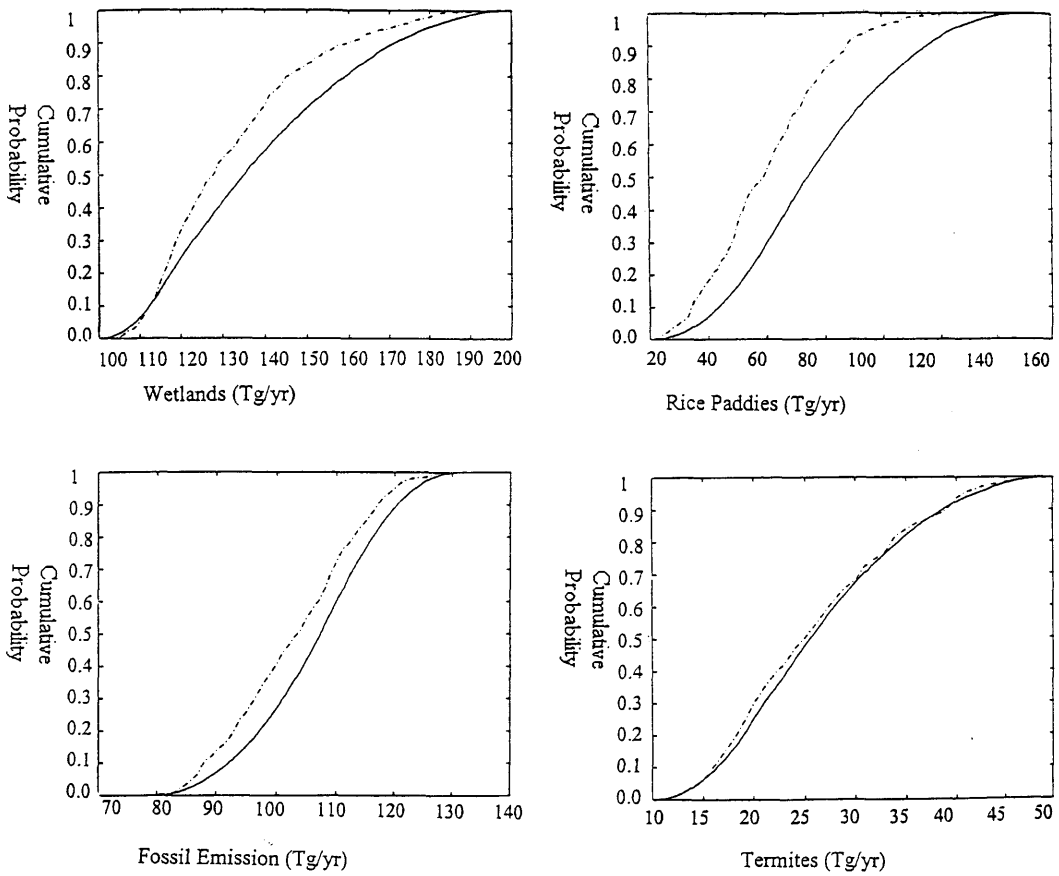

Fig. 2. Cumulative distribution functions for sources of methane: wetlands, rice paddies, fossil emissions, termites. Solid line is the informative prior (see Table 1) and the dashed line is the posterior. In each of the above graphs the x-axis gives the source term in Tg/Tr and the y-axis provides the probability that the source term is less than or equal to the correspondent value for the source.

From Tables 3, 4, and Figs. 2, 3 we observe that for rice paddies the posterior distribution shows a large reduction in uncertainty when compared to the prior distribution. This indicates that atmospheric observations and ice core data provide significant constraints on rice paddy emissions. Specifically, the upper and lower 90% confidence bounds for rice paddies were 37 and 85 Tg/yr respectively when informative priors are used. Using non informative priors, the resulting source range for rice paddies is equal to 25–90 Tg/yr. These results compare favorably with other estimates of rice paddy flux emissions. Bachelet and Neue (1993) have compared results of several studies with and without the inclusion of soil characteristics. They suggest that emissions from Asian rice paddies, which account for 90% of global emissions from rice were previously over estimated by 25% and cite a revised range of 57–82 Tg/yr. A more recent version of the IPCC (1994) report provides rice paddy source estimates of 20–100 Tg/yr based on these and other flux measurements. This is in fair agreement with the range from the non-informative priors case.

