

Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications

M. R. RAUPACH*, P. J. RAYNER†, D. J. BARRETT‡, R. S. DEFRIESS§, M. HEIMANN¶, D. S. OJIMA||, S. QUEGAN** and C. C. SCHMULLIUS††

*CSIRO Earth Observation Centre, GPO Box 3023, Canberra, ACT 2601, Australia, †LSCE-CEA de Saclay Orme des Merisiers, 91191 Gif/Yvette, France, ‡CSIRO Land and Water, Canberra, ACT 2601, Australia, §Department of Geography, University of Maryland, College Park, MD 20742, USA, ¶Department of “Biogeochemical Systems”, Max-Planck-Institut für Biogeochemie, D-07701, Jena, Germany, ||Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523-1499, USA, **Centre for Terrestrial Carbon Dynamics, University of Sheffield, Sheffield S37RH, UK, ††Institute für Geographie, Friedrich-Schiller-Universität, D-07743 Jena, Germany.

Abstract

Systematic, operational, long-term observations of the terrestrial carbon cycle (including its interactions with water, energy and nutrient cycles and ecosystem dynamics) are important for the prediction and management of climate, water resources, food resources, biodiversity and desertification. To contribute to these goals, a terrestrial carbon observing system requires the synthesis of several kinds of observation into terrestrial biosphere models encompassing the coupled cycles of carbon, water, energy and nutrients. Relevant observations include atmospheric composition (concentrations of CO₂ and other gases); remote sensing; flux and process measurements from intensive study sites; *in situ* vegetation and soil monitoring; weather, climate and hydrological data; and contemporary and historical data on land use, land use change and disturbance (grazing, harvest, clearing, fire).

A review of model–data synthesis tools for terrestrial carbon observation identifies ‘nonsequential’ and ‘sequential’ approaches as major categories, differing according to whether data are treated all at once or sequentially. The structure underlying both approaches is reviewed, highlighting several basic commonalities in formalism and data requirements.

An essential commonality is that for all model–data synthesis problems, both nonsequential and sequential, data uncertainties are as important as data values themselves and have a comparable role in determining the outcome.

Given the importance of data uncertainties, there is an urgent need for soundly based uncertainty characterizations for the main kinds of data used in terrestrial carbon observation. The first requirement is a specification of the main properties of the error covariance matrix.

As a step towards this goal, semi-quantitative estimates are made of the main properties of the error covariance matrix for four kinds of data essential for terrestrial carbon observation: remote sensing of land surface properties, atmospheric composition measurements, direct flux measurements, and measurements of carbon stores.

Received 8 June 2004; accepted 25 August 2004

Introduction

Systematic earth observation implies the collection and interpretation of multiple kinds of data about the evolving state of the earth system across wide spatial domains and over extended time periods. Three factors have caused a massive acceleration in earth observation activities over recent years. The first is need: global change is raising issues – such as greenhouse-induced climate change, water shortages and imbalances, land degradation, soil erosion, loss of biodiversity – which require informed human responses at both global and regional levels. Second, technological advances in sensors, satellite systems and data storage and processing capabilities are making possible observations and interpretations which were out of reach only a few years ago and unimaginable a few decades ago. Third, the synthesis of formerly discrete disciplines into a unified Earth System Science is driving new hypotheses about the dynamics of the earth system and the interconnectedness of its components, including humans. Systematic earth observation motivates and tests these hypotheses.

The focus of this paper is observation of the carbon cycle, and in particular its land-atmosphere components, as one part of an integrated earth observation system. It is a significant part because of the coupling between the carbon cycle and the terrestrial cycles of water, energy and nutrients, and the connections of all these biospheric processes with global climate and human activities (Field & Raupach, 2004; Raupach *et al.*, 2004). The carbon cycle is integral to the growth and decay of vegetation, maintains the water cycle through transpiration and provides habitat for maintaining biodiversity. Thus, terrestrial carbon observation is important for climate observation and prediction, for the management of water resources, nutrients and biodiversity, and for monitoring and managing the enhanced greenhouse effect.

It is increasingly recognized that strategies for earth observation (including terrestrial carbon observation) require methods for combining data and process models in systematic ways. This is leading to research towards the application in terrestrial carbon observation (and in earth observation more generally) of ‘model–data synthesis’, the combination of the information contained in both observations and models through both parameter-estimation and data-assimilation techniques. Motivations for model–data synthesis approaches include (1) model testing and data quality control (through systematic checks for agreement within specified uncertainty bands for both data and model); (2) interpolation of spatially and temporally sparse observations; (3) inference from available ob-

servations of quantities which are not directly observable (such as carbon stores and fluxes over large areas) and (4) forecasting (prediction forward in time on the basis of past and current observations).

The present paper arose from a workshop held in Sheffield, UK, 3–6 June 2003, to further the development of a Terrestrial Carbon Observation System (TCOS) with a particular emphasis on model–data synthesis. Antecedents for this effort were (1) preliminary steps toward a TCOS (Cihlar *et al.*, 2002a, b, c); (2) a wider concept for an Integrated Global Carbon Observing Strategy including atmosphere, oceans, land and human activities (Ciais *et al.*, 2004) and (3) the research program of the Global Carbon Project (Global Carbon Project, 2003).

The paper is founded on three themes arising from the Sheffield workshop. First, model–data synthesis, based on terrestrial biosphere models constrained with multiple kinds of observation, is an essential component of a TCOS. Second, from the standpoint of model–data synthesis, data uncertainties are as important as data values themselves and have a comparable role in determining the outcome. Third, and consequently, there is an urgent need for soundly based uncertainty specifications for the main kinds of data used in terrestrial carbon observation. These themes are developed as follows: the next section summarizes major purposes and attributes of a TCOS. ‘Model–data synthesis: methods’ provides an overview of model–data synthesis in the context of terrestrial carbon observation, by briefly describing some of the main methods, indicating their common characteristics, and highlighting the key role of data uncertainty. ‘Model–data synthesis: examples’ provides some examples. ‘Data characteristics: uncertainty in measurement and representation’ undertakes a survey of the uncertainty characteristics of the main kinds of relevant data.

