Part A

Principles

Chapter 1

Introduction

One cannot possibly study the disease unless one understands what it means to be healthy. We probably only have a few more decades to study a 'healthy' earth.

R. F. Keeling: Ph. D. Thesis [251].

1.1 Overview

Human activity is changing the composition of the atmosphere. This goes beyond the often obvious problems of local and regional pollution – even in remote locations there are changes in concentrations of minor atmospheric constituents such as carbon dioxide, methane and nitrous oxide. These and other long-lived gases affect the balance of radiation of the earth – they are the so-called greenhouse gases. Other long-lived gases are implicated in the decrease in concentration of ozone in the stratosphere.

The ability to understand the current atmospheric budgets of these trace gases is essential if we are to be able to project their future concentrations in the atmosphere. This book concentrates on one group of techniques that are being used to improve our knowledge – the interpretation of spatial distributions of trace-gas concentrations. An important theme of this book is the use of a statistical approach as being essential to obtaining realistic assessments of the uncertainties in the interpretation of trace-gas observations.

Modelling of the atmospheric transport of carbon dioxide (CO_2), methane (CH_4) and other greenhouse gases is used to interpret the observed spatial distributions of these gases. The spatial distribution of trace-gas concentrations represents a combination of the effect of spatially varying sources and sinks and the effect of atmospheric transport. Therefore a model of atmospheric transport is needed if the source/sink distributions are to be deduced from observed concentrations. The main reason for deducing the source/sink distributions is to help identify and quantify the processes responsible. We define 'tracer inversion' as the process of deducing sources and sinks from measurements of concentrations. We also consider a number of related inversion problems involving trace atmospheric constituents. This use of modelling is termed 'diagnostic' – the model is being used to interpret observations. The alternative use of models is in 'prognostic' operation, in which the model is used to make projections of future conditions.

Modelling of global atmospheric change has progressively widened its scope from the physical properties of the atmosphere to include atmospheric chemistry and biogeochemistry and is progressing to the currently emerging area of 'earth-system science'. This increase in scope has been motivated by the recognition of causal links between the components of the earth system. Realistic projections have to consider these connections and model their evolution in time. In contrast, diagnostic modelling is able to analyse components of the earth system, defining the linkages in terms of observations. Therefore, we can expect that inverse modelling in general (and inverse modelling of the atmosphere in particular) will become an increasingly important part of the development of earth-system science and the validation of earth-system models.

Recognition of the information present in global-scale spatial differences in concentration of CO_2 came soon after the establishment of high-precision measurement programmes at Mauna Loa (Hawaii) and the South Pole in 1958. The CO_2 records revealed a persistent difference and this mean spatial gradient has increased over the subsequent decades. Much of this difference is due to fossil-fuel use, which occurs mainly in the northern hemisphere. It is a measure of the difficulty of interpretation that there remain competing interpretations for the residual.

In order to achieve local air-quality objectives, many jurisdictions have established regulations controlling emissions. On a larger scale, cross-border transport of sulfur compounds has led to international agreements in some cases. On a global scale, the two objectives have been the control of ozone-depleting chemicals through the Montreal Protocol (see Box 16.3) and the restrictions on emission of greenhouse gases through the still-to-be-ratified Kyoto Protocol (see Box 6.1). The existence of these agreements creates a need to ensure that they are based on sound science, in order to ensure that the prescribed actions achieve the objectives of the agreements.

This book aims to capture my own experience and that of my colleagues in using atmospheric-transport modelling to help understand the global carbon cycle and other similar biogeochemical cycles. Since this activity is composed of so many interlinked parts, this introduction is designed to serve as a road-map to what lies ahead. The main division of this book is into Part A, which surveys general principles, and Part B, which reviews recent applications.

The components of tracer inversions are (i) a set of observations, (ii) an atmospheric-transport model and (iii) a set of mathematical/statistical techniques for matching observations to model results. This book is mainly about the matching process. It takes its context from the specific issues raised by the nature of atmospheric transport, the types of observations that are available and what we would like to learn about trace-gas fluxes.

Two important issues that we identify in developing practical inversion calculations are

- ill-conditioning, as introduced in Section 1.3, whereby the inversions are highly sensitive to errors and uncertainties in the inputs and assumptions; and
- the use of a statistical approach to the assessment of uncertainty.

1.2 Atmospheric inversion problems

As noted above, this book is divided into two parts, covering principles and applications, respectively. Nevertheless, principles need illustrative examples and most of the developments of techniques of trace-gas inversion have been in response to specific problems. The main classes of trace-gas inverse problem are the following.

