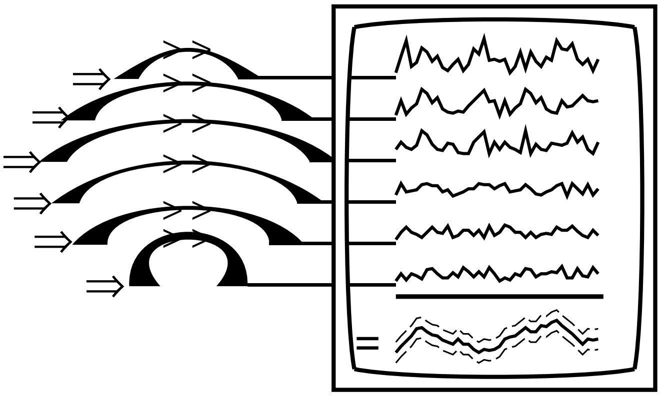
Inverse Problems in Atmospheric Constituent Transport: An Update

Version 0.6: October, 2006



I. G. Enting
MASCOS
139 Barry St
The University of Melbourne
Vic 3010, Australia

This is a preliminary draft for initial feedback — see distribution history. In particular, there are major gaps in my coverage of recent work. Often a * denotes a section, dot point or reference that is incomplete.

The proposed location is on http://www.unimelb.edu./au/∼enting/docs/update.pdf.

Contents

Distribution history

- Version 0.6 on website, 6 October 2006.
- Version 0.4 distributed at Carbon Data Assimilation workshop, MSRI, July 2006.

Preface

As I said in the preface to my book [22], extensive development of CO_2 inversions took place within the Cooperative Research Center for Southern Hemisphere Meteorology, which existed from 1993 to 2000. Indeed the 'lame-duck' period when we knew that funding for CRCSHM would not continue, provided me with a context in which to write the book.

At the end of CRCSHM, much of the CO_2 inversion work transferred to CSIRO Atmospheric Research (CAR). This involved an increased emphasis on regional-scale inversions, building on the work of the CSIRO Biospheric Working Group, and closely linking to the CSIRO observational programs. One reflection of the value of this linkage is that my most cited papers come from this work. Peter Rayner's extension of network design studies to the evaluation of potential satellite data, drawing on Denis O'Brien's expertise in radiative transfer took the 'experimental design' aspect of inversions from being purely theoretical studies into actual evaluation of proposed satellite missions. This linkage between observation and modelling was further enhanced by the development of the Lo-Flow analyser, for which Roger Francey and Paul Steele were awarded the Victoria Prize and Roger Francey was awarded a Federation Fellowship to undertake a research program exploiting these developments.

In the event, CSIRO chose not to go through with the Federation Fellowship activity, Roger retired, Denis went to Colorado State University, Peter went to LSCE, France, and I moved the the University of Melbourne. CSIRO Atmospheric Research merged to form CSIRO Marine and Atmospheric Research. (Note that, in his then role as Federal Minister for Science, the Honourable Peter McGaurin, appearing on national television, specifically denied that the cuts to this CSIRO carbon research were motivated by Australian government policy not to ratify the Kyoto Protocol). My objective — stated in the closing line of my book: sharing with my readers in further development — has been seriously curtailed.

A new section without a parallel in my book, is entitled 'So What?' and addresses the implications of CO_2 inversions. In the contemporary jargon of science management, it addresses *outcomes* rather than *outputs*.

Preparation of this update has been greatly facilitated by ongoing interaction with colleagues. Interaction with students and lecturers at the Carbon Data Assimilation workshop, at MSRI in July 2006 has expanded my views on these problems. My fellowship at MASCOS is supported in part by CSIRO. MASCOS (The Center of Excellence for Mathematics and Statistics of Complex Systems) is funded by the Australian Research Council.

Finally, the opinions expressed in this document do not represent the views, opinions or policies of the University of Melbourne.

1 Introduction

This document represents a partial update of my book *Inverse Problems in Atmospheric Constituent Transport* [22]. Although a paperback edition appeared in 2005, for production reasons this had to be an exact copy of the 2002 hardback edition.

What I want to do with this update document is:

- clarify some of the areas where I feel that I can now do better;
- explore some concepts in greater depth than was possible in the original writing;
- give an account of more recent work, mainly in the area of CO_2 inversions, since any attempt at significantly greater coverage would be beyond me;
- correct errors in my book that I (or others) have detected.

In other words, this is not a full update of my book as a whole, but rather an exploration of some of the themes that I find interesting.

Some aspects merely represent clarifications of my book. Others are reviews of recent work. However there is some new original work appearing here for the first time. This may create difficulty for people who wnat to cite it. Some of the new material is appearing in other places: my chapter in a book on complex systems science (Ed. D. Newth and I. Enting) and a chapter in abook on Turbulence and Air Pollution Modelling (Ed. D, Moriera and M Vilhena).

One important distinction in inverse problems can be illustrated by considering the behaviour of atmospheric CO_2 concentrations in response to a forcing from a flux $S(t)$:

$$C(t) = C(t=0) + \int_0^t R(t-t')S(t')\,dt' \tag{1.1}$$

This formalism was described by Oeschger and Heimann [76]. A change of variables, introducing $t'' = t - t'$ gives

$$C(t) = C(t=0) + \int_0^t S(t-t'')R(t'')\,dt'' \tag{1.2}$$

This shows that the two inverse problems (a) deduce $S(t)$ given $C(t)$ and $R(t)$ and (b) deduce $R(t)$ given $C(t)$ and $S(t)$ are formally equivalent. These problems, termed 'deconvolution' and 'calibration' respectively, are however quite different in their behaviour. The reason for this is that the statistical characterisation of the problems is very different. The calibration problem is looking for a smoothly-varying (and in this and many other contexts monotonically decreasing) function $R(t)$. The 'deconvolution' problem, so-called because (1) and (2) are convolutions of functions, can be trying to reproduce the full temporal variability of the forcing flux. The point is that the statistical characteristics of the inverse problem can, and usually will, play a dominant role in the form of solution. The two cases above: calibration and deconvolution serve as models for corresponding problems in much more complicated situations.

Table 1: Characteristics of the modelling spectrum with examples from carbon cycle studies and climate change.

Characteristics	Carbon Cycle	Climate System
Black box		
Empirical,	Curve fitting	Curve fitting
Stochastic	Airborne fraction	
Grey box		
Parametric,	Response function	Response function
Process-based	Box model	Energy balance model
White box		
Deterministic,		Atmos/ocean GCM
Reductionist,	Spatially resolved	
Mechanistic		Earth system model

2 Statistics

2.1 Mathematics and statistics

The underlying theme of my book is that inversion needs to be considered as a process of statistical inference. Although atmospheric transport models are formulated in terms of partial differentiation, and their implementation draws on sophisticated numerical mathematics, using the models for effective inversions requires techniques from statistics rather than applied mathematics.[1] The statistical approach to inversion underlies Tarantola's book [107] and is the basis of Kalnay's description of meteorological data assimilation [54]. A significant paper addressing this issue is by Evans and Stark [30] who integrate the applied mathematics and statistics approaches, relating concepts such as convergence etc. to statistical concepts such as consistency. In general, they advocate the use of non-parametric statistical techniques such as splines.

In discussing integrating applied mathematics and statistics, it is helpful to think in terms of a spectrum of models and modelling. A modelling spectrum was described by Karplus [57] as running from black-box models (characterised as being highly empirical) through to white-box models (characterised as being highly mechanistic), with black-box models generally being statistical and white-box models being generally deterministic. In between lies modelling represented by various shades of grey. Although Karplus envisaged different parts of the spectrum as being occupied by models from different fields of study, I argued that various forms of carbon cycle modelling spanned much of the spectrum [19]. The spectrum ran from curve fitting, through response functions to box models and on to models with full spatio-temporal and process resolution — ultimately with the carbon cycle as just one component of a coupled earth-system model. It also needs to be recognised that compound models may be assembled from components representing different parts of the modelling spectrum.

---

[1] In the names of both my departmental affiliation and my operational unit, 'Mathematics' and 'Statistics' are treated as separate entities.

Table 1 list characteristics of the modelling spectrum with examples of model classes from carbon cycle studies (based on [19]) and climate modelling. Developments over the last 2 decades have enlarged the range of modelling approach so that empirical 'black-box' techniques can also include approaches such as forms of data-mining [*****], and empirical fitting techniques such as artificial neural networks [1].

2.2 Inference

The theme of my book can be summarised as 'inversion needs to be considered as a process of statistical inference'. However, in retrospect, the descriptions of statistics are probably too abbreviated. One important distinction in statistical inference is between frequentist vs Bayesian inference [42]. Often the two approaches lead to performing the same inference calculations (as I pointed out) but the underlying concepts are different (which I didn't emphasise sufficiently) and use some different terminology (which I sometimes mixed incorrectly). A strictly frequentist description of inference, in a book oriented to atmospheric science, is given by von Storch and Zwiers [111, chapters 4 and 5]. In contrast, Tarantola's book [107, 108] is explicitly Bayesian. The Garthwaite et al. textboook [42] describes both frameworks.[1]

One classic definition is that Bayesian is someone who thinks it makes sense to talk about probability of a parameter value??? *check citation*, or [2] that all uncertainty should be expressed as probabilities.

For multivariate normal

$$\Theta = -\ln[\Pr(\mathbf{x}|\mathbf{z})] = [\mathbf{H}(\mathbf{x}) - \mathbf{z}]^{\mathsf{T}}\mathbf{X}[\mathbf{H}(\mathbf{x}) - \mathbf{z}] \tag{2.1}$$

One minor technical point that I encountered in related work concerns the use of the Kalman Filter as a recursive implementation of regression [22, Box 4.4]. If this is used to implement conventional regression (illustrated here as just a running mean) with a large prior (denoted P_0), stability issues can arise as the first increment of data with uncertainty R, is assigned a posterior uncertainty of

$$P_0 - \left(\frac{P_0}{P_0 + R}\right)^2 (P_0 + R)$$

Formally, this gives the correct limit of R as $P_0 \to \infty$, but this is calculated as a difference of two large ($\to \infty$) terms (with a matrix equivalent for more realistic cases) and so the initial steps in such applications need to be treated with caution.

Finally, I have a few comments about terminology for Bayesian priors. The first is the occasional description of such priors as an initial 'guess'. I think that this represents a poor description of what is happening. Many iterative computations, e.g. the Newton-Raphson algorithm for finding zeroes of functions, start with an initial guess and then iterate the calculation many times until the solution loses information about the initial guess (except for macro-features reflecting basins of attraction). In contrast, Bayes formula represents a single combination of an initial 'prior' estimate (expressed as a distribution) combined with data (again with a distribution) producing a posterior distribution.

---

[1] Participants at the 2006 MSRI-NCAR Carbon Data Assimilation workshop had Doug Nychka's lectures. These, and other workshop lectures, are available on-line.

[2] From Doug Nychka's MSRI-NCAR lecture

Similarly, since the Bayesian inversion formalism is envisaged as a way of adding atmospheric information to information from other types of measurement, my preference is to refer to ‘prior’ distributions, rather than ‘a priori’ distributions since the latter term has connotations (from Kant??) of knowledge existing prior to experience. (The term ‘a priori’ may be appropriate in cases where the Bayes priors genuinely represent ‘subjective probabilities’).