Purposes and attributes of a TCOS

A succinct statement of the overall purpose of a TCOS might be: to operationally monitor the cycles of carbon and related entities (water, energy, nutrients) in the terrestrial biosphere, in support of comprehensive, sustained earth observation and prediction, and hence sustainable environmental management and socio-economic development. These words are congruent with the Framework Document emerging from the Second Earth Observation Summit, Tokyo, April 2004 (<http://earthobservations.org/docs/Framework%20Doc%20Final.pdf>), which calls for a ‘Global Earth Observation System of Systems’ to serve nine areas of socio-economic benefit. A TCOS is a contributor to such a system with relevance to at least six of these areas:

- Understanding climate, and assessing and mitigating climate change impacts;
- Improving global water resource management and understanding of the water cycle;
- Improving weather information and prediction;
- Monitoring and managing inland ecosystems, including forests, and land use change;
- Supporting sustainable agriculture and combating desertification;
- Understanding, monitoring and preventing loss of biodiversity.

To make these contributions effectively, a TCOS must have a number of attributes (see also Running *et al.*, 1999; Cihlar *et al.*, 2002a; Ciais *et al.*, 2004). First, *scientific credibility* is needed to maintain methodological and observational rigour, and to include procedures for estimating uncertainties or confidence limits. Second, *consistency with global budgets* is necessary to respect constraints from global-scale carbon and related budgets incorporating terrestrial, atmospheric and oceanic pools and anthropogenic sources such as fossil fuel burning. Third, sufficient *spatial resolution* is necessary to resolve spatial variations in patterns of land use (typically tens of metres, consistent with high-resolution remote sensing). Fourth, enough *temporal resolution* is needed to resolve the influence of weather, inter-annual climate fluctuations and long-term climate change on carbon and related cycles. Fifth, the system needs to encompass a broad *range of entities*, eventually including CO₂, CH₄, CO, volatile organic carbons (VOCs) and aerosol black carbon. Of these, the highest priority is CO₂. Water is also a high priority because of its importance in modulating other terrestrial GHG fluxes. Sixth, a sufficient *range of processes* must be encompassed. A high priority is resolution of net land-air fluxes of greenhouse gases in which all terrestrial sources and sinks are lumped together. However, there is an equally high demand for identification of the terms contributing to the net fluxes, for example to partition a net flux between vegetation and soil storage changes. Finally, *quantification of uncertainty* is required. The 'demand side' of the uncertainty issue is: what level of uncertainty is acceptable for a TCOS to offer useful information? The answer is not simple and depends on the application, for example, from the areas mentioned above. This paper does not attempt to answer the demand-side question, but rather concentrates on the 'supply side' of uncertainty: that is, how uncertainty can be determined in a TCOS based on model-data synthesis and multiple observation sources, each with its own specified uncertainty.

Model-data synthesis: methods

In this section, we survey a range of model-data synthesis methods potentially applicable in a TCOS. More detail and further references can be found in a growing number of excellent sources, for instance Tarantola (1987) and Evans & Stark (2002) for high-level treatments of the general statistical problem of inverse estimation, Grewal & Andrews (1993) and Drécourt (2003) for introductions to the Kalman Filter, Reichle *et al.* (2002) for hydrological applications with an emphasis on the Kalman Filter and Enting (2002) and Kasibhatla *et al.* (2000) for applications of a range of methods to biogeochemical cycles.

Overview

The central problem is: using appropriate observations and models, we must determine the spatial distributions and temporal evolutions of the terrestrial stores and fluxes of carbon and related entities (water, nutrients, energy) across the earth. Important fluxes include land-air exchanges (atmospheric sources and sinks), exchanges with rivers and groundwater, and exchanges between terrestrial pools such as biomass and soil. We also need to determine the main processes influencing the fluxes, including those under human management. No single model or set of observations can supply this amount of information – hence the need for a synthesis approach. The task of combining observations and models can be carried out in many ways, encompassed by the umbrella terms 'model-data synthesis' or 'model-data fusion'. The general principle is to find an 'optimal match' between observations and model by varying one or more 'properties' of the model. (Words in quotes have specific meanings defined below). The optimal match is a choice of model properties, which minimizes the 'distance' between the model representations of a system and what we know about the real biophysical system from observational and prior 'data'. At this high level of generality, model-data synthesis encompasses both 'parameter estimation' and 'data assimilation'. All applications rest on three foundations: a model of the system, data about the system, and a synthesis approach.

Model. For a TCOS, the model is a terrestrial biosphere model describing the evolving stores and fluxes of carbon, water, energy and related entities. This *dynamic model* has the form

$$\frac{dx}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}) + \text{noise} \quad \text{or} \\ \mathbf{x}^{n+1} = \boldsymbol{\Phi}(\mathbf{x}^n, \mathbf{u}^n, \mathbf{p}) + \text{noise} = \mathbf{x}^n + \Delta t \mathbf{f}(\mathbf{x}^n, \mathbf{u}^n, \mathbf{p}) + \text{noise}, \quad (1)$$

where \mathbf{x} is a vector of *state variables* (such as stores of carbon, water and related entities, or store attributes such as age class distributions); \mathbf{f} is a vector of rates of change (net fluxes where components of \mathbf{x} are stores); $\boldsymbol{\varphi}$ is the discrete analogue for \mathbf{f} ; \mathbf{u} is a set of externally specified time-dependent forcing variables (such as meteorological variables and soil properties) and \mathbf{p} is a set of time-independent model parameters (such as rate constants and partition ratios). In the discrete formulation, time steps are denoted by superscripts. The noise terms account for both imperfections in model formulation and stochastic variability in forcings (\mathbf{u}) or parameters (\mathbf{p}). Once the model function $\mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p})$ or $\boldsymbol{\varphi}(\mathbf{x}^n, \mathbf{u}^n, \mathbf{p})$ is specified, then the system evolution $\mathbf{x}(t)$ can be determined by integrating Eqn (1) in time (with zero noise), from initial conditions $\mathbf{x}(0)$, with specified external forcing $\mathbf{u}(t)$ and parameters \mathbf{p} .