Estimation of atmospheric transport. Inversion calculations to determine atmospheric transport have played a relatively small role in trace-gas studies. An exception is early studies of ozone as a tracer of atmospheric motion. A few tracer studies have concentrated on estimating key indices of transport, such as interhemispheric exchange times. Some of these are reviewed in Chapter 18.

Estimation of sources and sinks of halocarbons. Studies of the various halocarbons have mainly been motivated by their role in ozone depletion. Initially, studies of chlorofluorocarbons (CFCs) concentrated on estimating the loss rates, expressed in terms of atmospheric lifetimes. Studies of methyl chloroform (CH_3CCl_3), for which there are good concentration data and quite good estimates of emissions, also aim to estimate the loss rate. CH_3CCl_3 is removed from the troposphere by reaction with the hydroxyl radical (OH) and so the CH_3CCl_3 loss rate can characterise the loss by reaction with OH of other trace gases, particularly methane [377]. More recently, studies of halocarbons have attempted to estimate the strengths and locations of unreported emissions. Inversions of halocarbon distributions are discussed in Chapter 16.

Estimation of sources and sinks of CO_2. The key issue in studies of CO_2 is the atmospheric carbon budget and, in particular, the partitioning of CO_2 exchanges between oceanic and biospheric processes. Atmospheric CO_2 inversions aim to use the spatial distribution of CO_2 to infer the spatial distribution of surface fluxes, the objective being to obtain sufficient detail to distinguish terrestrial from ocean fluxes. (Note that this book uses the term flux to mean both (i) exchange of mass per unit area, generally in the context of partial differential equations, and (ii) area-integrated exchange of mass, in contexts involving finite areas.) CO_2 inversions are discussed in Chapter 14.

Estimation of sources and sinks of CH_4. As with CO_2, the important questions for CH_4 are those concerning the atmospheric budget. Consequently, the main atmospheric inverse problem is that of estimating the spatial distribution of methane fluxes, mainly from surface-concentration data. The sink in the free atmosphere is an additional complication. Methane inversions are discussed in Chapter 15.

$\boxed{\text{Real world}} \rightarrow \boxed{\text{Mathematical model}} \rightarrow \boxed{\text{Computer model}}$

Figure 1.1 A schematic diagram of the relation among the real world, the mathematical model and the computer model. We adopt the terminology [423] of using 'validation' for testing the mathematical model against the real world and 'verification' for testing the computer model against the mathematical model.

Global-scale inversions of other trace gases are noted in Section 16.4 and regional-scale inversions are discussed in Chapter 17.

We consider the most common tracer inversion problem, that of deducing sources and sinks from concentration data. As noted above, this requires the use of a model of atmospheric transport. Figure 1.1 represents the relation among (i) the real world, (ii) a mathematical model of (some aspect of) the world and (iii) a computer implementation of the mathematical model. Identifying the mathematical model as an explicit intermediate step in model building allows us to use a wide range of mathematical techniques to analyse the modelling process. Much of this book is written in terms of such mathematical models.

The general mathematical form of the transport equation for a trace constituent describes the calculated rate of change with time of $m(\mathbf{r}, t)$, the (modelled) atmospheric concentration:

$$\frac{\partial}{\partial t} m(\mathbf{r}, t) = s(\mathbf{r}, t) + \mathcal{T}[m(\mathbf{r}, t), t] \tag{1.2.1}$$

where $s(\mathbf{r}, t)$ is the local source and $\mathcal{T}[., .]$ is a transport operator. Equation (1.2.1) expresses the rate of change of a trace-gas concentration at a point, $\mathbf{r}$, and time, t, as the sum of the net local source-minus-sink strength at that point, plus a contribution due to trace-gas transport from other locations. The transport is usually modelled with an advective component, $\nabla \cdot (\mathbf{v}m)$, often with a diffusive component to represent sub-grid-scale processes.

We can identify two main classes of inversion, which we denote 'differential' and 'integral'. The former works with equation (1.2.1). The latter uses Green's functions obtained by (numerical) integration of the transport equations. There are also various 'hybrid' techniques.

(a) **Differential inversions.** These are based on rewriting the transport equation (1.2.1) as

$$\hat{s}(\mathbf{r}, t) = \frac{\partial}{\partial t} \hat{m}(\mathbf{r}, t) - \mathcal{T}[\hat{m}(\mathbf{r}, t), t] \tag{1.2.2}$$

where $\hat{s}$ and $\hat{m}$ denote statistical estimates. The most common application is deducing surface sources from surface observations, so (1.2.2) is used at surface grid points, while (1.2.1) is numerically integrated throughout the free atmosphere. Equation (1.2.2) is applied with $\hat{m}(\mathbf{r}, t)$ as a statistically smoothed version of the observed concentration field, $c(\mathbf{r}, t)$ (hence the notation $\hat{m}$). This technique is described as a 'differential' form because of the $(\partial/\partial t)\hat{m}$ term – it is often referred to as the 'mass-balance' technique since the transport equations both in the original and in transformed forms are expressing local conservation of mass. Mass-balance inversion techniques are reviewed in Section 11.1.