In a politicised environment, using terms such as ‘guess’ and ‘a priori’ seems unwise as they imply less scientific objectivity than is actually the case.

2.3 Statistical modelling

Following from the theme of my book: that ‘inversion needs to be considered as a process of statistical inference’, is the proposition that ‘any statistical analysis requires a statistical model’. Alternatively, clarifying the words of Enting and Pearman [28] “whatever cannot be modelled deterministically needs to be modelled statistically”.

My book cites a small number of time series analyses of CO_2 data, expressing the results in standard forms such as ARIMA models, [69] and [110] (which was cited as Tunnicliffe-Wilson (1979) and missing from the book’s reference list). Another study noted was by Cleveland et al. [11] using robust regression techniques to decompose CO_2 series into trend, seasonal and irregular components. Some of this work is also described in the report [18] from a NOAA/WMO workshop in Hilo, 1988.

The studies in CAR Technical Paper 40 [20] cover a range of aspects of time series analysis of CO_2.

A recent example of a statistical analysis of CO_2 data, conceptualising the observations as signal plus noise, was given by Enting [21]. This analysed monthly mean CO_2 data, c_{xt} where x is a location index and t is a time index. The data were expressed as principle components:

$$c_{xt} = \sum_p \phi_{xp} \lambda_p \psi_{pt} \tag{2.2}$$

with the location and time factors (ϕ_{xp} and ψ_{pt} respectively) normalised and the λ_p decreasing. The dominant coherent behaviour is then captured by the smallest p values (i.e. the largest λ_p). An analysis of this type had been presented by Dargaville [14] and found that about 5 coherent modes could be identified. A similar study by Lintner [68] found only 2 significant non-seasonal modes. The Enting analysis [21] concentrated on the residuals:

$$g_{xt} = c_{xt} - \sum_{p=1}^{M} \phi_{xp} \lambda_p \psi_{pt} \tag{2.3}$$

for small M, regarded as time series in t. For most locations, the autocorrelations were characteristic of a moving average process. An exception was Samoa, which exhibited long-term autocorrelations — this was interpreted as reflecting an ENSO signal. Enting noted that the ‘moving average’ correlations in monthly data were what could be expected when constructing monthly means from weekly data if the weekly data had autoregressive structure.

An important requirement of statistical modelling is consistency. Considering the stationary case, for simplicity, one often expresses observations, $z(t)$, as a sum of signal, $s(t)$, and noise,

$n(t)$, with power spectra related by

$$f_z(\theta) = f_s(\theta) + f_n(\theta) \tag{2.4}$$

Consistency requires that the assumed noise model (here characterised by $f_n(\theta)$) and the statistical characteristics of the signal (here characterised by $f_s(\theta)$), which derives from the statistics of the system's forcing, transformed by the system's response) are consistent with the statistics of the observations (here characterised by $f_z(\theta)$) in the sense that equation (2.4) is satisfied. More generally, one wants statistical consistency as one moves across the modelling spectrum.

The type of statistical study discussed in this section is mainly applicable to the observations that feed into inversion studies. Note in particular, that while there is a formal equivalence between model and data error, the actual covariance structure is likely to be very different and model error is far less susceptible to statistical analysis. One way of addressing model error is intercomparison studies, most notably TRANSCOM in the present case (see Section 5). Other approaches are based on analysis of the results of inversions, and are described in the 'Methodology' section that follows.

3 Inversion methodology

3.1 General

For a general account of inversion methodology, the book by Tarantola [107] has been updated [108] and is also available on-line. This book is squarely based on a statistical approach (specifically using Bayesian statistics) to inversions. An alternative view of a statistical framework of inverse problems is given by Evans and Stark [30].

In my book I note a split in trace gas inversion techniques between synthesis and mass balance techniques. These are characterised respectively by flux interpolation vs concentration interpolation. Increasingly, there is a convergence between these techniques.[1] The synthesis analysis of CO_2 by Enting et al [29] can be characterised as a statistical form of the synthesis used for CH_4 by Fung et al. [39], noting an earlier description by Prahm et al. [83]. As a Bayesian statistical technique, the uncertainty analysis was a defining characteristic of the Bayesian synthesis method. In contrast, the initial mass-balance inversions in 2 or 3 dimensions did not include such formal uncertainty analysis. In my book [22, Section 11.2], I list a number of ways in which mass balance inversions can be mapped onto a state space representation for time-evolving estimation with the Kalman filter as the estimation procedure.

The availability of adjoint techniques has facilitated such calculations. In particular a recent study by Rödenback [96] [[** and other refs] uses a variational technique for the fitting procedure. This appears to correspond to 3D-VAR techniques from meteorology. The 4D-VAR technique seems to be appropriate for non-linear problems, i.e. in cases where a current state is not a 'sufficient statistic' for estimation of future states. Kalnay [54] gives the history of meteorological data assimilation, describing in particular the most recent techniques of 4D-VAR and Ensemble Kalman Filter.

Rayner et al. [94] performed a joint assimilation of atmospheric concentration data and remotely sensed AVHRR vegetation data.

---

[1] For this characterisation, I am indebted to Peter Rayner.

Other methodological studies include:

- the geostatistics approach by Michalak et al. [71] that regularises the problem in terms of a specified spatial correlation structure;
- studies treating data uncertainty as an additional factor to be estimated [72], finding a standard deviation of 0.71 ppm for marine boundary layer sites, 1.49 for desert and high mountain sites and 3.16 for other sites. A study by Krakauer et al. citeTC:krakauer04 use generalised cross-validation to jointly estimate a weighting for scaling TRANSCOM priors and a factor scaling the onservational weights expressed as a proportional scaling between TRANSCOM weighting and uniform weighting.
- 'shrinkage estimators' that deliberately introduce a bias in order to reduce the variance and the mean-square error [102];
- studies for the TRANSCOM intercomparison which is described in Section 5;
- studies using trace gas inversion for 'experimental design' — various cases are described in Section 7.

3.2 Multiple constraints

In recent years, analyses of the carbon cycle have introduced terms like 'multiple constraints', 'model-data fusion', 'model-data synthesis', etc., for the concept of broad-scale intergation of information, especially in carbon cycle studies. Discussions include [60] and [85].

In terms of inverse problems, these discussions are capturing two primary inverse problems:

model calibration (also termed tuning) — estimating $R(.)$ in (1.1), (1.2);

state estimation (also termed data assimilation) — estimating $S(.)$ in (1.1), (1.2).

Often, of course, both activities are required.

Within, this broader concept, inversions based on the transport of trace atmospheric constituents (i.e. the topic of my book [22] and the present update) are but one part of the study.

Of course, the various Bayesian synthesis analyses of CO_2 always involve 'multiple constraints', the additional information being that used to specify the 'priors'. The analysis is based in the cost function

$$\Theta = (\mathbf{Gx} - \mathbf{z})^{\mathsf{T}}\mathbf{X}(\mathbf{Gx} - \mathbf{z}) + (\mathbf{x} - \mathbf{x}_0)\mathbf{W}(\mathbf{x} - \mathbf{x}_0) \tag{3.1}$$

where $\mathbf{W}$ is the inverse covariance matrix for the prior estimates of the fluxes, $\mathbf{X}$ is the inverse covariance matrix for the concentration observations and $\mathbf{G}$ is the Green's function for atmospheric transport, i.e. the relation between fluxes and concentrations.

A key result for the Bayesian synthesis inversion is the expression of uncertainties as

$$\mathbf{Y} = \mathbf{G}^{\mathsf{T}}\mathbf{XG} + \mathbf{W} \tag{3.2}$$

where $\mathbf{Y}$ is the inverse covariance matrix for the posterior estimates of the fluxes. This has formed the basis of 'experimental design' studies described in Section 7 below.

In early applications of Bayesian synthesis, the components of the prior flux estimates have been taken as independent, i.e. $\mathbf{W}$ was taken as diagonal. This leads to simplifications in programming, and in assembling the relevant data, but it is in no way an essential part of the synthesis inversion formalism. Indeed, when the synthesis inversion formalism was extended from the cyclo-stationary case to genuinely time-dependent inversions, it was found that the inversions had to take account of the implicit time-correlation of the prior estimates of fluxes.

The requirements of multiple constraints is to consider the synthesis relation for multiple data sets $\mathbf{z}_j$ related to parameters $\mathbf{x}$ by

$$\mathbf{z}_j = \mathbf{G}_j\mathbf{x} + \mathbf{e}_j \tag{3.3}$$

with inverse covariances, $\mathbf{X}_j$ for the $\mathbf{e}_j$.

Equation (3.2) generalises to give

$$\mathbf{Y} = \mathbf{W}_{\text{prior}} + \sum_j \mathbf{G}_j{}^{\mathsf{T}}\mathbf{X}_j\mathbf{G}_j \tag{3.4}$$

as the inverse covariance matrix for the estimate

$$\hat{\mathbf{x}} = [\mathbf{W}_{\text{prior}} + \sum_j \mathbf{G}_j{}^{\mathsf{T}}\mathbf{X}_j\mathbf{G}_j]^{-1}[\mathbf{W}\,\mathbf{x}_0 + \sum_j \mathbf{G}_j{}^{\mathsf{T}}\mathbf{X}_j\mathbf{z}_j] \tag{3.5}$$

This makes it clear that the results do not depend on the order in which the data sets are added and that the combination of N data sets can be performed sequentially using the posterior estimate (and its variance) from each stage as the prior for the next stage. In terms of the inverse covariance matrix,

$$\mathbf{W}_n = \mathbf{Y}_{n-1} \tag{3.6}$$

with $\mathbf{W}_1 = \mathbf{W}_{\text{prior}}$, giving

$$\mathbf{Y}_n = \mathbf{W}_n + \mathbf{G}_n{}^{\mathsf{T}}\mathbf{X}_n\mathbf{G}_n = \mathbf{W}_{n+1} \tag{3.7}$$

for estimates

$$\hat{\mathbf{x}}_n = [\mathbf{W}_{n+1}]^{-1}[\mathbf{W}_n\,\mathbf{x}_{n-1} + \mathbf{G}_n{}^{\mathsf{T}}\mathbf{X}_n\mathbf{z}_n] \tag{3.8}$$

In statistical terminology, $\hat{\mathbf{x}}_n$ is a ‘sufficient statistic’ that contains all the information about $\mathbf{x}$ that can be obtained from $\mathbf{z}_1$ to $\mathbf{z}_n$, i.e. all that estimates based on $\mathbf{z}_m$ (for $m > n$) need to know about $\mathbf{z}_1$ to $\mathbf{z}_n$ is contained in $\hat{\mathbf{x}}_n$.

This convenient property is a consequence of the relations being linear with multivariate normal distributions with known covariance.

More generally, the relations are defined in terms of combining probability distributions, where, for 2 data sets, one has the Bayesian relation

$$\Pr(\mathbf{x}|\mathbf{z}) \propto \Pr(\mathbf{z}|\mathbf{x})\,\Pr_0(\mathbf{x}) \tag{3.9}$$

generalise to

$$\Pr(\mathbf{x}|\mathbf{z}_1,\mathbf{z}_2) \propto \Pr(\mathbf{z}_2|\mathbf{x})\,\Pr(\mathbf{z}_1|\mathbf{x})\,\Pr_0(\mathbf{x}) \propto \Pr(\mathbf{z}_2|\mathbf{x})\,\Pr_1(\mathbf{x}) \tag{3.10}$$

This is still assuming that the data sets $\mathbf{z}_1$ and $\mathbf{z}_2$ are independent, given $\mathbf{x}$.