Wetlands are the largest single source of methane and constitute much of the natural fraction. From Table 4, the range for wetland sources is 110–155 Tg/yr with an informative prior and 110–180 Tg/yr using a non-informative prior. This is higher than extrapolations from point flux estim-

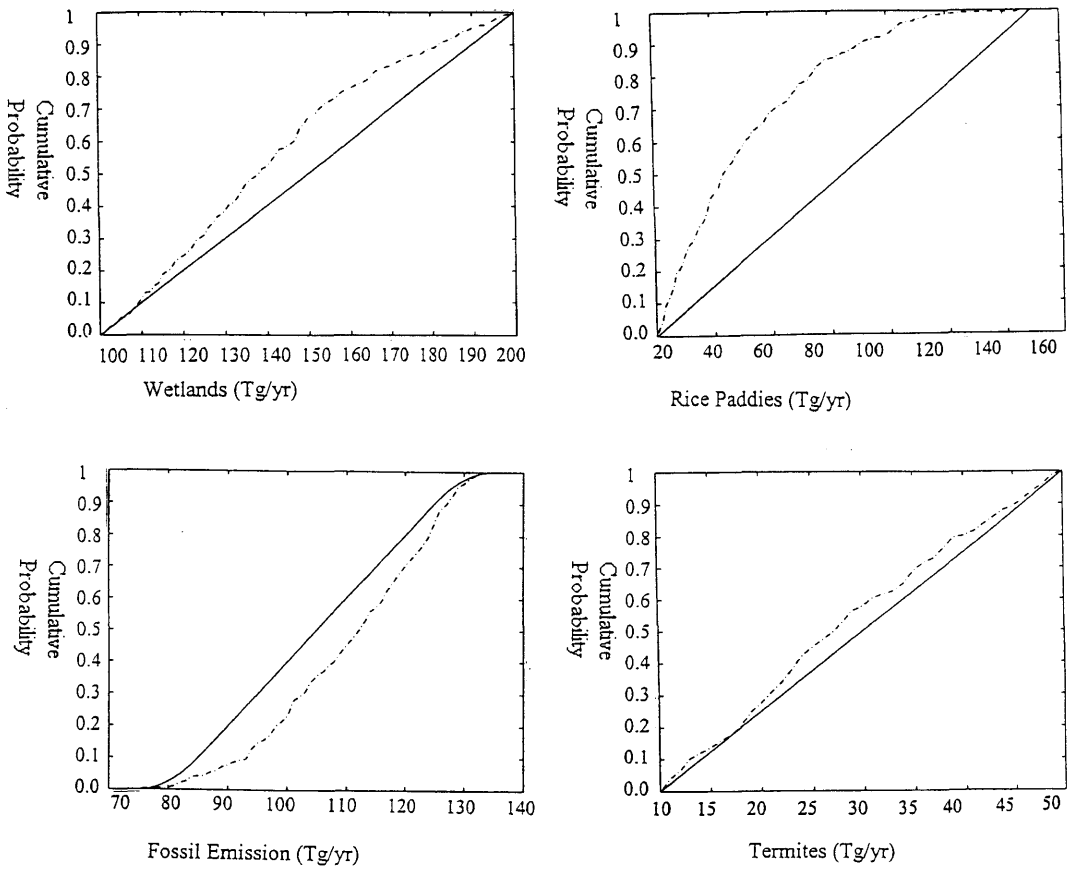

Fig. 3. Cumulative distribution functions for sources of methane: wetlands, rice paddies, fossil emissions, termites. Solid line is the non-informative prior (see Table 1) and the dashed line is the posterior. In each of the above graphs the *x*-axis gives the source term in Tg/yr and the *y*-axis provides the probability that the source term is less than or equal to the corresponding value for the source.

ates cited by Bartlett and Harris (1993), all in the range of 85–115 Tg/yr. When the analysis was repeated using a broader prior distribution for wetlands source (50–200 Tg/yr with a mode at 105 Tg/yr) with all other sources held at their informative priors, the posterior 10%–90% confidence bounds were given by 100–150 Tg/yr. This suggests that the range of wetland sources might be higher than flux based estimates compiled by Bartlett and Harris (1993) but lower than the subjective range of 100–200 Tg/yr provided in IPCC (1992).