(b) **Green-function methods.** These are expressed formally through the Green function, $G(\mathbf{r}, t, \mathbf{r}', t')$, relating modelled concentrations, $m(\mathbf{r}, t)$, to source strengths, $s(\mathbf{r}, t)$,

$$m(\mathbf{r}, t) = m_0(\mathbf{r}, t) + \int_{t_0}^{t} G(\mathbf{r}, t, \mathbf{r}', t')s(\mathbf{r}', t')\, \mathrm{d}^3r'\, \mathrm{d}t' \tag{1.2.3}$$

where $m_0(\mathbf{r}, t)$ describes the way in which the initial state, $m(\mathbf{r}, t_0)$, evolves in the absence of sources. Of necessity, actual calculations are performed using some discretisation of (1.2.3). This is expressed as the generic relation

$$c_j = \sum_\mu G_{j\mu} s_\mu + \epsilon_j = m_j + \epsilon_j \tag{1.2.4}$$

where c_j is an item of observational data, m_j is the model prediction for this item of data, ϵ_j is the error in c_j, s_μ is a source strength and $G_{j\mu}$ is a discretisation of $G(\mathbf{r}, t, \mathbf{r}', t')$.

The discretisation is based on decomposing the sources as

$$s(\mathbf{r}, t) = \sum_\mu s_\mu \sigma_\mu(\mathbf{r}, t) \tag{1.2.5}$$

so that the $G_{j\mu}$ are the responses (for observation j) to a source defined by the distribution $\sigma_\mu(\mathbf{r}, t)$. (Often, for convenience, $G_{j\mu}$ includes pseudo-sources defining the m_0 of (1.2.3), which is assumed to be constant for each species.) The sources are estimated by using (1.2.4) to fit the coefficients, s_μ. For this reason, these Green-function methods that work in terms of pre-defined components, $\sigma_\mu(\mathbf{r}, t)$, have been termed 'synthesis' calculations [165], since the source estimate is synthesised from these pre-defined components.

The most important point for the development of Green-function methods is that (1.2.1) defines a linear relation between the concentrations, $m(\mathbf{r}, t)$, and the sources, $s(\mathbf{r}, t)$, so the full machinery of linear algebra can be applied to solving (1.2.4). Green-function techniques are discussed in Chapter 10.

(c) **Hybrid techniques.** These techniques lie between the differential (mass-balance) and the integral (synthesis) inversions. Generally, they take the form of synthesis inversions over a sequence of relatively short time-steps. Examples of this are the techniques of Brown [56], Hartley and Prinn [197] and Ramonet *et al.* [394]. These and other similar techniques are reviewed in Section 11.2. In addition, there is an exploratory discussion in Chapter 12 of techniques involving non-linear estimation.

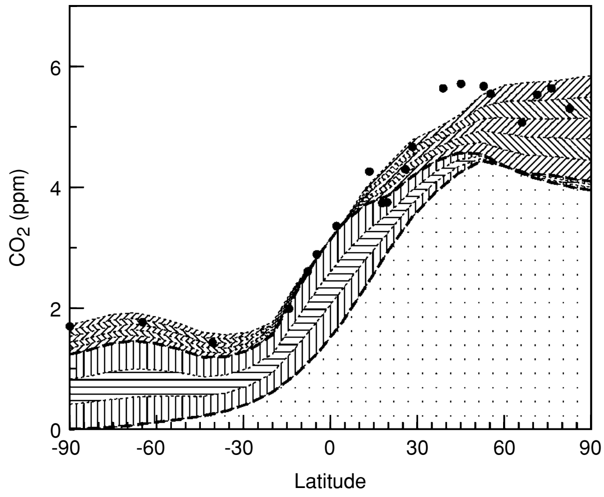

Figure 1.2 A schematic representation of synthesis inversion as analogous to a jigsaw, fitting an unknown number of differently shaped components (fossil, ocean and terrestrial) to find the best fit to observations (solid points).