Consideration of particular instances of multiple constraints approaches need to be considered in terms of the types of data, and the appropriate distributions for use in (3.10).

Canadell et al. [9] listed the various types of data input to carbon cycle studies as:

- air sampling networks interpreted by inverse modelling;
- satellite data, for quantities such as leaf-area index and phenology;
- terrestrial biosphere models;
- convective boundary layer measurements;
- stand-level flux networks;
- ecosystem experiments;
- small cuvettes.

In addition, satellite measurements of concentrations are now being produced and will expand with the launch of OCO.

The inversion methodology must also adapt to the statistical characteristics of the relations. In particular, approaches drawing on data assimilation (e.g. CCDAS [94]) may have less computational flexibility about the order in which the data are combined.

Raupach et al. [85] list key characteristics of the error, $\Pr(\mathbf{z}|\mathbf{x})$ in the present notation, as

- magnitude;
- degree of correlation between components;
- temporal correlation structure;
- spatial correlation structure;
- distribution;
- mismatches in averaging;
- contribution from model representativeness error.

3.3 Gradient methods

Two difficulties are concealed by the mathematical elegance of the linear estimation equations that come from multi-variate normal distributions.

- the formalism does not apply if either the linear relation, $\mathbf{z} = \mathbf{G}\mathbf{x} + \epsilon$, or the normality assumption for the ϵ is invalid;
- even if these assumptions apply, using the linear equations may not be the best way to find the minimum of the cost function.

Gradient techniques aim to minimise the cost function directly, using generic minimisation techniques, based on the gradient. For the linear case derived from (3.1),

$$\frac{1}{2}\nabla_{\mathrm{x}}\Theta = \mathbf{G}^{\mathsf{T}}\mathbf{X}[\mathbf{G}\mathbf{x} - \mathbf{z}] + \mathbf{W}[\mathbf{x} - \mathbf{x}_{\mathrm{prior}}] \tag{3.11}$$

In terms of computational efficiency, $\mathbf{G}\mathbf{x} - \mathbf{z}$ is easy to evaluate, since $\mathbf{G}\mathbf{x}$ is obtained by a single model integration with sources $\mathbf{x}$. Multiplying this vector by $\mathbf{X}$ is a simple matrix times vector operation in general, although specific cases are often much simpler if $\mathbf{X}$ has a block-diagonal (or even diagonal) structure due to independence of various subsets of data. The difficulty comes from multiplying this vector by $\mathbf{G}^{\mathsf{T}}$. The direct approach, used in synthesis inversion, is to calculate the full matrix $\mathbf{G}$ by integrating the model with a set of basis functions. The number of such integrations is the dimensionality of $\mathbf{x}$. For high-resolution inversions this becomes computationally infeasible.

The alternative to such a 'brute-force' approach is to use what is known as an adjoint model. This is a model whose operation corresponds to the effect of $\mathbf{G}^{\mathsf{T}}$. This adjoint model is then run with $\mathbf{X}[\mathbf{G}\mathbf{x} - \mathbf{z}]$ as its input. A more detailed discussion of adjoints is given in Section 3.5. This gives three different perspectives: that of vector spaces (i.e. the formal defintion of an adjoint relation), that of differential equations (in terms of relations for sensitivities), and that of computer science, since there are software tools [43, 44] that take the computer code (that formally computes $\mathbf{G}\mathbf{x}$ for arbitrary $\mathbf{x}$) and automatically generates code that calculates $\mathbf{G}^{\mathsf{T}}\mathbf{z}'$ for arbitrary $\mathbf{z}$. Adjoint techniques apply to linearised calculations as well as the fully-linear case described here.

These adjoint techniques have been the basis for obtaining high-resolution inversions, regularised, at least in part, by smoothness constraints rather than relying on fixed spatial structure in low-resolution basis functions [97].

3.4 Algorithmic differentiation

Section 10.6 of the book describes the role of adjoint models in inversions and data assimilation. Adjoint modelling is based on algorithmic differentiation, the process of transforming a computer model into a derivative model that calculates gradients. A key reference is by Griewank [44].

Algorithmic differentiation provides an important tool for transforming computer models into a 'model analysis system' for performing the various tasks involving differentiation (as opposed to integration of the basic equations). These include calibration, sensitivity analysis and data assimilation.

As described in my presentation *Automatic differentation in the analysis of strategies for mitigation of global change* at MODSIM 2005 [24] there are several ways of obtaining code to calculate derivatives:

A: A program to calculate the derivatives can be derived from a model program by hand-coding the line-by line transformations. This is laborious, error-prone and needs to be repeated each time the model changes. (It is this possibility of hand-coding that makes the term 'algorithmic differentiation' preferable to 'automatic differentiation').

B: In many cases, symbolic algebra systems such as Mathematica can be programmed to produce derivatives by adding a single command to the program – producing adjoint models in this way would seem problematic.
C: Tangent/adjoint compilers are tools that analyse the source code for a model and produce code that implements the tangent-linear or adjoint models.
D: Rather than analysing the program directly, operator overloading of steps in a model can be used to produce a 'script' that can be analysed to produce code for the derivatives — projects to implement this approach are ADOL-F and ADOL-C (using Fortran and C++ respectively).
E: Another approach is to use the operator overloading capabilities directly – this is straightforward for the tangent-linear-model, but the application to adjoint modelling is more difficult and more restricted in scope. Apart form my own work, described below, Straka [103] has described a system for Fortran-95 with the code available on-line.

Adjoint modelling is playing an increasing role in trace gas inversions by allowing consideration of highly-resolved fluxes. Early examples were the study by Kaminski et al. of aggregation corrections [56] and study involving data assimilation into process models. Adjoint modelling has formed the basis of extensive high-resolution inversions by Rödenbeck and co-workers [98, 97, 99, 96].

Adjoint modelling has also played a major role in wider earth system modelling by the MIT group [see appendix for URL].

The use of the capability of obtaining derivatives to produce a 'carbon cycle data assimilation system' (CCDAS) is described by Rayner et al. [94].

My own research is developing approach E, with the initial application (using C++) applied to analysis of the Brazilian Proposal [24]. Ongoing research (using Fortran-95) aims to apply automatic differentiation to the calibration of terrestrial carbon models. The operator overloading approach, described by by Griewank [44], makes it particularly easy to convert numerical models to calculate an associated 'tangent linear model' (see definition in [22, eqn 10.6.8]). Using this method for adjoint calculations is probably only possible in special cases.

3.5 Adjoints

The discussion of adjoints can be expressed in three different forms: that of abstract vector spaces, that of transformations of specific classes of differential equations, and finally in terms of the software operation of adjoint compilers.

In vector space terms, we often have gradients, expressed as some form of inner product

$$\Theta = \langle \mathbf{x}'(.)|\mathbf{x}'(.)\rangle \tag{3.12}$$

(displaying my physicist past by using the Dirac notation)..

Sensitivities are given as

$$\frac{1}{2}\frac{\partial}{\partial p_\alpha}\Theta = \langle \mathbf{x}'(.)|\mathbf{u}^{[\alpha]}(.)\rangle \tag{3.13}$$

where $\mathbf{x}(.)$ represents the state vector $\mathbf{x}$ regarded as a function evolving over time, $\mathbf{x}'(.) = \mathbf{x}(.) - \mathbf{z}(.)$ represents a 'data-mismatch, and $\mathbf{u}^{[\alpha]}$ represents its sensitivity with respect to parameter p_α.

The overall gradient is given by:

$$\frac{1}{2}\sum_{\alpha=1}^{N}\frac{\partial}{\partial\alpha}\Theta=\sum_{\alpha=1}^{N}\langle\mathbf{x}'(.)|\mathbf{u}^{[\alpha]}(.)\rangle \tag{3.14}$$

These sensitivities can be written in terms of differential operators, $\mathcal{L}$:

$$\mathcal{L}\mathbf{u}^{[\alpha]}(.)=\mathbf{y}^{[\alpha]}(.) \tag{3.15}$$

We can (formally) define an adjoint operator $\mathcal{L}^{\mathsf{T}}$ and consider the equation

$$\mathcal{L}^{\mathsf{T}}\mathbf{v}(.)=\mathbf{x}'(.) \tag{3.16}$$

To show that this can be (incredibly) useful, we re-write (***) as:

$$\sum_{\alpha=1}^{N}\langle\mathbf{x}'(.)|\mathbf{u}^{[\alpha]}(.)\rangle=\sum_{\alpha=1}^{N}\langle\mathcal{L}^{\mathsf{T}}\mathbf{v}(.)|\mathbf{u}^{[\alpha]}(.)\rangle=\sum_{\alpha=1}^{N}\langle\mathbf{v}(.)|\mathcal{L}\mathbf{u}^{[\alpha]}(.)\rangle=\sum_{\alpha=1}^{N}\langle\mathbf{v}(.)|\mathbf{y}^{[\alpha]}(.)\rangle \tag{3.17}$$

This shows that if we can find the operator $\mathcal{L}^{\mathsf{T}}$ and solve (3.16) by using the data-mismatch as a forcing function for these adjoint equations, then the solution $\mathbf{v}(.)$ is a *single* (vector) function that can be combined with each of the sensitivities $\mathbf{y}^{[\alpha]}(.)$ which do not involve a model integration. If $\mathcal{L}$ is a differential operator, then evaluating $\mathbf{v}(.)$ is requires a single (vector) integration, in addition to the integration needed to calculate $\mathbf{x}(.)$, rather than needing N integrations of the $\mathbf{u}^{[\alpha]}(.)$, one for each value of α.

4 CO$_2$

4.1 Flux inversions

At the TRANSCOM meeting in Tsukuba in 2004, I gave a presentation[1] entitled *Forty years of CO$_2$ inversions: What have we learned?*, the 40 years reflecting (approximately) the period since the inverse calculation by Bolin and Keeling [3].

The main conclusions were:

- the space-time distribution of CO$_2$ contains extensive information about the carbon cycle, but not necessarily about the things of most interest;
- in spite of the north-south asymmetry in land-sea distribution, spatial data provide only a weak constraint on land-ocean partitioning of global fluxes;
- the ill-conditioned nature of the inversion problem implies a need for careful analysis;
- in spite of the work of TRANSCOM (see section 5), we still have a poor grasp of model error.

---

[1] Versions of this talk have been given subsequently at other places, including the MSRI carbon data assimilation workshop.

Specific results noted were:

- The net tropical source was smaller than expected from the combination of ocean outgassing (derived from Δp_{CO2} measurements) and fluxes from land-use change;
- there was a robustly-determined northern hemisphere sink, with less-reliable partitioning between land and ocean;
- weak indications of a high-latitude southern (ocean) source;
- greater precision in estimates of interannual variation, compared to long-term means.