For many of the other sources the updating procedure does not lead to posterior distributions that are significantly different from the priors. In such cases, the prior distributions are relatively consistent with the observations and the observations contain little extra information that could lead to a reduction in the source uncertainties. This is illustrated in Figs. 2, 3. The posterior distribution for fossil sources is marginally less uncertain compared to the prior; for the methane source from termites the two distributions are almost identical.

The budget(s) of methane derived in this study agree reasonably with our earlier work (Kandlikar and McRae, 1995) which uses chance constrained programming. In the chance constrained programming method the inversion attempts to find the set of normally distributed posteriors for the

source terms that best fit the observational data, for a given set of normally distributed set of priors. There is, however, one key difference between the two approaches. In the earlier work, the source term for wetland was 70–130 Tg/yr, which is systematically lower than the estimated range in this paper. Some of this difference results from differences in model assumptions and priors*. However, the difference persists even when the same set of prior distributions are used for the model inputs. The systematic difference results from the different requirements that the two methods place on the posterior distributions. The Bayes Monte Carlo approach does not place any restrictions on the mathematical form of the solution for posterior distributions; in the case of wetland source, posterior distributions show skewness and significantly deviate from the normal form. As a result, the budget presented in this work may be more reflective of the real uncertainties than the budgets derived in Kandlikar and McRae(1995) and Hein and Heimann(1994) who also use normally distributions for priors and posteriors.

An important element of the Bayesian approach is the need for careful selection of prior distributions especially when there are large uncertainties in the data. In such cases, it may be necessary to perform the analyses with multiple sets of priors. We illustrate this with the example below. The sources of methane derived from animal wastes and sewage treatment both poorly known quantities and as such have only recently been identified. The prior distributions assigned to them in Table 1, are based only a few studies and are highly uncertain. To test the dependence of the results of our analysis on the choice of these prior distributions, we perform a set of analyses based on an alternative assumption: that the prior distribution for landfills/waste management source category adequately captures emissions from animal wastes and sewage treatment, i.e., the source terms for these emissions categories are not included in the Bayes Monte Carlo analysis. The effect of this alternative assumption on the posterior source ranges for wetlands and rice paddies is observable. The 10% to 90% confidence range using informative priors for wetlands change from 110–155 Tg/yr to 105–155 Tg/yr and rice paddies emissions range shifts from 37–85 Tg/yr to 58–125 Tg/yr. Rice paddy emissions increase to account for the reduction in anthropogenic sources under alternate assumptions regarding emissions from animal wastes and sewage treatment.

As with other inversion approaches, an additional benefit of the Bayes Monte Carlo method is that it allows the modeler to ask questions about the value of gathering additional information, which is a central theme in Bayesian decision theory (Morgan and Henrion, 1990). For instance, one can determine the additional reduction in the uncertainty in posterior source distributions if uncertainties in observational data were to be reduced. A systematic and rigorous demonstration of value of information techniques using Bayes Monte Carlo simulation is outside the scope of this paper. We note, however, that the flexibility of the method allows modelers to carry out the analyses for non-linear models with skewed output distributions.

Below, we illustrate the concept using a simple example. Using best guess priors, we determine the posteriors for methane sources, assuming that the uncertainty of the OH sink is reduced by 50%, i.e., the range 475 ± 25 Tg/yr is assumed to capture the 1σ range of the OH source term. By comparing the posterior distributions of these source terms to the posterior distributions obtained with an OH source range of 475 ± 50 Tg/yr, reduction in source uncertainties from a 50% reduction in OH sink uncertainty can be calculated. The results of

Table 5. *The effect of reducing uncertainty in the global OH sink on source uncertainties*

Sources	Base case (Tg/yr)	Reduced uncertainty case (Tg/yr)
wetlands	110–155 (132)	100–140 (120)
rice	37–85 (57)	30–70 (50)
fossil emissions	85–115 (104)	88–115 (103)
termites	16–40 (26)	16–40 (26)

The base case numbers correspond to the "best guess" budget, i.e., they are the inferred source values with informative priors from Table 1 and an OH sink corresponding to 475 ± 50 Tg/yr. The reduced uncertainty case corresponds to the inferred source values with informative priors from Table 1 and an OH sink corresponding to 475 ± 25 Tg/yr.

* For example, in Kandlikar and McRae (1995) the soil sink was not considered.

the analysis are provided in Table 5 for 4 source types of methane. Some sources, particularly rice paddies show an observable reduction in the inferred uncertainty bounds.