Data. These are generally of two broad kinds: (1) observations or measurements of a set of quantities \mathbf{z} and (2) prior estimates for model quantities (\mathbf{x} , \mathbf{u} and \mathbf{p}). Both include uncertainty, through errors and noise. In this paper, the term ‘data’ includes both observations and prior estimates, and incorporates the uncertainty inherent in each.

The measured quantities (\mathbf{z}) are related to the system state and external forcing variables by an *observation model* of the form

$$\mathbf{z} = \mathbf{h}(\mathbf{x}, \mathbf{u}) + \text{noise}, \quad (2)$$

where the operator \mathbf{h} specifies the deterministic relationship between the measured quantities and the system state. The noise term accounts for both ‘measurement error’ (instrumental and processing errors in the measurements \mathbf{z}), and ‘representation error’ (errors in the model representation of \mathbf{z} , introduced by shortcomings in the observation model \mathbf{h}). In the rare case where we can observe all state variables directly, \mathbf{h} reduces to the identity operator, so $\mathbf{z} = \mathbf{x} +$ (measurement) noise. In time-discrete form, Eqn (2) becomes $\mathbf{z}^n = \mathbf{h}(\mathbf{x}^n, \mathbf{u}^n) + \text{noise}$. Note the interpretation of the time-step superscripts: \mathbf{x}^n and \mathbf{u}^n are simply the model state and forcings at time step n , whereas \mathbf{z}^n is the set of new observations introduced at time step n , whatever the actual time of its measurement. However, no observations may be used more than once.

Examples of potential observations in a TCOS include (1) atmospheric composition (concentrations of CO_2 and other gases); (2) remote sensing of terrestrial and atmospheric properties; (3) fluxes of carbon and related entities, with supporting process observations, at intensive study sites; (4) vegetation and soil stores of carbon from forest and ecological inventories; (5)

hydrological data on river flows, groundwater, and concentrations of C, N and other entities; (6) soil properties and topography; (6) disturbance records (both contemporary and historical) including land management, land use, land use change and fire and (8) climate and weather data (precipitation, solar radiation, temperature and humidity). Of these, some (especially the first five) typically provide observational constraints (\mathbf{z}), while others provide model drivers (\mathbf{u}). Examples of observation models (Eqn (2)) include radiative transfer models to map modelled surface states into the radiances observed by satellites; atmospheric transport models to transform modelled surface fluxes to measured atmospheric concentrations; and allometric relations to transform modelled biomass to observed tree diameters.

Synthesis. The final requirement is a synthesis process, or a systematic method for finding the optimal match between the data (including observations and prior estimates) and the model. This process needs to provide three kinds of output: optimal estimates for the model properties to be adjusted, uncertainty statements about these estimates, and an assessment of how well the model fits the data, given the data uncertainties. In any synthesis process, there are three basic choices: (1) the model properties to be adjusted or ‘target variables’, (2) the measure of distance between data and model or ‘cost function’ and (3) the search strategy for finding the optimum values. Search strategies can be classified broadly into (3a) ‘nonsequential’ or ‘batch’ strategies in which the data are treated all at once, and (3b) ‘sequential’ strategies in which the data arrive in a time sequence and are incorporated into the model–data synthesis step by step. The rest of this section explores the choices (1), (2), (3a) and (3b).

Target variables

The target variables are the properties of the model to be adjusted in the optimization process. They include any model property considered to be sufficiently uncertain as to benefit from constraint by the data. Model properties which can be target variables include: (1) model parameters (\mathbf{p}); (2) forcing variables (\mathbf{u}^n), if there is substantial uncertainty about them; (3) initial conditions on the state variables (\mathbf{x}^0) and (4) time-dependent components of the state vector \mathbf{x}^n . The inclusion of the state vector \mathbf{x}^n as a possible target variable is for the following reason: in a purely deterministic model the trajectory \mathbf{x}^n is determined by the dynamical model (\mathbf{f} or $\boldsymbol{\varphi}$), the values of \mathbf{p} and \mathbf{u}^n , and the initial value \mathbf{x}^0 . It might seem sufficient, therefore, to estimate these and allow integration of

the model to take care of \mathbf{x}^n . However, the model itself may not be perfect, as indicated by the noise term in Eqn (1), so there may be advantage in adjusting values of \mathbf{x}^n through the model integration.

To maintain generality, we denote the vector of target variables by \mathbf{y} . This vector may or may not be a function of time, and will usually be a subset of all model variables ($\mathbf{x}^n, \mathbf{u}^n, \mathbf{p}$). Broadly speaking, parameter estimation problems are those where the target variables are restricted to model parameters (\mathbf{p}), while data assimilation problems may include any model property as a target variable, usually with an emphasis on state variables (\mathbf{x}^n).

Cost function

The cost or objective function J (a function of the target variables \mathbf{y}) defines the mismatch or distance between the model and the data. It can take a wide range of forms, but must have certain properties (for example, it must be monotonic in the absolute difference between data and model-predicted values). A common choice is the quadratic cost function:

$$J(\mathbf{y}) = (\mathbf{z} - \mathbf{h}(\mathbf{y}))^T [\text{Cov } \mathbf{z}]^{-1} (\mathbf{z} - \mathbf{h}(\mathbf{y})) + (\mathbf{y} - \hat{\mathbf{y}})^T [\text{Cov } \hat{\mathbf{y}}]^{-1} (\mathbf{y} - \hat{\mathbf{y}}), \quad (3)$$

where $\hat{\mathbf{y}}$ is the vector of 'priors' (*a priori* estimates) for the target variables, and $[\text{Cov } \mathbf{z}]$ and $[\text{Cov } \hat{\mathbf{y}}]$ are covariance matrices for \mathbf{z} and $\hat{\mathbf{y}}$, respectively ($[\text{Cov } \mathbf{z}]_{mn} = \langle z'_m z'_n \rangle$, with $z'_m = z_m - \langle z_m \rangle$, angle brackets denoting the expectation operator). The first term in Eqn (3) is a sum of the squared distances between measured components of the observation vector (\mathbf{z}) and their model predictions ($\mathbf{h}(\mathbf{y})$), while the second is a corresponding sum of distances between target variables and their prior estimates. The matrices $[\text{Cov } \mathbf{z}]^{-1}$ and $[\text{Cov } \hat{\mathbf{y}}]^{-1}$ represent the weights accorded to the observations and the priors, and thus scale the confidences accorded to each. Their role can be clarified by considering the simple case in which components z_m of the observation vector \mathbf{z} are independent, with variances σ_m^2 ; then $[\text{Cov } \mathbf{z}]^{-1}$ is the diagonal matrix $\text{diag} [1/\sigma_m^2]$ and the squared departures of the measurements (z_m) from the predictions ($h_m(\mathbf{y})$) are seen to be weighted by the confidence measure $1/\sigma_m^2$ for each component.