More recently, Heimann et al. [50] have reviewed what is known about the space-time distribution of sources and sinks of CO_2. The role of inversions is derived mainly from TRANSCOM and Rödenbeck [97]. Some of the particular points noted were:

- $<$ 30 % reduction in variance of posterior flux estimates relative to priors (but less error when integrated over larger regions);
- the expected outgassing in tropical oceans and uptake in the North Atlantic;
- terrestrial sinks in the tropics;
- an apparent weak ocean source at high southern latitudes;
- a generally weak sink at northern temperate latitudes except for a possibly transient source in eastern Europe.

Another important result [7, 97] is that estimates of temporal anomalies seem far more robust than estimates of means.

Appendices to the present 'update' lists a number of recent CO_2 inversion studies and give comparisons.

Studies related to other aspects of the carbon cycle are:

- Suntharalingam et al. [104] considered the role of non-CO_2 carbon emissions (i.e. various forms of reduced carbon, mainly CO and CH_4) and the effect on CO_2 inversions. Their conclusions were generally similar to those of Enting and Mansbridge [27] based on a 2-D model:
    - the inclusion of a non-surface source has only a small effect on the inversion;
    - other factors involved in the conversion from CO_2 fluxes to carbon fluxes can be done as off-line calculations.
- Note also the C4MIP activities described in Section 9.

4.2 Process inversions

In my book [22, Section 12.4] I note the scope for going beyond inversions of fluxes in calculations that directly address processes. Studies by Fung et al. [40] and Tans et al. [106] were noted as fore-runners of this approach, and applications to non-CO_2 gases were also noted. The calculation by Knorr and Heimann [59], using seasonal cycle of CO_2 to fit parameters for light-use-efficiency and the temperature sensitivity of respiration in a terrestrial model. An overview of 'process-inversion' was given by Rayner [86].

The gradient from the non-linear relation (**) is

$$\frac{1}{2}\nabla_{\mathbf{x}}\Theta = \mathbf{G}^{\mathsf{T}}\mathbf{X}[\mathbf{H}(\mathbf{x}) - \mathbf{z}] \tag{4.1}$$

where $\mathbf{G}^{\mathsf{T}}$ is the adjoint of the linearisation of the reation $\mathbf{H}(\mathbf{x})$.

Recent work in this area includes a study by Vukićević et al. [112] estimating parameters in a globally-aggregated carbon pool model using inverse techniques based on an adjoint model.

A list of process inversions is given in an appendix.

5 TransCom

Section 9.2 of my book addresses the issue of transport model error. Model error is a troublesome issue across a wide range of inverse problems. One initiative for trace gas inversions has been the TransCom model intercomparison, set up to parallel other earth system intercomparisons such as AMIP and OCMIP.

There have been several phases of TransCom, targeted at assessing the effect of model error on CO_2 inversions:

Phase 1: CO_2 transport This study compared 'forward' calculations for the two main source components: fossil emissions and the seasonal biotic exhange. The results were published by Law et al. [63] with further detail in a technical paper [91]. An important issue was the so-called 'rectifier effect': the mean gradients established due to covariance between fluxes and transport. It was found that models differed significantly in their estimates of the strength of the rectifier. Further discussion was given by Denning et al. [16].

Phase 2: SF_6 transport This was undertaken as a transport experiment using a compound for which there were good observations and well-known sources [17].

Phase 3: CO_2 inversions This is the 'core' TransCom experiment.

Some of the published results have been:

- an initial summary [48] and more detailed studies of annual means [46] and seasonal cycles [47];
- a study of interannual variability by Baker et al. [2].

On-going TransCom activities are:

non-surface data looking at the role of aircraft data and similar data sets;

TCCON The continuous data experiment, looking at the role of high-frequency (typically hourly) data, as opposed to the monthly mean data used in most inversions;

Statistical modelling This aims to explore aspects of the statistical modelling including diagnostics to detect breakdown of the statistical assumptions. [1]

Related activities, OptIC and C4MIP, are noted in Section 9.

TRANSCOM exists primarily to study the impact of model error in CO_2 inversions (although inversion needs to be regarded more broadly than the Bayesian synthesis that is the 'standard' TRANSCOM case). However, one of the consequences of working to a template negotiated well in advance of the calculations is that as inversions, TRANSCOM calculations will not be state-of-the-art, but rather closer to a lowest-common-denominator of what can be achieved by a number of groups.

Some of the features where recent TRANSCOM presentations lag current-best-practice inversions are:

- omission of carbon fluxes into the atmosphere in forms other than CO_2 (and the omission of a CO_2 source in the free atmosphere from CO oxidation);
- omission of any uncertainty in the fossil component;
- the calculations have low spatial resolution without any specification of truncation error. (An appropriate formulation is given in [109] – the applicability to CO_2 inversions was analysed by Kaminski et al. [56] — a schematic is given in [22, Figure 8.3] and this truncation correction is described in mode detail in an appendix below).

Indeed, the first two omissions are specific simplifications for TRANSCOM since a fossil fuel uncertainty and a CO flux were included in the first published Bayesian inversions of CO_2 [25, 29], and increasing resolution in inversion calculations is reducing the discretisation problems.

6 Observations

My book [22] did not give extensive details of observational data sets, and the present 'update' is correspondingly short. Some recent developments have been:

- expansion of NOAA/CMDL flask network,[2] with an increase in land sites and some new ocean sites, including Easter Island, which replaced Cape Grim, Tasmania as being the sampling site with the lowest annual mean surface CO_2 concentration;
- expansion of flux-tower networks;
- development of the Lo-Flow CO_2 analyser [33], with improved instrument stability and precision giving the potential for sensitive determination of spatial gradients, and low requirements for reference gases leading to reduced costs for deploying such instruments;

---

[1] This is on-going work by Anna Michalak and the present author, with an initial presentation at the AGU Fall Meeting, 2004.

[2] But a contraction of the Australian network, in association with the changes noted in the preface.

- improved understanding of issues, especially ocean outgassing, involved in interpreting oxygen observations;

Satellite data have the potential to totally change the nature of trace gas inversions. However since launches of dedicated satellite CO_2 missions lie in the future (as of 2006) the main role of CO_2 inversion is in 'experimental design' (see Section 7.3). However this has been a particularly important role.

7 Experimental design

7.1 Overview

Chapter 13 of my book covers 'Experimental design'. The idea is that an inversion technique that includes a systematic assessment of uncertainty can be applied to assess the utility (as measured by reduction in uncertainty) of putative new data.

This particularly simple in the case of a linear model with Gaussian error because in this case the posterior uncertainty does not depend on the values of the putative data, but only on its uncertainty (as indicated by equation 3.4). Examples using Bayesian synthesis inversion of CO_2 were given in the initial paper [29], considering aspects such as the utility of ^{13}C data, and improved precision in CO_2 data. The book explored joint reductions in the land-ocean partitioning from various forms of ^{13}C data [22, Fig13.1].

A more extensive application of 'experimental design' was the evaluation of the CO_2 sampling network, initially as optimal location of additional stations and also as re-configuration of the entire network [90]. The latter case involved a complex optimisation with multiple local minima, and a simulated annealing technique was used, drawing on applications in seismology [49]. A similar approach to network design, using simulated annealing as the optimisation technique, has also been applied in oceanography. Further network design studies were reported by Patra and co-workers [78, 80]. Rayner [88] has extended network design calculations through optimisation in the presence of model error, this being based on results from TRANSCOM.

Two areas of on-going research where trace inversions are being used for 'experimental design', described in the following sections are:

- analysing the utility of continuous data (hourly or better time-resolution);
- analysing the utility of satellite data.

7.2 High time-resolution

Most global-scale inversions of CO_2 have used monthly mean data, obtained mainly from flask sampling. However, there are long-term continuous measurements from the NOAA CMDL (NOAA GMCC) program at the South Pole, Samoa, Mauna Loa and Point Barrow. Other national programs also maintain continuous analysers.

The development of the Lo-Flow analyser potentially allows wider deployment of continuous analysers by reducing the demands for calibration gases.

Law and collaborators have undertaken a series of experimental design studies, exploring the utility of various modes of deployment [64, 65, 66].

Inversions using continuous data (often inverting hourly values) place severe demands on the modelling:

i. the transport modelling needs to be based on analysed winds in order to match specific events;

ii. the statistical modelling needs to capture a complicated space-time structure of the transport fields.

For experimental design studies, point (i) is not a problem, since the analysis can be done using synthetic data. Point (ii) remains a challenge to researchers. Additional research is being conducted through the TRANSCOM continuous data activity.

7.3 Satellite data

The potential for measuring CO_2 concentrations from satellites brings the prospect of global-scale coverage at high spatial resolution and regular time sampling. Against this stand the technical difficulties of achieving reliable measurements from space. An overview of the OCO (Orbiting Carbon Observatory), scheduled for launch in 2008, is given by Crisp et al. [13].

This creates a valuable role for 'experimental design' studies that assess the potential utility of satellite data. Initial studies [92, 93, 75] suggested that a useful reduction in uncertainty in estimated fluxes could be achieved with column-integrated values with an 8° by 10° footprint and 2.5 ppm uncertainty on monthly mean values.

Pak and Prather [77] compared the relative utility of various cases such as tropospheric vs. whole column and spatial resolution vs. coverage. They noted in particular that the most useful data would be for tropical regions, i.e. those that currently have least surface data.

Houweling et al. [52] compared OCO and SCIAMACHY. In particular they found that techniques based on near-infra-red rather than thermal infra-red were preferable because of better performance near the surface.

8 So what?

Section 4 has summarised some of what we have learned about the carbon cycle. This section begins to address some of the significance in the context of CO_2 being the most important of the anthropogenic greenhouse gases.

** see also Field and Raupach volume ***

Some of the issues are:

- the scale of concentrations over which the response of CO_2 concentrations is quasi-linear response,
- the issues of carbon systems, especially on land, as being vulnerable to climate change. Inversions can play a role, both in detection of such change and for assessing sensitivities that might, via models, be used to predict such vulnerability;
- the point at which such vulnerability leads to feedbacks on the coupled CO_2-climate system and to thresholds and instabilities – again, inversion can play a role in both early detection and model validation. C4MIP [37] studies are described in section 9 below.

A starting point for assessing implications of inversion results is to consider fluxes in terms of their 'functional role' in the carbon cycle. The 'functional role' was used in a two-way classification [22, Fig 14.2] of 'role' vs 'reservoir', emphasising that, part from the fossil component, the relation between resevoir and role was ambiguous. The primary classification of roles was equilibrium vs. perturbation, with perturbations split into deterministic and random (noting that this split was context-dependent). The deterministic perturbations were further split into forcing and response.

In identifying and characterising the 'roles' of various processes, it can be helpful to use the concept of the modelling spectrum and to map mechanistic models from the white box end of the spectrum onto descriptive models from the darker end of the spectrum.

The importance of inversion techniques is the characterisation of uncertainties. My book notes the book by Morgan and Henrion [73] as a comprehensive discussion of uncertainty. Realistic characterisation of uncertainty is particularly important at the present when it is being widely argued that existing uncertainty is an excuse for not taking any action on climate change. Pollack [82] presents a powerful counter-agument.

9 Other topics

There are many other important areas of continuing research in trace gas inversions. This section flags a few of these — hopefully I will be able to expand some of these in later versions of this document.