The emphasis given to Green-function, or synthesis, techniques in this book primarily reflects the scope for error analysis. This also underlies the second reason, which is the extensive experience of synthesis inversions in our research group. In its simplest form, the synthesis approach corresponds to multiple regression: a function $c(x)$ is expressed as a linear combination of specified functions, $G_\mu(x)$, in terms of unknown coefficients, s_μ, by fitting a set of observations at points x_j as $c(x_j) \approx \sum_\mu s_\mu G_\mu(x_j)$.

A visual illustration of the technique can be obtained by regarding the fitting process as a 'jigsaw'. Figure 1.2 gives a schematic representation of the regression (in terms of latitudinal variation only) for CO_2 distributions. Rather than fit pieces of unknown size but known shape, the 'jigsaw' analogy approximates this by fitting an unknown number of pieces of fixed shape and size, shown by alternating hatching. Figure 1.2 demonstrates fitting an unknown number of land pieces (five in this example, above the upper dashed line and shown with diagonal hatching) and an unknown number of ocean pieces (three in this case, between the dashed lines) plus a fairly well-known fossil piece (below the lower dashed line with dot fill) to fit the observed spatial distribution (solid points). (Note that the higher concentrations in the 'land' pieces at high southern latitudes reflect the transport of northern air southwards through the upper troposphere.)

1.3 Uncertainty analysis

A key focus of this book is the estimation of uncertainties. It is particularly important in ill-conditioned problems that are subject to large error-amplification. Uncertainty analysis is required on general grounds of scientific integrity and the needs of policy-related science, as well as for input into further levels of scientific (or policy-related) analysis. In addition, as described in Chapter 13, we have used a systematic uncertainty analysis as the basis of experimental design.

The underlying principle is that *any statistical analysis requires a statistical model.* Generally we need to *assume* such a model. We can test (but never prove) the validity of the assumptions. In discussing the various types of error that can affect trace-gas inversions, Enting and Pearman [141] noted that "any variability that cannot be modelled [$\cdots$] must be treated as part of the 'noise' [$\cdots$]". In other words, the noise in the statistical model is whatever is not being modelled deterministically. Statistical estimation is described in Chapter 3; the special case of time series is described in Chapter 4.

This book follows Tarantola [471] in being firmly based on the use of prior information. This is both an optimal use of available information and an essential part of stabilising ill-conditioned problems. The standard Bayesian formalism has proved adequate for the problems that we have encountered in practice, without the need to adopt the extensions proposed by Tarantola (see Box 12.1).

The emphasis on the statistical modelling of uncertainty leads us to express the results from inversion calculations in the terminology of statistical estimation. The results of inverting (1.2.4) are *estimates*, denoted $\hat{s}_\mu$, of the source components, s_μ, or more generally estimates, $\hat{x}_\mu$, of parameters, x_μ. As noted above, the essential requirement for any statistical analysis is that one must have a statistical model of the problem.

The word 'model' has been used both for the transport model and for the statistical model. This multiple usage needs to be recognised since both types of model are needed for tracer inversions. The transport model represents a deterministic relation between the sources and the concentrations. However, our overall knowledge of the sources and concentrations is incomplete. Most obviously, our knowledge is not infinitely precise. This incompleteness in our knowledge is expressed in statistical terms. With this terminology, the quote from Enting and Pearman [141] needs to be reworded as *any variability that cannot be modelled deterministically* [$\cdots$] *must be treated as part of the 'noise' and modelled statistically*.

In order to address these issues of uncertainty, inversion studies need to use an overall model with both deterministic and statistical aspects. A common form of the combined model is one in which the statistical aspects appear as noise added to the outputs of a deterministic model and (when one is using a Bayesian approach) finite uncertainties on the inputs. With such a structure, there can be an apparent distinction between the deterministic (transport) model and the statistical model. The deterministic model never 'sees' the statistical model – all the statistical analysis occurs somewhere outside. Conversely, the statistical model sees the deterministic model as a functional relation and need take no account of the immense complexity that may lie inside a functional representation of atmospheric transport. This convenient separation between deterministic and statistical models becomes rather less tenable when we wish to apply a statistical approach to considering uncertainties in the deterministic model itself. State-space modelling (see Section 4.4) provides one framework in which statistical and deterministic aspects of modelling can be integrated.

The simplest statistical model of the observations is to assume independent normally distributed errors. For the case of linear relations between observations and parameters, this leads to a weighted least-squares fit giving the optimal (minimum-variance) estimates. Conversely, adopting a least-squares fitting procedure and associated error analysis is equivalent to assuming, whether implicitly or explicitly, that the errors are normally distributed. Errors with a multivariate normal distribution lead to a weighted least-squares fit with 'off-diagonal' weights (see Section 3.3) as the optimal estimates. Other error distributions (and associated non-linear estimation) have been considered in a highly simplified zonally averaged inversion [130]. Additional discussion of non-linear estimation is given in Chapter 12.