The model-data synthesis problem now becomes: vary \mathbf{y} to minimize $J(\mathbf{y})$, subject to the constraint that $\mathbf{x}(t)$ must satisfy the dynamic model, Eqn (1). The value of \mathbf{y} at the minimum is the *a posteriori* estimate of \mathbf{y} , including information from the observations as well as the priors. We denote it by $\tilde{\mathbf{y}}$ (so frowns and smiles respectively designate prior and posterior estimates).

Equation (3) defines the generalized least squares cost function minimized by the *minimum-variance* estimate ($\tilde{\mathbf{y}}$) for \mathbf{y} . For any distribution of the errors in the data (observations \mathbf{z} and priors $\hat{\mathbf{y}}$), this estimate is unbiased, and has the minimum error covariance among all linear (in \mathbf{z}), unbiased estimates (Tarantola 1987). Use of Eqn (3) has another, additional foundation: provided that the probability distributions for data errors are Gaussian, it yields a *maximum-likelihood* estimate for \mathbf{y} , conditional on the data and the model dynamics (Press *et al.*, 1992, p. 652; Todling 2000). Outside the restriction of Gaussian distributions, $\tilde{\mathbf{y}}$ as defined by minimizing a quadratic J is not exactly the maximum-likelihood estimate, but it is often not far from it. A quadratic J is widely used even when the data errors are not Gaussian; see Press *et al.* (1992, p. 690) for discussion. There are alternative cost functions J in which model-measurement differences ($\mathbf{z} - \mathbf{h}(\mathbf{y})$) are raised to powers other than 2, the choice in Eqn (3) (Tarantola, 1987; Gershensfeld, 1999). For example, in flood event modelling, the absolute maximum error is needed to capture peak flow rates, while for modelling base flow rates, the mean absolute deviation ($|\mathbf{z} - \mathbf{h}(\mathbf{y})|$ to the power 1) has the desirable property of being less sensitive to outliers than a power 2. Different powers for $|\mathbf{z} - \mathbf{h}(\mathbf{y})|$ produce maximum-likelihood estimates for $\tilde{\mathbf{y}}$ with different distributions for data errors; for example, a power 1 J yields a maximum-likelihood estimate when the data errors are distributed exponentially, and a high-power J preferentially weights outliers such as peak flows. Here, we use a power 2 J exclusively.

Search strategies for nonsequential problems

In nonsequential or batch problems, all data are treated simultaneously and the minimization problem is solved only once. A familiar case is least-squares parameter estimation.

Example. Some of the attributes of these problems are demonstrated by considering a simple linear example, which extends the parameter-estimation problem. Although mathematically straightforward, this case finds important application in the atmospheric inversion methods used to estimate trace gas sources from atmospheric composition observations (see 'Model-data synthesis: Examples'). Here the target variables (\mathbf{y}) are a set of surface-air fluxes, averaged over suitable areas; there is no dynamic model relating fluxes at different times and places to each other; and the observation operator (\mathbf{h}) is a model of atmospheric transport. From the linearity of the conservation equation for an inert trace gas, it follows that \mathbf{h} is linear and can hence be represented by a matrix \mathbf{H}

(Raupach, 2001), thus, $\mathbf{z} = \mathbf{H}\mathbf{y} + \text{noise}$. For now, the noise is assumed to be Gaussian with zero mean and no temporal correlation, and thus completely characterised by an observation error covariance matrix $[\text{Cov } \mathbf{z}]$. By minimizing J analytically, one obtains the expression (Tarantola, 1987, p. 196; Enting, 2002):

$$\tilde{\mathbf{y}} = \hat{\mathbf{y}} + [\text{Cov } \tilde{\mathbf{y}}] \mathbf{H}^T [\text{Cov } \mathbf{z}]^{-1} (\mathbf{z} - \mathbf{H}\hat{\mathbf{y}}), \quad (4)$$

where $[\text{Cov } \tilde{\mathbf{y}}]$, the estimated error covariance of the a posteriori estimate $\tilde{\mathbf{y}}$, is given by

$$[\text{Cov } \tilde{\mathbf{y}}]^{-1} = [\text{Cov } \hat{\mathbf{y}}]^{-1} + \mathbf{H}^T [\text{Cov } \mathbf{z}]^{-1} \mathbf{H}. \quad (5)$$

These expressions already tell us some important things. The posterior estimates are given by the prior estimates plus a term depending on the mismatch between the experimental observations and the observations as predicted by the prior estimates. This mismatch is weighted by our confidence in the observations, $[\text{Cov } \mathbf{z}]^{-1}$. Thus, observations with little weight hardly shift the posterior estimate from the prior, and *vice versa*. Furthermore, the transpose of the observation operator (\mathbf{H}^T) multiplies the weighted mismatch. If this operator is very weak, that is if the available observations are only weakly related to the target variables, then the update to the initial estimate is also small. Finally the posterior covariance $[\text{Cov } \tilde{\mathbf{y}}]$ (Eqn (5)) is bounded above, in some sense, by the prior covariance $[\text{Cov } \hat{\mathbf{y}}]$. If the prior covariance is small (suggesting substantial confidence in the initial estimate) then the increment $\tilde{\mathbf{y}} - \hat{\mathbf{y}}$ (the difference between the posterior and prior estimates, a measure of the information added by the observations \mathbf{z} , and equal to the second term in Eqn (4) in the present case) is also small.