North American carbon sink The question of a large carbon sink in North America arose in the work of Fan et al. [32].[1] While this work attracted considerable attention, there are several technical reasons for questioning the result:

- concentrations were calculated with respect to the South Pole, thus giving a correlated data error that was not taken into account in the calculations;
- the discretisation of flux regions was mis-matched to the data set, with high spatial resolution not being associated with a correspondingly high data density;
- within the TRANSCOM group, the particular model had an interhemispheric transport that was higher than most others.

In addition, it should be noted that:

- the particular years that were analysed included a period of widespread reduction in CO_2 emissions from natural systems, following the Pinatubo eruption;
- the analysis in [32] was a CO_2 inversion, and consequently conversion to a carbon budget requires accounting for carbon transport by rivers and carbon emissions in non-CO_2 gases. This issue was raised by Sarmiento and Sundquist [100] and is further discussed in [22, Section 14.1].

---

[1] A quantitative analysis of the extent to which each of these factors contributes to the anomalous result in [32] has been undertaken by Rachel Law and has contributed to this section.

rectifier effect The extent to which inversions would be affected by the degree of seasonal covariance between emissions and transport was apparent from Phase I of TRANSCOM, where it appeared that models tended to be divided in groups with large rectifier effects and thos with low rectifier effects, mainly on the basis of how the model treated the boundary layer. The issue was emphasised by Denning et al. [16].

A set of comparisons by [62] suggested that the modelling of the rectifier could have a significant effect on estimates of land-ocean partitioning of fluxes within zones, but a lesser effect on larger (semi-hemispheric) scales. They also suggested that estimates of interannual variability were more robust (with respect to rectifier uncertainty) than estimates of long-term means.

analysis of specific events Specific events that appear to have had a global-scale impact on CO_2 are the Pinatubo eruption (as noted above) and extensive fires in Indonesian and the European drought of 2003.

Feedbacks An important area of carbon cycle research, although one not currently analysed in terms of inversions, concerns the issue of climate change feeding back into changes in the carbon cycle. Overviews of the issue are given by the conference volume [114], various chapters [70, 15] of the IPCC second assessment [51]. A cooperative study (termed 'flying-leap') of CO_2-climate feedbacks was initiated [37] and later formalised as C4MIP [87].

The feedbacks can be split into oceanic and terrestrial, important examples being, respectively, a decrease in the thermohaline circulation and a climate-induced transition from rainforest to savannah in Amazonia. A calculation by Cox et al. [12] predicted strong feedbacks from both these processes. In contrast a study by Friedlingstein et al. [35] showed little feedback from either process. Comparitive reviews have been presented by Friedlingstein et al. [36, 34] and by Fung et al. [38]. These feedback processes are two of what the IGBP describes as potential 'tipping points' in the climate system, others being:

- reduction of the thermohaline circulation;
- sudden release of methane held as clathrates in the ocean;
- release of methane as tundras warm.

A review of carbon cycle vulnerabilties is given by [45].

Complexity The study of complexity and complex systems has been the basis for a number of attempts at finding common viewpoint spanning disparate fields. Such concepts include multiple interacting scales, thresholds and emergent phenomena.[2]

However claims for the existence of a unifying theory of complexity are controversial. A seminal event in integrating studies of complex systems was the establishment of the Santa Fe Institute [113]. Scoping reviews covering complex systems have appeared in *Science* [41, and subsequent articles] and *Nature* [116, and subsequent articles]. The

---

[2] My own involvement in complex systems has been through the CSIRO Complex Systems Emerging Science Initiative and through MASCOS: the ARC Centre of Excellence for Mathematics and Statistics of Complex Systems.

Science review included an article by Rind [95] on complexity and climate, although this tended to identify complexity with chaos. Falkowski et al. [31] addressed the carbon cycle as a complex system, and Scheffer et al. [101] considered ecosystem intability, a topic of considerable importance given the feedback issues noted above.

A technical paper of mine [23] discusses the role of data assimilation in complex systems. The main thrust is that those studying complex systems could well learn from the earth system sciences, using data assimilation to study systems that are not ammenable to controlled experiment. Finally, it is worth noting the comment by Kadanoff [53] that while there may be no such thing as "complex systems science", there is good science to be done by studying complex systems.

OptIC This is an on-going data-assimilation intercomparision, conducted under the auspices of the Global Carbon Project.

Methane For CH_4, the main difference from CO_2 inversions is the presence of a sink process (oxidation by hydroxyl radicals) in the free atmosphere. If the sink rate was known, and strictly proportional to methane concentrations, then it would be possible to define Green's functions that described the surface sources and the resulting sink, and combine these in a synthesis [22, Section 15.2]. However, most if not all studies, have treated the sink as an unknown that must be estimated and performed iterative calculations to obtain a consistent budget.

Methane inversions are described in Chapter 15 of my book, noting the study [39] where the authors apparently coined the term "synthesis" for this type of study, and mass-balance inversions using a prescribed sink [67]. A more recent mass-balance study [8] uses a chemical transport model. The use of a full chemical model can take account of the effect of methane concentration changes affecting the strength of the methane sink, though reduction of hydroxyl concentrations.

OH trend Possible trends in atmospheric hydroxyl has remained an important research issue, both on the time-scales of direct observations and the longer time-scales of ice-core data. Going beyond the brief account in my book is outside the scope of this draft update, except to note work by Bousquet et al. [6].

10 Summary

This first draft leaves many key issues unexplored. Some of these are listed in Section 9. My own research direction, which was one of the attractions of the move to MASCOS, is in refining the statistical models that underlie inversions. Put simply, I believe that, for trace gas inversions, the use of Gaussian distributions is a simplification that is rapidly approaching its 'use-by' date.

References

[1] G. Abramowitz. Towards a benchmark for land surface models. *Geophys. Res. Lett.*, 32:L22702, doi:1029/20005GK024419, 2005.

[2] D. F. Baker, R. M.Law, K. R. Gurney, P. Rayner, P. Peylin, A. S. Denning, P. Bousquet, L. Bruhwiler, Y.-H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, K. Masarie, K. Masarie, M. Prather, B. Pak, S. Taguchi, and Z. Zhu. TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO_2 fluxes, 1988–2003. *Global Biogeochemical Cycles*, 20:GB1002, doi:10.1029/2004GB002439, 2006.

[3] B. Bolin and C. D. Keeling. Large-scale atmospheric mixing as deduced from the seasonal and meridional variations of carbon dioxide. *J. Geophys. Res.*, 68:3899–3920, 1963.

[4] P. Bousquet, P. Ciais, P. Peylin, and P. Monfray. Inverse modelling of annual atmospheric CO_2 sources and sinks 1. Method and control inversion. *J. Geophys. Res.*, 104D:26161–26178, 1999.

[5] P. Bousquet, P. Ciais, P. Peylin, and P. Monfray. Inverse modelling of annual atmospheric CO_2 sources and sinks 2. Sensitivity study. *J. Geophys. Res.*, 104D:26179–26193, 1999.

[6] P. Bousquet and Others. Two decades of OH variability as inferred by an inversion of amospheric transport and chemistry of methyl chloroform. *Atmos. Chem. Phys. Discuss.*, 5:1679–1731, 2005.

[7] P. Bousquet, P. Peylin, P. Ciais, C. Le Quéré, P. Friedlingstein, and P. P Tans. Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science*, 290:1342–1346, 2000.

[8] T. M. Butler, I. Simmonds, and P. J. Rayner. Mass balance inverse modelling of methane in the 1990s using a chemistry transport model. *Atmos. Chem. Phys.*, 4:2561–2580, 2004.

[9] J. P. Canadell, H. A. Mooney, D. D. Baldocchi, J. A. Berry, J. R. Ehleringer, C. B. Field, S. T. Gower, D. Y. Hollinger, J. E. Hunt, R. B. Jackson, S. W. Running, G. R. Shaver, W. Steffen, S. E. Trumbore, R. Valentini, and B. Y. Bond. Carbon metabolism of the terrestrial biosphere: a multitechnique approach for improved understanding. *Ecosystems*, 3:115–130, doi: 10:1007/s100210000014, 2000.

[10] P. Ciais, P. P. Tans, J. W. C. White, M. Trolier, R. J. Francey, J. A. Berry, D. R. Randall, P. J. Sellers, J. G. Collatz, and D. S. Schimel. Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}C$ measurements from the NOAA climate monitoring and diagnostics laboratory global air sampling network. *J. Geophys. Res.*, 100D:5051–5070, 1995.

[11] W. S. Cleveland, A. E. Freeny, and T. E. Graedel. The seasonal component of atmospheric CO_2: information from new approaches to the decomposition of seasonal time series. *J. Geophys. Res.*, 88C:10934–10946, 1983.

[12] P. M. Cox, R. A. Betts, and C. D. Jones. Acceleration of global warming due to carbon-cycle feedback in a coupled climate model. *Nature*, 408:184–187, 2000.

[13] D. Crisp, R. M. Atlas, F.-M. Breon, L. R. Brown, J. P. Burrows, P. Ciais, B. J. Connor, S. C. Doney, I. Y. Fung, D. J. Jacob, C. E. Miller, D. O'Brien, S. Pawson, J. T. Randerson, P. Rayner, R. J. Salawitch, S. P. Sander, B. Sen, G. L. Stephens, P. P. Tans, G. C. Toon, P. O. Wennberg, S. C. Wofsy, Y. L. Yung, Z. Kuang, B. Chudasama, G. Sprague, B. Weiss, R. Pollock, D. Kenyon, and S. Schroll. The Orbiting Carbon Observatory (OCO) mission. *Advanc. Space Res.*, 34:700–709, 2004.

[14] R. Dargaville. *The variability of atmospheric CO_2 and its surface sources*. PhD thesis, University of Melbourne, 1999.

[15] K. Denman, E. Hofmann, and H. Marchant. Marine biotic responses to environmental change and feedbacks to climate. In Houghton et al. [51], chapter 10, pages 483–516.

[16] A. S. Denning, I. Y. Fung, and D. Randall. Latitudinal gradient of atmospheric CO_2 due to seasonal exchange with the land biota. *Nature*, 376:240–243, 1995.

[17] A. S. Denning, M. Holzer, K. R. Gurney, M. Heimann, R. M. Law, P. J. Rayner, I. Y. Fung, S.-M. Fan, S. Taguchi, P. Friedlingstein, Y. Balkanski, J. Taylor, M. Maiss, and I. Levin. Three-dimensional transport and concentration of SF_6: A model intercomparison study (TransCom 2). *Tellus*, 51B:266–297, 1999.

[18] W. P. Elliott, editor. *The Statistical Treatment of CO_2 Data Records*. NOAA ERL, Silver Spring, Md, 1989.

[19] I. G. Enting. A modelling spectrum for carbon cycle studies. *Math. Comput. Simulation*, 29:75–85, 1987.

[20] I. G. Enting. Characterising the temporal variability of the global carbon cycle. CSIRO Atmospheric Research Technical Paper no. 40 CSIRO :Australia), 1999. Electronic edition at:
http://www.dar.csiro.au/publications/Enting_2000a.pdf.

[21] I. G. Enting. An empirical characterisation of signal vs. noise in CO_2 data. *Tellus*, 54B:301–306, 2002.

[22] I. G. Enting. *Inverse Problems in Atmospheric Constituent Transport*. CUP, Cambridge, UK, 2002.

[23] I. G. Enting. *Inverse problems in earth system science: A complex systems perspective*. CSIRO Atmospheric Research Technical Paper no. 62. CSIRO, Australia, 2002. http://www.dar.csiro.au/publications/enting_2002c.pdf.