Although this work demonstrates an application of the technique to a 0-D globally aggregated model, Bayes Monte Carlo simulation can be just as easily applied to models with greater spatial and temporal specificity. The method can be used to perform atmospheric "snap shot" atmospheric synthesis inversions using 2-D and 3-D transport models (Hein and Heimann, 1994; Fung et al., 1991, Enting et al., 1995), as well ocean modeling approaches that give longer term trends. The primary advantage of the approach results from its ability systematically synthesize multiple types of information, to reconcile all the different uncertainties and facilitate value of information calculation. Additionally, by allowing flexibility in the choice of the prior distributions it forces modelers to carefully provide statistical interpretation of all available information.

5. Acknowledgements

The author would like to acknowledge discussions with Drs. M. Small, G. Ramachandran, J. Kalagnanam. In addition, two anonymous reviewers provided detailed comments that significantly improved the paper. The responsibility of all errors, however, lies entirely with the author. The funding for this work was provided by core support from US National Science Foundation and US Department of Energy.

REFERENCES

Bachelet, D. and Neue, H. U. 1993. Methane emissions from wetland rice areas of Asia. *Chemosphere* **26**, 219–238.

Bartlett, K. and Harriss, R. 1993. Review and assessment of methane emissions from wetlands. *Chemosphere* **26**, 1–4, 261–320.

Cicerone, R. J. and Oremland, R. S. 1988. Biogeochemical aspects of atmospheric methane. *Global Biogeochemical Cycles* **2**, 229–327.

Dlugokencky, E. J., Lang, P. M., Masarie K. A. and Steele, L. P. 1994. Atmospheric CH_4 records from sites in the NOAA/CDML sampling network, pp. 247–350. In: Boden, T. A., Kaiser, D. P., Sepanski, R. J. and Stoss F. W. (eds.): *Trends '93. A compendium of data on global change.* ORNL/CDIAC-65. Carbon Dioxide Information Analysis Centre, Oak Ridge National Laboratory, Oak Ridge, Tenn., U.S.A.

Enting, I. 1985. Principles of constrained inversion in the calibration of carbon cycle models. *Tellus* **37B**, 7–27.

Enting, I. and Pearman, G. 1987. Description of a one-dimensional model of the carbon cycle calibrated using techniques of constrained inversion. *Tellus* **39B**, 459–46.

Enting, I. and Newsam, G. N. 1990. Atmospheric constituent inversion problems: Implications for baseline monitoring. *J. Atmos. Chem.* **11**, 69–87.

Enting, I., Trudinger C. M. and Francey, R. J. 1995. A synthesis inversion of the concentration and $\delta^{13}C$ of atmospheric CO_2. *Tellus* **47B**, 35–52.

EPA, 1990. Methane emissions and opportunities for control, *EPA/400/9–90/007*. September 1990.

Fung, I., John, J., Matthews, E., Prather, M., Steele, L. P. and Fraser, P. J. 1991. Three-dimensional model synthesis of global methane cycle. *J. Geophys. Res.* **96**, 13033–13065.

Gardner, R. H. and Trabalka, J. R. 1985. Methods of uncertainty analysis for a global carbon dioxide model. *DOE report TR024 — DOE/OR/21400–4*, US Department of Energy, Oak Ridge National Labs, Oak ridge, Tenn., USA.

Gelb, A. (ed.), 1988. *Applied optimal estimation.* MIT Press, Cambridge, MA.

Guthrie, P. and Yarwood, G. 1991. *Analysis of the IPCC future methane simulations, SYSA-91/114.* Systems Application International, San Rafael, CA.

Hartley, D. and Prinn, R. 1993. Feasibility of determining surface emissions of trace gases using an inverse method in a three-dimensional chemical transport model. *J. Geophys. Res.* **98**, 5183–5197.

Hein, R. and Heinmann, M. 1994. Determination of global scale emissions of atmospheric methane using an inverse modelling method. pp. 271–281. In: *Non-CO_2 greenhouse gases*, J. Van Ham et al. (eds). Kluwer Academic, The Netherlands.

Hogan, K., Hoffman, J. and Thompson, A. 1991. Methane on the greenhouse agenda. *Nature* **354**, 181–182.

IPCC, 1992. *Supplementary report to the Scientific Assessment.* Cambridge University Press, Cambridge, UK.