All the above is reasonable. More surprising is the relationship between the data, its uncertainty and the cost function. We can decompose the (positive definite) matrix $[\text{Cov } \mathbf{z}]^{-1}$ into a matrix product $\mathbf{A}^T \mathbf{A}$, using the Cholesky decomposition for a positive definite matrix. For a diagonal covariance matrix, $\text{diag} [\sigma_m^2]$, the decomposition is trivial: $\mathbf{A} = \text{diag} [1/\sigma_m^2]$. Likewise, we can write $[\text{Cov } \hat{\mathbf{y}}]^{-1} = \mathbf{B}^T \mathbf{B}$. The cost function, Eqn (3), can then be rewritten as

$$J(\mathbf{y}) = (\mathbf{a} - \mathbf{A}\mathbf{h}(\mathbf{y}))^T (\mathbf{a} - \mathbf{A}\mathbf{h}(\mathbf{y})) + (\mathbf{b} - \hat{\mathbf{b}})^T (\mathbf{b} - \hat{\mathbf{b}}), \quad (6)$$

where $\mathbf{a} = \mathbf{A}\mathbf{z}$, $\mathbf{b} = \mathbf{B}\mathbf{y}$ and $\hat{\mathbf{b}} = \mathbf{B}\hat{\mathbf{y}}$. Thus the cost function, and thence the entire minimization, takes a form in which neither the observations nor the prior estimates appear; they are replaced by quantities \mathbf{a} and \mathbf{b} scaled by the square roots of the inverse covariance matrices, which are measures of confidence. This is no mathematical nicety; rather it demonstrates that the

data and the uncertainties are completely inseparable in the formalism. To put the point provocatively, providing data and allowing another researcher to provide the uncertainty is indistinguishable from allowing the second researcher to make up the data in the first place. This realization informs the emphasis on uncertainty throughout this paper.

Algorithms for nonsequential problems

The task in general is to find the target variables \mathbf{y} which minimize $J(\mathbf{y})$. Clearly, the shape of $J(\mathbf{y})$ is all important: it may have a single minimum or multiple separated local minima, only one of which is the true global minimum. Near the minimum, J may be shaped like a long, narrow ellipsoidal valley. If this valley has a flat floor tracing out some line in \mathbf{y} space, then all points along that line are equally acceptable and these \mathbf{y} coordinates cannot be distinguished in terms of optimality, so such combinations of target variables cannot be resolved by model–data synthesis with the available data and model. Diagnostic indicators about these issues are provided by the Hessian or curvature matrix $\mathbf{D} = \partial^2 J / \partial y_j \partial y_k$, a measure of the local curvature of $J(\mathbf{y})$. The degree of orthogonality among columns of \mathbf{D} indicates the extent to which it is possible to find a unique local minimum to $J(\mathbf{y})$ in the vicinity of the point at which \mathbf{D} is evaluated. A high ‘condition number’ (ratio of largest to smallest eigenvalue) for \mathbf{D} indicates that some linear combination(s) of the columns of \mathbf{D} are nearly zero, that is, that the curvature is nearly zero in some direction(s), so that the minimization problem is ill-conditioned, as in the case of a valley with a flat floor.

Given these considerations, classes of method for finding the minimum in $J(\mathbf{y})$ include the following.

1. *Analytic solution* is possible when the observation operator $\mathbf{h}(\mathbf{y})$ is linear ($\mathbf{z} = \mathbf{H}\mathbf{y} + \text{noise}$). In this case $J(\mathbf{y})$ is a quadratic form shaped like a parabolic bowl, and the minimization can be carried out analytically as in the example of Eqns (4) and (5). This ‘direct’ or ‘one-step’ solution is highly efficient when applicable; however, most problems are nonlinear and require a nonlinear method.

2. *Gradient descent algorithms* are the most familiar search algorithms for nonlinear optimization. They include (for example) steepest-descent, conjugate-gradient, quasi-Newton and Levenberg–Marquardt algorithms (Press *et al.*, 1992). Gradient-descent methods are easily implemented, provided that the gradient vector $\nabla_y J = \partial J / \partial y_k$ can be calculated. The main advantages of gradient-descent algorithms are relative simplicity and low cost; the main disadvantage is that if the surface $J(\mathbf{y})$ has multiple minima, they tend to find local

minima near the starting value of \mathbf{y} rather than the global minimum.

3. *Global search methods* find the global minimum in a function $J(\mathbf{y})$ by searching (effectively) the whole of \mathbf{y} space. They overcome the local-minimum pitfall (so to speak) of gradient-descent methods, but have the disadvantage of higher computational costs. Simulated annealing and genetic algorithms are two examples. These methods are efficient at finding the vicinity of a global minimum where there may be multiple local minima, but do not locate an exact local minimum. They may be combined with gradient-descent methods for finding an exact global minimum once in the right vicinity.

Search strategies for sequential problems

In sequential problems, the task is to solve for a set of target variables \mathbf{y}^n associated with a particular time step, usually including the state variables of the dynamic model (\mathbf{x}^n). The process is then repeated sequentially to give a time history for \mathbf{y}^n . Information about \mathbf{y}^n can come from two sources: evolution of the dynamic model from the previous time step, and comparison between the observations at the current time step (\mathbf{z}^n) and the model predictions ($\mathbf{h}(\mathbf{y}^n)$).

Kalman filter. Introduced by Kalman (1960), the Kalman filter is by now a group of algorithms for the sequential combination of dynamic and observational information, using a ‘prediction’ step and an ‘analysis’ step. In the prediction step, the dynamic model is used to calculate prior estimates $\hat{\mathbf{y}}^n$ for the target variables at time step n , from the best (posterior) estimates $\hat{\mathbf{y}}^{n-1}$ at the previous step. In the analysis step, posterior estimates $\tilde{\mathbf{y}}^n$ at step n are obtained by ‘improving’ the prior estimates with data. The model state is then ready for evolution to the next ($n + 1$) time step. A key point is that the confidence in the current state, embodied in the error covariance for the target variables \mathbf{y} , is also evolved with the dynamic model and improved with observations. A schematic diagram of the information flow in the Kalman filter is given in Fig. 1.