[24] I. G. Enting. Automatic differentiation in the analysis of strategies for mitigation of global change. MODSIM05 (International Congress on Modelling and Simulation), Melbourne, 2005. http://www.mssanz.org.au/modsim05/papers/enting.pdf, 2005.

[25] I. G. Enting, R. J. Francey, C. M. Trudinger, and H. Granek. Synthesis inversion of atmospheric CO_2 using the GISS tracer transport model. Technical Report Technical Paper no. 29, CSIRO Division of Atmospheric Research, 1993.

[26] I. G. Enting and J. V. Mansbridge. Seasonal sources and sinks of atmospheric CO_2: Direct inversion of filtered data. *Tellus*, 41B:111–126, 1989.

[27] I. G. Enting and J. V. Mansbridge. Latitudinal distribution of sources and sinks of CO_2: Results of an inversion study. *Tellus*, 43B:156–170, 1991.

[28] I. G. Enting and G. I. Pearman. Average global distributions of CO_2. In M. Heimann, editor, *The Global Carbon Cycle*, pages 31–64. Springer-Verlag, Heidelberg, 1993.

[29] I. G. Enting, C. M. Trudinger, and R. J. Francey. A synthesis inversion of the concentration and δ^{13}C of atmospheric CO_2. *Tellus*, 47B:35–52, 1995.

[30] S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Problems*, 18:R55–R97, 2002.

[31] P. Falkowski, R. J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Höberg, S. Linder, F. T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, and W. Steffan. The global carbon cycle: A test of our knowledge of the earth as a system. *Science*, 290:291–296, 2000.

[32] S.-M. Fan, M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento, T. Takahashi, and P. Tans. A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models. *Science*, 282:442–446, 1998.

[33] R. J. Francey and L. P. Steele. Measuring atmospheric carbon dioxide — the calibration challenge. *Accred. Qual. Assur.*, 8:200–204, doi: 10.1007/s00769–003–0620–1, 2003.

[34] P. Friedlingstein. Climate-carbon cycle interactions. In C. B. Field and M. R. Raupach, editors, *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*, chapter 10, pages 217–224. Island Press, Washington DC, 2004.

[35] P. Friedlingstein, L. Bopp, P. Ciais, J.-L. Dufresne, L. Fairhead, H. le Treut, P. Monfray, and J. Orr. Positive feedback between future climate change and carbon cycle. *Geophys. Res. Lett.*, 28:1543–1546, 2001.

[36] P. Friedlingstein, J. L. Dufresne, P. M. Cox, and P. J. Rayner. How positive is the feedback between climate change and the carbon cycle? *Tellus*, 55B:692–700, 2003.

[37] I. Fung, P. J. Rayner, P. Friedlingstein, and D. Sahagian. Full-form earth system models: coupled carbon-climate interaction experiment (the “flying leap”). *Global Change Newsletter*, 41:78, 2000.

[38] I. Y. Fung, S. C. Doney, and J. John. Evolution of carbon sinks in a changing climate. *Proc. Nat. Acad. Sci.*, 102:11201–11206, 2005.

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

[40] I. Y. Fung, C. J. Tucker, and K. C. Prentice. Application of advanced very high resolution radiometer vegetation index to study atmosphere-biosphere exchange of CO_2. *J. Geophys. Res.*, 92D:2999–3015, 1987.

[41] R. Gallagher and T. Appenzeller. Beyond reductionism. *Science*, 284:79, 1999.

[42] P. H. Garthwaite, I. T. Jolliffe, and B. Jones. *Statistical Inference*. OUP, Oxford, U.K., 2nd edition, 2002.

[43] R. Giering. Tangent linear and adjoint biogeochemical models. In P. Kasibhatla, M. Heimann, P. Rayner, N. Mahowald, R. G. Prinn, and D. E. Hartley, editors, *Inverse Methods in Global Biogeochemical Cycles. (Geophysical Monograph no. 114)*, pages 33–48. AGU, Washington, DC, 2000.

[44] A. Griewank. *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*. SIAM, Philadelphia, 2000.

[45] N. Gruber, P. Friedlingstein, C. B. Field, R. Valnetini, M. Heimann, J. E. Richey, P. Romero Lankao, E.-D. Schulze, and C.-T. A. Chen. The vulnerability of the global carbon cycle in the 21st century: An assessment of carbon-climate-human interactions. In C. B. Field and M. R. Raupach, editors, *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*, chapter 3, pages 45–76. Island Press, Washington DC, 2004.

[46] K. R. Gurney, R. M. Law, A. S. Denning, P. J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, S. Fan, I. Y. Fung, M. Gloor, M. Heimann, K. Higuchi, J. John, E. A. Kowalczyk, T. Maki, S. Maksyutov, P. Peylin, M. Prather, B. C. Pak, J. Sarmiento, S. Taguchi, T. Takahashi, and C. W. Yuen. TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information. *Tellus*, 55B:555–579, 2003.

[47] K. R. Gurney, R. M. Law, A. S. Denning, P. J. Rayner, B. C. Pak, D. Baker, P. Bousquet, L. Bruhwiler, Y. H. Chen, P. Ciais, I. Y. Fung, M. Heimann, J. John, T. Maki, S. Maksyutov, P. Peylin, M. Prather, and S. Taguchi. TransCom 3 inversion intercomparison: model mean results for the estimation of seasonal carbon sources and sinks. *Global Biogeochemical Cycles*, 18:1010, 2004. doi:10.1029/2003GB002111.

[48] K. R. Gurney, R. M. Law, A.S. Denning, P. J. Rayner, D. Baker, P. Bousquet, L. Bruhwiler, Y.-H. Chen, P. Ciais, S.M. Fan, I.Y. Fung, M. Gloor, M. Heimann, K. Higuchi, J. John, T. Maki, S. Maksyutov, K. Masarie, P. Peylin, M. Prather, B.C. Pak, J. Randerson, J. Sarmiento, S. Taguchi, T. Takahashi, and C.W. Yuen. Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature*, 415:626–630, 2002.

[49] M. Hardt and F. Scherbaum. The design of optimum networks for aftershock recordings. *Geophys. J. Int.*, 117:716–726, 1994.

[50] M. Heimann, C. Rödenbeck, and M. Gloor. Spatial and temporal distribution of sources and sinks of carbon dioxide. In C. B. Field and M. R. Raupach, editors, *The Global Carbon Cycle: Integrating Humans, Climate and the Natural World*, chapter 8, pages 187–204. Island Press, Washington DC, 2004.

[51] J. T. Houghton, L. G. Meira Filho, N. Callander, B. A. Harris, A. Kattenberg, and K. Maskell, editors. *Climate Change 1995: The Science of Climate Change*. Published for the IPCC by CUP, Cambridge, UK, 1996.

[52] S. Houweling, F.-M. Breon, I. Aben, C. Rödenbeck, M. Heimann, and P. Ciais. Inverse modelling of CO_2 sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their performance as a function of space and time. *Atmos. Chem. Phys.*, 4:523–538, 2004.

[53] L. P. Kadanoff. Turbulent heat flow: Structures and scaling. *Physics Today*, 54(8):34–39, 2001.

[54] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability*. CUP, Cambridge, 2003.

[55] T. Kaminski, W. Knorr, P. J. Rayner, and M. Heimann. Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Global Biogeochemical Cycles*, 16:1066, 2002. doi:10.1029/2001GB001463.

[56] T. Kaminski, P. J. Rayner, M. Heimann, and I. G. Enting. On aggregation errors in atmospheric transport inversions. *J. Geophys. Res.*, 106:4703–4715, 2001.

[57] W. J. Karplus. The spectrum of mathematical modelling and systems simulation. *Math. Comput. Simulation*, 19:3–10, 1977.

[58] C. D. Keeling, S. C. Piper, and M. Heimann. A three-dimensional model of atmospheric CO_2 transport based on observed winds. 4: Mean annual gradients and interannual variations. In D. H. Peterson, editor, *Aspects of Climate Variability of the Pacific and Western Americas . Geophysical Monograph 55*. AGU, Washington, 1989.

[59] W. Knorr and M. Heimann. Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model. *Tellus*, 47B:471–489, 1995.

[60] B. Kruijt, A. J. Dolman, J. Lloyd, J. Ehleringer, M. Raupach, and J. Finnigan. Assessing the regional carbon balance: Towards an integrated, multiple constraints approach. *Change*, 56, (March–April 2001):9–12, 2001.

[61] R. M. Law, Y.-H. Chen, K. R. Gurney, and TransCom3 modellers. Transcom 3 CO_2 inversion intercomparison: 2. Sensitivities of annual mean results to data choices. *Tellus*, 55B:580–595, 2003.

[62] R. M. Law and P. J. Rayner. Impacts of seasonal covariance on CO_2 inversions. *Global Biogeochemical Cycles*, 13:845–856, 1999.

[63] R. M. Law, P. J. Rayner, A. S. Denning, D. Erickson, I. Y. Fung, M. Heimann, S. C. Piper, M. Ramonet, S. Taguchi, J. A. Taylor, C. M. Trudinger, and I. G. Watterson. Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions. *Global Biogeochemical Cycles*, 10:783–796, 1996.

[64] R. M. Law, P. J. Rayner, L. P. Steele, and I. G. Enting. Using high temporal frequency data for CO_2 inversions. *Global Biogeochemical Cycles*, 16, 2002. doi:10.1029/2001GB001593.

[65] R. M. Law, P. J. Rayner, L. P. Steele, and I. G. Enting. Data and modelling requirements for CO_2 inversions using high-frequency data. *Tellus*, 55B:512–521, 2003.

[66] R. M. Law, P. J. Rayner, and Y. P. Wang. Inversion of diurnally varying synthetic CO_2 : Network optimization for an Australian test case. *Global Biogeochemical Cycles*, 18:1044, 2004. doi:10.1029/2003GB002136.

[67] R. M. Law and P. Vohralik. Methane sources from mass-balance inversions: Sensitivity to transport. CSIRO Atmospheric Research Technical Paper no. 50, 2001. Electronic edition at:
http://www.dar.csiro.au/publications/Law_2001a.pdf.

[68] B. R. Lintner. Characterizing global CO_2 interannual variability with empirical orthogonal function/principle component (EOF/PC) analysis. *Geophys. Res. Lett.*, 29:doi:10.1029/2001GL014419, 2002.

[69] F. Martín and A. Díaz. Different methods of modeling the variability in the monthly mean concentrations of atmospheric CO_2 at Mauna Loa. *J. Geophys. Res.*, 96D:18689–18704, 1991.

[70] J. M. Melillo, I. C. Prentice, G. D. Farquhar, E.-D. Schulze, and O. E. Sala. Terrestrial biotic responses to environmental change and feedbacks to climate. In Houghton et al. [51], chapter 9, pages 445–481.

[71] A. Michalak, L. Bruhwiler, and P. P. Tans. A geostatistical approach to surface flux estimation of atmospheric trace gases. *J. Geophys. Res.*, 109:D14109, doi:10.1029/2003/JD004422, 2004.

[72] A. Michalak, A. Hirsch, L. Bruhwiler, K. R. Gurney, and P. P. Tans. Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions. *J. Geophys. Res.*, 110:D24107, doi:10.1029/205/JD005970, 2005.