IPCC, 1994. *Climate Change 1994: radiative forcing of climate.* Cambridge University Press, Cambridge, UK.

Iman, R. L., Davenport, J. M. and Zeigler, D. K. 1980. *Latin hypercube sampling (program users guide).* Report number SAND79–1473, Sandia National Labs, Albuquerque, New Mexico, USA.

Kandlikar, M. 1994. *Reconciling uncertainties in science and policy models. Applications to global climate change.* Dissertation, Department of Engineering and Public Policy, Carnegie Mellon University.

Kandlikar, M. and McRae, G. J. 1995. Inversion of the

global methane cycle using chance constrained programming. *Chemosphere* **30**, 1151–1170.

Khalil, M. A. K. and Shearer, M. J. 1993. Sources of methane: an overview. In: *The atmospheric methane cycle: sources, sinks and rôle in global change* (ed. Khalil, M.A.K.). Springer Verlag, Berlin.

Khalil, M. A. K. and Rasmussen, R. A. 1993. Decreasing trend of methane: unpredictability of future concentrations. *Chemosphere* **26**, 803–814

Khalil, M. A. K. and Rasmussen, R. A. 1990. Constraints on methane global mass balance and an analysis of recent budgets. *Tellus* **42B**, 229–236.

King, S. L., Quay, P. D. and Landsdown, J. M. 1989. $^{13}C/^{12}C$ kinetic isotope effect for soil oxidation of methane at ambient concentrations. *J. Geophys. Res.* **94**, 18 373–18 277.

Lacroix, A. V. 1993. Unaccounted for sources of fossil and isotopically enriched methane and their contribution to the emissions inventory: a review and synthesis. *Chemosphere* **26**, 507–559.

Logan, J. A., Prather, M. J., Wofsy, S. C. and McElroy, M. B. 1981. Tropospheric chemistry. A global perspective. *J. Geophys. Res.* **86**, 7210–7254.

Morgan, M G. and Henrion, M. 1990. *Uncertainty. A guide to dealing with uncertainty in quantitative risk and policy analysis.* Cambridge University Press, Cambridge.

Patwardhan, A. and Small, M. Bayesian methods for model uncertainty analysis with applications to future sea level rise. *Risk Analysis* **124**, 513–523.

Prinn, R., Cunnold, D., Simmonds, P., Alyea, F., Boldi, R., Crawford, A., Fraser, P., Gutzler, D., Hartley, D., Rosen R. and Rasmussen, R. 1992. Global average concentration and trend for hydroxyl radicals deduced from ALE/GAGE trichloroethane data for 1978–1990. *J. Geophys. Res.* **97**, 2445–2461.

Quay, P. D., King, S. L., Stutsman, J., Wilbur, D. O., Steele, L. P., Fung, I., Gammon, R. H., Brown, T. A., Farrel, G. W., Grootes, P. M. and Schmidt, F. H. 1991. Carbon isotopic composition of atmospheric methane: Fossil and biomass burning source strengths. *Global Biogeochemical Cycles* **5**, 25–47.

Rubin, E. S, Cooper, R. N., Forsch, R. A., Lee, T. H., Marland, G., Rosenfeld, A. H. and Stine, D. D. 1992. Realistic mitigation options for global warming. *Science* **257**, 148–266.

Shearer, M. J. and Khalil, M. A. K. 1993. Rice agriculture: emissions. In: *The atmospheric methane cycle: sources, sinks and rôle in global change* (ed. Khalil, M. A. K.). Springer Verlag, Berlin.

Stevens, C. 1993. Isotopic abundances in the atmosphere and sources. In: *The atmospheric methane cycle: sources, sinks and rôle in global change* (ed. Khalil, M. A. K.). Springer Verlag, Berlin.

Stuiver, M. and Polach, H. A. Discussion. Reporting of ^{14}C Data. *Radiocarbon* **19**, 355–363.

Tans, P., Fung, I. and Takahashi, T. 1990. Observational constraints on the global atmospheric CO_2 budget. *Science* **247**, 1431–1438.

Thompson, A. M. and Cicerone, R. J. 1986. Possible perturbations to atmospheric CO, CH_4, and OH. *J. Geophys. Res.* **91**, 10 853–10 864.

Vaghjiani, G. L. and Ravishankara, A. R. 1991. New measurement of the rate coefficient for the reaction of OH with methane. *Nature* **350**, 406–409.