In the prediction step, the task of evolving \mathbf{y} is common across all implementations of the Kalman filter since it involves only a normal forward step of the dynamic model: $\hat{\mathbf{y}}^n = \boldsymbol{\Phi}(\hat{\mathbf{y}}^{n-1})$. The prior estimate for the covariance at time step n evolves according to

$$[\text{Cov } \hat{\mathbf{y}}^n] = \boldsymbol{\Phi}[\text{Cov } \hat{\mathbf{y}}^{n-1}]\boldsymbol{\Phi}^T + \mathbf{Q}, \quad (7)$$

where $\boldsymbol{\Phi} = \partial\boldsymbol{\phi}/\partial\mathbf{y}$, the Jacobian matrix of the dynamic model $\boldsymbol{\phi}(\mathbf{y})$. The first term on the right represents the propagation of the error covariance in the target

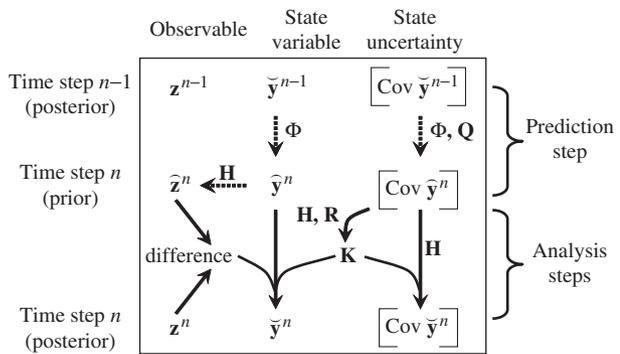


Fig. 1 Information flow in the linear Kalman filter. Linear operators are indicated next to arrows. Operations in prediction and analysis steps are shown as dashed and solid lines, respectively.

variables \mathbf{y} from one time step to the next, by a linearized version of the model. The second term (\mathbf{Q}) is the covariance of the noise term in the dynamic model, Eqn (1), which includes both model imperfections and stochastic variability in forcings and parameters. This term plays a crucial role in the Kalman filter: it quantifies our lack of confidence in the ability of the dynamic model to propagate the model state, and is usually referred to as model error. In most implementations of the Kalman filter the model error is assumed to be Gaussian with zero mean and no temporal correlation, and thus completely characterized by the covariance matrix \mathbf{Q} .

In the analysis step, the prior estimates are refined by the inclusion of data. This is done using the prior estimate for the predicted observation vector, $\hat{\mathbf{z}}^n = \mathbf{h}(\hat{\mathbf{y}}^n)$, and its covariance

$$[\text{Cov } \hat{\mathbf{z}}^n] = \mathbf{H}[\text{Cov } \hat{\mathbf{y}}^n]\mathbf{H}^T + \mathbf{R}, \quad (8)$$

where $\mathbf{H} = \partial\mathbf{h}/\partial\mathbf{y}$ is the Jacobian matrix of the observation model $\mathbf{h}(\mathbf{y})$, and \mathbf{R} is the data covariance matrix $[\text{Cov } \mathbf{z}]$, indicating lack of confidence in the data and often called the data error. Again it is usually assumed that the data error is Gaussian with zero mean and no temporal correlation, and thus completely characterized by $\mathbf{R} = [\text{Cov } \mathbf{z}]$.

The expressions for the final (posterior) estimates for \mathbf{y} and its covariance are now exactly as for the nonsequential mode, except that the operation is carried out for one time step only:

$$\begin{aligned} \tilde{\mathbf{y}}^n &= \hat{\mathbf{y}}^n + [\text{Cov } \hat{\mathbf{y}}^n]\mathbf{H}^T[\text{Cov } \hat{\mathbf{z}}^n]^{-1}(\mathbf{z}^n - \mathbf{h}(\hat{\mathbf{y}}^n)) \\ &= \hat{\mathbf{y}}^n + \mathbf{K}(\mathbf{z}^n - \mathbf{h}(\hat{\mathbf{y}}^n)), \end{aligned} \quad (9)$$

$$\begin{aligned} [\text{Cov } \tilde{\mathbf{y}}^n]^{-1} &= [\text{Cov } \hat{\mathbf{y}}^n]^{-1} + \mathbf{H}^T\mathbf{R}\mathbf{H} \text{ or} \\ [\text{Cov } \tilde{\mathbf{y}}^n] &= (\mathbf{I} - \mathbf{K}\mathbf{H})[\text{Cov } \hat{\mathbf{y}}^n], \end{aligned} \quad (10)$$

where $\mathbf{K} = [\text{Cov } \hat{\mathbf{y}}^n] \mathbf{H}^T [\text{Cov } \hat{\mathbf{z}}^n]^{-1}$ is the Kalman gain matrix. The two equalities in Eqn (10) are equivalent. Time step n is now complete, and we are ready for the next time step. We note that the ratio of the magnitudes of \mathbf{Q} and \mathbf{R} (model and data error covariances) is critical, since it largely determines how closely the evolution of \mathbf{y} follows that suggested by the dynamic model ($\mathbf{Q} \ll \mathbf{R}$) or the data. The tuning of \mathbf{Q} and \mathbf{R} is a crucial part of Kalman filter implementation; see Grewal & Andrews (1993) for an excellent extended discussion.

The concepts underlying the Kalman filter are now implemented in several different ways (see for example Grewal & Andrews, 1993; Evensen, 1994, 2003; Kasibhatla *et al.*, 2000; Reichle *et al.*, 2002; Drécourt, 2003), including the following:

1. The linear Kalman filter (LKF), in which both $\phi(\mathbf{y})$ and $\mathbf{h}(\mathbf{y})$ are linear in \mathbf{y} , can be shown to be an optimal solution for appropriate linear problems.

2. The extended Kalman filter (EKF) applies for nonlinear $\phi(\mathbf{y})$ and $\mathbf{h}(\mathbf{y})$, by linearizing the covariance propagation part of the analysis step (Eqn (7)), but not the prediction step, at each point. This is the algorithm sketched above.