[73] M. G. Morgan and M. Henrion. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* CUP, Cambridge, UK, 1990. (with M. Small).

[74] B. Nemry, L. Francois, J.-C. Gérard, A. Bondeau, M. Heimann, and THE PARTICIPANTS OF THE POTSDAM NPP MODEL INTERCOMPARISON. Comparing global models of terrestrial net primary productivity (NPP): analysis of the seasonal atmospheric CO_2 signal. *Global Change Biology*, 5 (suppl.):65–76, 1999.

[75] D. M. O'Brien and P. J. Rayner. Global observations of the carbon budget. 2. CO_2 column from differential absorption of reflected sunlight in the 1.61 μm band of CO_2. *Journal of Geophysical Research*, 107D, 2002. doi:10.1029/2001JD000617.

[76] H. Oeschger and M. Heimann. Uncertainties of predictions of future atmospheric CO_2 concentrations. *J. Geophys. Res.*, 88C:1258–1262, 1983.

[77] B. C. Pak and M. J. Prather. CO_2 source inversions using satellite observations of the upper troposphere. *Geophys. Res. Lett.*, 28:4571–4574, 2001.

[78] P. K. Patra and S. Maksyutov. Incremental approach to the optimal network design for CO_2 surface source inversion. *Geophys. Res. Lett.*, 29:1459, 2002. doi:10.1029/2001/GL013943.

[79] P. K. Patra, S. Maksyutov, M. Ishizawa, T. Nakazawa, and J. Ukita. Interannual and decadal changes in the sea-air CO_2 flux from atmospheric CO_2 inverse modelling. *Global Biogeochemical Cycles*, 19:GBC4013, doi:10/1029/2004GB002257, 2005.

[80] P. K. Patra, S. Maksyutov, and TransCom 3 Modellers. Optimal network design for improved CO_2 source inversion. *Tellus*, 55B:498–511, 2003.

[81] P. Peylin, P. Bousquet, C. Le Quéré, S. Sitch, G. McKinley, N. Gruber, P. Rayner, and P. Ciais. Multiple constraints on regional CO_2 flux variations over land and oceans. *Global Biogeochmeical Cycles*, pages GB1011, doi: 10.1029/2003GB002214, 2005.

[82] H. N. Pollack. *Uncertain Science … Uncertain World.* CUP, Cambridge, UK, 2003.

[83] L. P. Prahm, K. Conradsen, and L. B. Nielsen. Regional source quantification model for sulphur oxides in Europe. *Atmos. Environment*, 14:1027–1054, 1980.

[84] M. J. Prather. Time-scales in atmospheric chemistry: Theory for GWPs for CH_4 and CO, and runaway growth. *Geophys. Res. Lett.*, 23:2597–2600, 1996.

[85] M. R. Raupach, P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biology*, 11:378–397, doi: 10.1111/j.1365–2486.2005.00917.x, 2005.

[86] P. J. Rayner. Atmospheric perspectives on the ocean carbon cycle. In E.-D. Schulze, M. Heimann, S. Harrison, E. Holland, J. Lloyd, I. C. Prentice, and D. Schimel, editors, *Global Biogeochemical Cycles in the Climate System.* Academic, San Diego, 2001.

[87] P. J. Rayner. Flying leap becomes C4MIP. *Research GAIM*, 4:2 (winter 2001):8, 2001.

[88] P. J. Rayner. Optimizing CO_2 observing networks in the presence of model error: results from TransCom 3. *Atmospheric Chemistry and Physics*, 4:413–421, 2004.

[89] P. J. Rayner, I. G. Enting, R. J. Francey, and R. Langenfelds. Reconstructing the recent carbon cycle from atmospheric CO_2, δ^{13}C and O_2/N_2 observations. *Tellus*, 51B:213–232, 1999.

[90] P. J. Rayner, I. G. Enting, and C. M. Trudinger. Optimizing the CO_2 observing network for constraining sources and sinks. *Tellus*, 48B:433–444, 1996.

[91] P. J. Rayner and R. M. Law. A comparison of modelled responses to prescribed CO_2 sources. CRCSHM Technical Paper no. 1, 1995. (and CSIRO Division of Atmospheric Research Technical Paper no. 36) (CSIRO: Australia).

[92] P. J. Rayner and D. O'Brien. The utility of remotely sensed CO_2 concentration data in surface source inversions. *Geophys. Res. Lett.*, 28:175–178, 2001.

[93] P. J. Rayner and D. M. O'Brien. Correction to "The utility of remotely sensed CO_2 concentration data in surface source inversions". *Geophysical Research Letters*, 28:2429, 2001.

[94] P. J. Rayner, M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CC-DAS). *Global Biogeochemical Cycles*, 19:GB2026, doi: 10.1029/2004GB002254, 2005.

[95] D. Rind. Complexity and climate. *Science*, 284:105–107, 1999.

[96] C. Rödenbeck. Estimating CO_2 sources and sinks from atmospheric mixing ratio measurements using a global inversion of atmospheric transport. Max-Planck-Institut für Biogeochemie: Technical Paper 6, 2005.

[97] C. Rödenbeck, S. Houwerling, M. Gloor, and M. Heimann. CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics Discussions*, 3:2575–2659, 2003.

[98] C. Rödenbeck, S. Houwerling, M. Gloor, and M. Heimann. CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, 3:1919–1964, 2003.

[99] C. Rödenbeck, S. Houwerling, M. Gloor, and M. Heimann. Time-dependent atmospheric CO_2 inversions based on interannually varying tracer transport. *Tellus*, 55B:488–497, 2003.

[100] J. L. Sarmiento and E. T. Sundquist. Revised budget for the oceanic uptake of anthropogenic carbon dioxide. *Nature*, 356:589–593, 1992.

[101] M. Scheffer, S. Carpenter, J.A. Foley, C. Folke, and B. Walker. Catastrophic shifts in ecosystems. *Nature*, 413:591–596, 2001.

[102] B. A. Shaby and C. B. Field. Regression tools for CO_2 inversions: application of a shrinkage estimator to process attribution. *Tellus*, 58B:279–292, 2006.

[103] C. W. Straka. ADF95: Tool for automatic differentiation of a FORTRAN code designed for large numbers of independent variables. arXiv.cs:MS/0503014 v1 — submitted to Computer Phys. Comms., preprint 2005.

[104] P. Suntharalingam, J. T. Randerson, N. Krakaeur, D. J. Jacob, and J. A. Logan. Influence of reduced carbon emissions and oxidation on the distribution of atmospheric CO_2: Implications for inversion analyses. *Global Biogeochemical Cycles*, 19:GB4003, doi:10.1029/2005GB002493, 2005.

[105] P. P. Tans, T. J. Conway, and T. Nakazawa. Latitudinal distribution of the sources and sinks of atmospheric carbon dioxide derived from surface observations and an atmospheric transport model. *J. Geophys. Res.*, 94D:5151–5172, 1989.

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

[107] A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.* Elsevier, Amsterdam, 1987.

[108] A. Tarantola. *Inverse Problem Theory and Methods for Data Fitting and Model Parameter Estimation.* SIAM, Philadelphia, 2005.

[109] J. Trampert and R. Sneider. Model estimation biased by truncated expansions: Possible artifacts in seismic tomography. *Science*, 271:1257–1260, 1996.

[110] G. Tunnicliffe-Wilson. Analysis of selected time-series of atmospheric carbon dioxided concentrations. In W. P. Elliott, editor, *The Statistical Treatment of CO_2 Data Records*, pages 82–90. NOAA ERL, Silver Spring, Md, 1989.

[111] H. von Storch and F. W. Zwiers. *Statistical Analysis in Climate Research.* CUP, Cambridge, UK, 1999.

[112] T. Vukićević, B. H. Braswell, and D. Schimel. A diagnostic study of temperature controls on global terrestrial carbon exchange. *Tellus*, 53B:150–170, 2001.

[113] M. M. Waldrop. *Complexity: The Emerging Science at the Edge of Order and Chaos.* Viking, London, 1992.

[114] G. M. Woodwell and F. T. MacKenzie, editors. *Biotic Feedbacks in the Global Climatic System: Will the Warming Feed the Warming?* OUP, New York, 1995.

[115] C. Wunsch and J.-F. Minster. Methods for box models and ocean circulation tracers: Mathematical programming and non-linear inverse theory. *J. Geophys. Res.*, 87C:5647–5662, 1982.

[116] K. Ziemelis. Nature insight: Complex systems. *Nature*, 410:241, 2001.

Note that references [3, 4, 5, 10, 16, 17, 20, 27, 25, 29, 32, 40, 49, 56, 60, 62, 67, 63, 73, 74, 76, 91, 92, 90, 100, 111, 106, 107, 109, 114, 115] are cited in my book — [22] is the book.

Notation

In general, the notation in this update follows the book [22].

$\mathbf{e}$ errors in data values, $\mathbf{z}$.

$\mathbf{G}$ The Green's function for atmospheric transport, i.e. the relation between fluxes and concentrations.

t Time, generally in units of years.

$\mathbf{W}$ the inverse covariance matrix for the prior estimates of the fluxes.

$\mathbf{x}$ The set of model parameters.

$\mathbf{X}$ the inverse covariance matrix for $\mathbf{e}$, the errors in $\mathbf{z}$, the concentration observations or other data.

$\mathbf{Y}$ the inverse covariance matrix for the posterior estimates of the fluxes.

$\mathbf{z}$ set of data values.

Θ Cost funtion, to be minimised in inversion calculation.

Appendix: Summary of studies

This appendix groups a number of recent inversion studies according to the area of interest. Studies particularly related to the TRANSCOM intercomparison are cited in Section 5, experimental design studies for satellite data in section 7.3.

CO_2 flux inversions

TransCom interannual variability [2]

Peylin et al., 2005 A comparison of inversion estimates (with 7 methodological variations) and 'bottom-up' models of oceanic and terrestrial models (2 of each) [81].

Patra et al., 2005 A 64-region (42 land, 22 ocean) time-dependent inversion for 1988 to 2001, [79], concentrating on interannual (and longer) variations in air-sea flux. [***]

Rödenbeck et al., 2003 A set of inversions based on 730 regions (sources at 8° by 10°, with transport calculated at 4° by 5°) with monthly time resolution over 1982–2001. Responses calculated using adjoint model. Compared inversions regularised by (a) spatial aggregation (as in original synthesis inversions), and (b) specified spatial correlation (c.f. geostatistics). Presented at Sendai [99] with update [98] and discussion [97].

TransCom 3 inversions Preliminary overview [48], annual mean (with sensitivity to priors) [46] and sensitivity to data choices [61].

Bousquet et al., 1999 A cyclo-stationary inversion [4] and associated sensitivity analysis [5] using 77 sites for the period 1985–1995.

Rayner et al., 1999 Time-dependent Bayesian synthesis inversion, using Green's functions [89].

Enting et al., 1995 Introduction of Bayesian synthesis inversion. Cyclo-stationary case [29].

Ciais et al., 1995 Two-dimensional mass-balance inversion for 1992, including δ^{13}C, [10].

Enting and Mansbridge, 1991 A two-dimensional mass-balance inversion of CO_2, including a prescribed free-atmosphere source from CO oxidation [27].