Many of the inversions use a Bayesian approach, i.e. independent prior estimates of the sources are included in the estimation procedure. Detailed discussions of applications of Bayesian estimation in carbon-cycle studies have been given for global model calibration [140], for synthesis inversion [143, 145], for methane [242, 206] and in subsequent work.

Some of the key aspects of error analysis are the following.

Measurement error. Most error analyses in trace-gas studies have assumed that the errors in the various observations, c_j, are independent. To the extent that 'error' includes the effect of small-scale sources that are omitted from the model, the assumption of independence of distinct sites can easily fail. Probably, a more important omission is the time correlation in the errors for a single site. Most inversion studies have ignored this problem. Two early exceptions are the three-dimensional model study reported by Bloomfield [34], in which an autoregressive error model was used, and the synthesis inversion by Enting *et al*. [145], in which the issue of autocorrelation in the data was avoided because time-dependence was expressed in the frequency domain.

Model error. The problem of determining the effect of model error remains largely unsolved. The difficulty is particularly great in ill-conditioned inverse problems with their large sensitivity to errors. Tarantola [471] describes a formalism in which model error becomes an extra component added to observational error (see equation (9.1.2)). Enting and Pearman [141] considered such a formalism with particular reference to the 'truncation error' when the small-scale degrees of freedom are excluded from the process of synthesis inversion (see also Section 8.3). One difficulty with this approach is that these errors are unlikely to be independent and there is little basis for defining an appropriate error covariance.

Numerical modelling involves an initial discretisation of the spatial and temporal variations both of sources and of concentrations. Further discretisation of source distributions may be needed because of the limited information contained in a sparse observational network and the loss of information associated with the ill-conditioning of the inversion. The 'synthesis-inversion' approach defined by equation (1.2.4) is usually based on a coarse discretisation of source/sink processes.

In recent years there has been a collaborative project of the International Geosphere–Biosphere Program (IGBP) known as TransCom, which compares some of the atmospheric-transport models used to study CO_2 [398, 277; see also Section 9.2 and Box 9.1]. These studies have confirmed the importance of the problem but have not yielded a 'magic-bullet' solution. A further aspect of 'model error' that must be considered is that of errors in the statistical model.

Source statistics. The Bayesian approach requires, as one of its inputs, prior statistics of the source strengths. One of the most critical issues is the time correlation of these prior source estimates in time-dependent Bayesian inversions. Initially, Rayner *et al.* [402] used time pulses whose prior distributions were assumed to be independent. Later calculations (e.g. those in Figure 14.9) used 'mean-plus-anomaly' representation of data error and prior estimates. Mulquiney *et al.* [332] used a random-walk model for the prior statistical characterisation of sources.

In addition, consideration of the spatial statistics of the sources is required in order to assess the discretisation error inherent in the synthesis approach. The particular importance of spatial statistics is quite explicit for inversions using adjoint calculations [240], in which very large numbers of source components are involved. Assumptions of independent uncertainties for a large number of small regions could imply an unrealistically small uncertainty in the total. Similar issues of spatial statistics are implicit in synthesis inversions based on a small number of highly aggregated components.

The process of deducing trace-gas fluxes from concentration data is severely hampered by a mathematical characteristic known as ill-conditioning. Figure 1.3 gives a schematic representation of the source of the difficulty: dissipative processes in the real world lead to a loss of detail in the information available for analysis. In cases in which the model calculations (or, more precisely, the model-based inferences) are in the opposite direction to the real-world chain of causality, the attenuation of detailed

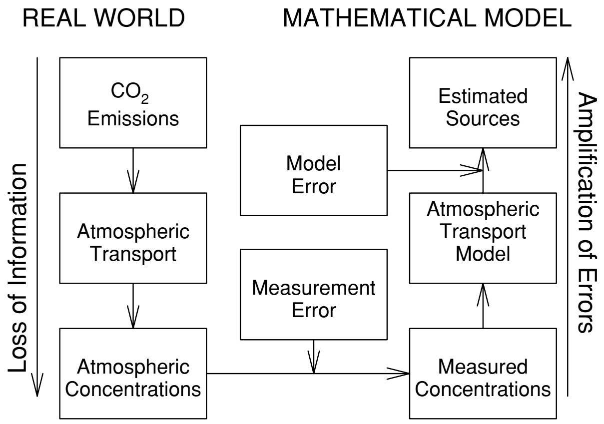

Figure 1.3 The origin of ill-conditioning in the tracer inversion problem. Figure from [132].