3. The ensemble Kalman filter (EnKF) (Evensen, 1994, 2003) is appropriate for high-dimensional problems such as data assimilation into atmospheric and ocean models, where the error covariance matrix for \mathbf{y} is too large to store, let alone integrate forward. The EnKF uses stochastic methods based on multiple model runs to propagate the covariance matrix without storing it. Also, the EnKF does not explicitly require the Jacobian matrices $\phi(\mathbf{y})$ and $\mathbf{h}(\mathbf{y})$, which can be difficult to derive analytically and expensive to calculate numerically. Reichle *et al.*, (2002) summarize the differences between the EKF and the EnKF.

4. The Kalman smoother assimilates multitemporal information to constrain \mathbf{y}^n at each time point, by running both forward and backward in time (Todling, 2000). It produces an estimate of target variables at time step n based on the entire record, rather than only the record up to time step n . This gives the Kalman smoother the attributes of a nonsequential method, as data at all times are used together.

Adjoint methods. These form an additional group of methods applicable to sequential problems. The principle (le Dimet & Talagrand, 1986; Giering 2000) is to update the target variables (including the model state) by using measurements at nearby times such as the interval between steps n and $n + 1$, and an estimate of the gradient $\nabla_{\mathbf{y}} J$ obtained by backward integration of an 'adjoint model' over that interval. The target

variables are effectively the initial state variables for integration of the model from step n to $n + 1$. This approach underpins four-dimensional data assimilation (4DVAR) methods for assimilating data into atmospheric and oceanic circulation models on weather and climate time scales (Chen & Lamb, 2000; Park & Zupanski, 2003).

Discussion of model–data synthesis methods

Differences between nonsequential and sequential strategies. Parameter estimation and data assimilation problems tend to be amenable to solution by nonsequential and sequential search strategies, respectively. However, this is not an absolute correspondence: many problems can be solved using either nonsequential or sequential strategies.

The most important advantage for sequential methods is the ability of the optimal state to differ from that embodied in the model equations. This requires that the evolving model state \mathbf{x}^n be included among the target variables \mathbf{y} . In principle, \mathbf{y} can also include \mathbf{x}^n in nonsequential methods but, since all time steps are considered simultaneously, the size of the problem is usually intractable. Sequential methods also have the computational advantages that their size does not grow with the length of the model integration, and that they can easily handle incremental extensions to time series observations.

The advantages of nonsequential methods come, naturally, from their ability to treat all data at once. This is a direct advantage in itself. It is, for example, difficult for a sequential method to treat the impact of a datum on a state variable some time in the past, as can occur when, for example, signals are transported through the atmosphere so that the model state at some time is only observed later. This problem is often handled with the Kalman smoother.

Model and data error structures. The noise terms in the dynamic and observation models, Eqns (1) and (2), can in principle be quite general in form, including biases, drifts, temporal correlations, extreme outliers and so on. Many extant methods take these noise terms to be Gaussian with zero mean and no temporal correlation, as assumed in Eqns (4)–(10). However, more general error structures are very common, and the development of methods for dealing with such errors is an active area of current research. In the case of biases, drifts and temporal correlations, a promising approach is to introduce extra target variables to represent these features of the model or data error. Evensen (2003) showed how this approach can be used to treat both temporal correlation and bias in the model

error. Wang & Bentley (2002) introduced a target variable representing the temporally correlated part of the data error.

Nonsequential and sequential parameter estimation. Although parameter estimation is typically carried out with nonsequential strategies such as least-squares fitting, there can be advantages in using sequential methods such as Kalman filtering for parameter estimation. The approach is to treat parameters \mathbf{p} as components of the target vector \mathbf{y} (in addition to the state variables \mathbf{x}), with \mathbf{p} governed by the dynamic equation $d\mathbf{p}/dt = 0$ (+ noise) (Grewal & Andrews 1993). This means that the problem is almost always nonlinear and must be solved with the EKF or EnKF. Annan & Hargreaves (2004) show how this technique can be used to estimate parameters in the Lorenz system with chaotic dynamics. A potential advantage of this approach is that parameters can drift through time toward new values, in response to observations. This offers a means for model–data synthesis to respond to exogenous catastrophic events (such as fire, windthrow or clearing) which suddenly change the parameters in a terrestrial biosphere model, since exogenous changes in parameters are the usual way that catastrophic events are incorporated in the absence of a full dynamic model for the processes governing the catastrophe.

Model–data synthesis: examples

The methods outlined above are being applied in several fields relevant to terrestrial carbon observation. The first major example is parameter estimation. Most biogeochemical models contain parameters (\mathbf{p}) determining photosynthetic capacities, light use efficiencies, temperature and nutrient controls on photosynthesis and respiration, pool turnover times and so on. It is almost always necessary to choose \mathbf{p} to optimize the fit of the model to test data, usually obtained from multiple study sites. Techniques for doing this range from simple graphical fits ('chi-by-eye') to least-squares fitting procedures based on Eqn (3) or other cost functions.

A second example is provided by atmospheric inversion methods for inferring the surface–atmosphere fluxes of CO₂ and other trace gases from atmospheric composition observations. The data come from global flask networks and continuous *in situ* analysers, upper-air measurements from aircraft and tall towers, and potentially in the future from remote sensing of atmospheric composition. The observation model is a model of global atmospheric transport. The basic approach has been sketched in Eqns (4) and (5). There is now a significant literature on this technique (Enting

et al., 1995; Ciais & Meijer, 1998; Enting, 1999a,b; Rayner *et al.*, 1999; Rayner, 2001; Schimel *et al.*, 2001; Enting, 2002; Gurney *et al.*, 2002). In summary, atmospheric inversions at global scale provide good constraints on total global sources and sinks, but (presently) with very coarse spatial resolution (continental to hemispheric). In addition to global applications, atmospheric inversion methods have been applied regionally (Gloor *et al.*, 2001), in the atmospheric boundary layer (Lloyd *et al.*, 1996), and in vegetation canopies (Raupach, 2001).