Tans et al., 1990 Ad hoc synthesis fit using 3-D model with ocaen p_{CO2} constraint on northern and tropical oceans [106].

Keeling et al., 1989 Ad hoc synthesis fit using 3-D model, with ocean constraint from double deconvolution of box-model [58].

Tans, Conway and Nakazawa, 1989 A 2-dimensional mass-balance inversion of CO_2 [105].

Enting and Mansbridge, 1989 A 2-dimensional mass-balance inversion of CO_2 [26].

CO_2 process inversions

My book (Section 12.4) notes a number of studies [40, 74, 84] that can be regarded as precursors of process inversion. More specific studies are:

Rayner et al. This used a variational technique to analyse the BETHY biospheric model and estimate 3 parameters for each of 13 plant-functional-types as well as 18 generic parameters 9and the concentration offset) [94].

Kaminski et al. Estimation of 2 parameters in SDBM (simple diagnostic Biosphere Model) [55].

Vukićević et al Parameter estimation in a globally-aggregated model [112].

Knorr and Heimann This is a two-parameter inversion, estimating light-use-efficiency and the temperature-sensitivity of respiration [59].

The term 'process-inversion' has also been applied to synthesis inversions with basis components based on specific processes.

Shaby and Field This used 8 processes with estimated geogaphical spread as basis functions in a synthesis study, with variance reduction by using shrikage estimators [102].

A comparable study was presented by Michalak at the MSRI-NCAR workshop, using statistical estimates from proxy variables,r regularised by geostaistical techniques.

Inversion methodology

Estimation of data variances Using maximum likelihood, to estimate variances for classes of sites [72].

Special topics

Reduced carbon A recent study [104] analysed the role of CO, CH_4, etc. in atmospheric transport of carbon and implications for calculation and interpretation of CO_2 inversions. Results are broadly consistent with analysis from earlier 2-D modelling [27].

Comparison of flux inversions

Comparison of inversion results from different groups has been difficult, except under highly-standardised cases such as TRANSCOM. My book proposes the use of nort-south integrated fluxes as (a) coping with different grids and resolution, and (b) reducing the spatial covariance in the estimates. Difficulties from calculations covering different time periods still remain. Comparisons are further complicated by cases that do do consider the role of CO and its precursors. While initial synthesis inversions considered CO, it has generally been neglected in subsequent work.

The effects of neglecting CO and its precursors depend on what is required:

1. neglect of a free atmosphere source in the set of basis functions means that other sources will be correspondingly increase. Other things being equal, this error will have a fairly uniform distribution, somewhat peaked towards to tropics. However a Bayesian inversion will preferentially distribute the error into the regios with the largest priors. A correction for this term will lead to a surface CO_2 inversion that is formally correct.

2. more seriously, many CO_2 inversions are presented in terms of non-fossil fluxes by subracting the full fossil carbon emissions, rather than fossil CO_2 emissions. To correct such a budget to produce non-fossil CO_2 emissions requires adding back on the non-CO_2 fossil emissions. This produces a surface CO_2 flux budget.

3. If a surface carbon flux budget is required, then one needs to add the surface biotic emissions of non-CO_2 carbon (excluding components that are rapidly oxidised to products that return to the ecosystem via rain-out).

4. A carbon storage budget, would require an additional correction for carbon cycled from biosphere to oceans via rivers before returning to the atmosphere.

There are three ways of evaluating these corrections

- from the 2-D analysis of [27];

- from the 3-D study [104];

- by running systhesis inversions with and without proper treatment of the various effects.

Note that all but item (1) is, in principle, an off-line correction, and really constitutes a change in what is being estimated from the inversion. The point where failure to be consistent in the treatment of (2) (3) and (4) will affect the inversion calcualtion (as opposed to its interpretation) is if priors (especially tight priors) are based on the wrong sort of budget.

Truncation example

Figure 8.3 of my book, reproduced below, gives a schematic 2-component example of truncation issues, specifically truncating the expansion to omit small-scale details. However, the description was itself truncated to the extent of being difficult to understand. (This was confirmed at the MSRI-NCAR workshop).

The problem was defined as seeking estimates of $c = a + b$ in a system with two degrees of freedom. (For example, we regarded a and b as CO_2 fluxes from east and west of USA). This reflects the original context of having basis functions fixed, apart from a scale factor, within a geographical region. However, the same issues will be at least as important in terrestrial modelling when a fixed set of parameters is used within a particular biome.

The problem:

How does one address the bias due to making the assumption $a = b$ when performing the inversion?

The different parts of the figure represent different ways of looking at the problem. The dashed $a = b$ line represents the truncated framework in which the problem is considered, e.g. a synthesis case where the basis function does not allow variation between a and b. The dotted lines are contours of constant values of c, whose value is being sort. The solid dot shows the true value, i.e. the figure considers the problem of esimating $a + b$ when the true value violates the $a = b$ assumption built into the modelling.

i This is a representation of the description above. We want a way of projecting the true value onto the correct point on the $a = b$ axis.

ii This shows what happens with limited sampling. A single measurement captures some combination of a and b, but not exactly $a+b$ and so when interpreted as if $a = b$ (the projection onto the dashed line, indicated by heavy segment) it gives a biased estimate. The dotted lines show the measurment uncertainty.

iii The hope is that with enough data (shown here as two cases) the biases cancel out. Hope, as they say, springs eternal.

iv This part shows how it would play out if one acknowledged that $a = b$ may not be true and and some knowledge (with large uncertainty) of $a - b$. This is the Wunsch and Minster [115] recipe of doing the calculations in a larger space and then projecting onto the requisite low-dimensional space at the end.

v If one actually knows the correct $a : b$ ratio and works in a representation that has the correct ratio, then the problem goes away.

vi This is the Trampert and Sneider [109] solution (whose applicability to CO_2 inversions was confirmed by Kaminski et al. [56]) of adding an uncertainty that reflects statistical beliefs about $a - b$. The bias is not removed, but the uncertainty range is expanded to be more likely to inc;ude the true value, i.e. a more realistic uncertainty is produced.

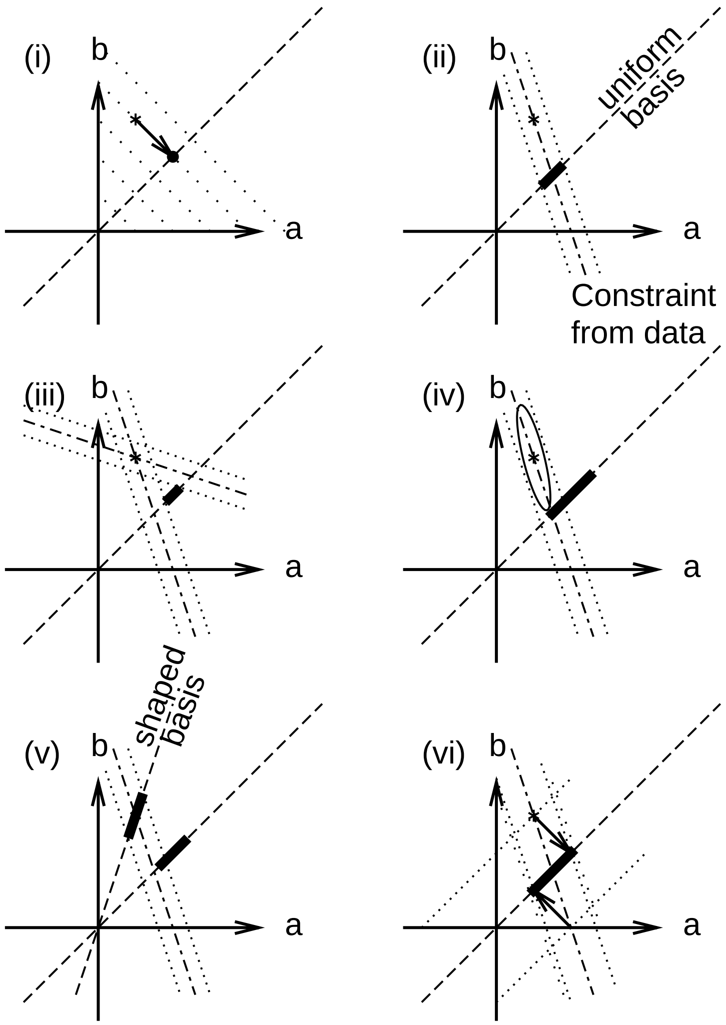


Figure 1: Comparison of treatment of truncation error, from Figure 8.3 of my book [22].

Links

The book [22, Appendix B] gave URLs for data sources from CDIAC, NOAA CMDL, GDCGG, GEIA and EDGAR.

Other web-sites of interest include:

TransCom http://www.purdue.edu/transcom is the TRANSCOM home page. (relocated from Colorado State U).

Global Carbon Project http://www.globalcarbonproject.org

OptIC Within the Global Carbon Project, OptIC

C4MIP http://www.c4mip.cnrs-gif.fr/background.html

Albert Tarantola http://www.ipgp.jussieu.fr/∼tarantola/ is Tarantola's home page, with access to the book [108] and other related publications.

FastOpt http://www.fastopt.com is FastOpt site.
http://www.fastopt.com/topics/publications/html is of particular interest, being their bibliography of publications associated with automatic differentiation .

Andreas Griewank http://www.math.hu-berlin.de/∼griewank/ is Prof. Griewank's home page, now moved from Dresden to Berlin, including material on automatic differentiation.

MIT ESS http://ecco.mit.edu/autodiff.html describes MIT research in ocean data assimilation, using adjoint modelling.

Carbon data assimilation workshop Presentations from the MSRI-NCAR workshop are online at
http://www.atmos.berkeley.edu/∼inez/MSRI-NCAR_CarbonDA/Lectures

In addition, many of the journals cited in the reference list are available on-line (most-commonly on a subscription basis), with varying degrees of back-conversion of print editions for on-line access. IPCC TAR chapters are available on-line. Specific documents are:

- The Rödenbeck technical paper [96].
- My own technical papers [20, 23]. These have been relocated with the formation of the combined division: CSIRO Marine and Atmospheric Research. For the moment the original URL is being re-directed.

Corrections to the book

My intention is to update this document from time to time. Notification of further errors would be appreciated.[1]

eqn 3.3.9a: Description on following line should indicate that $\mathbf{X}$ is the **inverse** covariance matrix, not the covariance matrix.

dobson units ** need to check

In addition, there are a number of corrections to references, with the following list bold-face numbered as in the book with the bracketed numbers referring to the reference list in the present document:

131 My former CSIRO technical papers, including [20], have been relocated with the formation of the combined division: CSIRO Marine and Atmospheric Research. For the moment the original URL is being re-directed.

137 Missing year: 1991 [27].

484 Sneider, not Snieder [109].

missing Cited in chapter 5 as Tunnicliffe-Wilson (1979) [110].

---

[1] Of course, reports of looking hard and not finding additional errors would be appreciated even more.

Production information

This draft typeset (i.e. processed by LaTeX) October 6, 2006.
Last change to text: October, 2006. File: 0-inv/ubook.tex

Status

- 27/4/06: 35.6 kb, 72 refs.
- 16/6/06: 42.08 kb 89 refs.
- 20/7/06: 57.8 kb, 95 refs. Version 0.4.
- 31/7/06 59.8 kb Version 0.5
- 6/10/06 80.4 kb, 116 references: version 0.6