A third example, combining the previous two, is the use of multiple constraints. This involves model–data synthesis with the simultaneous use of multiple kinds of observations (for example, atmospheric composition measurements, remote sensing, eddy-covariance fluxes, vegetation and soil stores, and hydrological data). This approach has two advantages: first, different kinds of observation constrain different processes. For example, atmospheric composition measurements and eddy fluxes directly determine net CO₂ exchanges (net ecosystem exchange, NEE) at large and small spatial scales, respectively, while remote sensing provides indirect constraints on gross exchanges (gross primary production, GPP) through indices such as the normalized difference vegetation index (NDVI). Second, different observations have different resolutions in space and time. Through assimilation into a terrestrial biosphere model, the high space–time resolution of environmental remote sensing can add space–time texture to estimates of NEE from methods such as atmospheric inversions or eddy-covariance fluxes. Some difficulties must also be noted: for example, handling data sources with quite different spatial and temporal scales of measurement (discussed further in 'Scale mismatches between measurements and models'), and also with very different sample numbers (remotely sensed data can swamp *in situ* data with realistic error specifications, because the former has a factor of 10³–10⁶ more data points).

Applications of the multiple-constraint concept include the combined use of atmospheric CO₂ concentrations and surface data at continental scales (Wang & Barrett, 2003; Wang & McGregor, 2003) and global scales (Kaminski *et al.*, 2001, 2002); use of genetic algorithms to constrain terrestrial ecosystem models of the global carbon cycle with multiple ecological data (Barrett, 2002); and discriminating vegetation and soil sources and sinks in forest canopies with concentration, isotopic and physiological data (Styles *et al.*, 2002).

A fourth example deserves more space than is available here: the use of data assimilation in atmospheric and ocean circulation models. This is now

well-developed and applied routinely in weather forecasting. A variety of techniques are employed, including ‘nudging’, three-dimensional and four-dimensional variational data assimilation (3DVAR and 4DVAR) based on adjoint methods, and use of the ensemble Kalman filter. Recent reviews are provided by Chen & Lamb (2000) and Park & Zupanski (2003).

Data characteristics: uncertainty in measurement and representation

We have emphasized that data uncertainties affect not only the predicted uncertainty of the eventual result of a model–data synthesis process, but also the predicted best estimate. This realization raises the challenge of evaluating the uncertainty properties of the main kinds of observation relevant to a TCOS, in forms directly usable for model–data synthesis. This is a very large goal, which embraces all categories of observation identified at the beginning of ‘Model–data synthesis: Methods’, and also a range of issues:

- The error magnitude σ_m for an observation z_m , inclusive of all error sources (in other words, the diagonal elements $[\text{Cov } \mathbf{z}]_{mmm} = \sigma_m^2$ of the covariance matrix);
- The correlations $[\text{Cov } \mathbf{z}]_{mn}/(\sigma_m\sigma_n)$ among errors in different observations, quantified by the off-diagonal elements of the covariance matrix;
- The temporal structure of the errors: whether they are random in time or temporally correlated, and the possible presence of unknown long-term drifts or biases;
- The spatial structure of errors (random, slowly varying or bias as for temporal structure);
- The error distribution: normal (Gaussian), log-normal, skewed or the sum of multiple error sources with different distributions, such as a small Gaussian noise together with occasional large outliers because of measurement corruption events;
- Possible mismatches between the spatial and temporal averaging implicit in the model and the measurements (the ‘scaling problem’);
- The separate contributions to all the above error properties of measurement error (the distribution of the measurements \mathbf{z} around their true values) and representation error (the distribution of the error in the model representation of the measurement, $\mathbf{z} = \mathbf{h}(\mathbf{y})$).

This challenge is too large to meet fully here. To make a start, we consider (in the next four subsections) a

selection of observations from four categories of data: remote sensing of land surface properties, atmospheric composition measurements, direct flux measurements, and direct measurements of carbon stores. The aim is to make estimates of error properties for these categories of measurement. The discussion does not address all of the above issues, largely omitting questions of spatial and temporal error structure. We present tables indicating ranges for the diagonal elements $[\text{Cov } \mathbf{z}]_{mmm} = \sigma_m^2$ of the error covariance matrix for measurement error, and the qualitative behaviour of the correlations which determine the off-diagonal elements. The entries in these tables are mostly ‘expert judgements’ by the authors and their colleagues, backed up by quantitative evidence where possible. There is, of course, no claim that our estimates are definitive; the intention is rather to indicate the kinds of uncertainty information required of observations for model–data synthesis purposes. The tables characterize measurement errors only; representation errors, which often exceed measurement errors, are discussed separately in qualitative terms only. The issue of scale mismatches between measurements and models, which arises in all cases as a significant contribution to representation error, is treated generically in a fifth subsection ‘Scale mismatches between measurements and models’.

Remote sensing of land surface properties

The main satellite-borne remotely sensed data on land surface properties come from two kinds of sensor, both polar-orbiting to provide frequent global coverage: moderate-spatial-resolution (~ 250 – 1000 m) and high-temporal-resolution (~ 1 day repeat interval) sensors such as AVHRR and MODIS; and high-spatial-resolution (~ 10 – 30 m) and moderate-temporal-resolution (~ 16 day repeat interval) sensors such as SPOT and LANDSAT. All these sensors provide multi-year records. One major application (among many) for the AVHRR-MODIS family is assessment of vegetation dynamics with indices such as NDVI (defined as $(\text{NIR} - \text{Red})/(\text{NIR} + \text{Red})$, where NIR and Red are radiances in the near-infrared and visible red spectral bands) and measures such as surface temperature. Applications for the SPOT-LANDSAT family include detection of land cover change and vegetation clearing and regrowth. In all cases, the measurements are at-sensor reflected radiances from the earth in several spectral bands (5 for AVHRR, 37 for MODIS).

In using these forms of remote sensing data for model–data synthesis applications, three kinds of error need to be considered: (1) errors associated with the measurement and spatial attribution of radiances at the sensor; (2) errors in relating radiances at sensor to

radiances at surface and (3) errors in relating radiances at the surface to biophysical quantities represented in a terrestrial biosphere model. The first of these is measurement error, the third is representation error, and the second can contribute to either depending on how the problem is formulated.

Measurement error. The primary measurements are radiances at the sensor, attributed to elements (pixels) on the earth's surface. Measurement errors arise from sensor noise, calibration drift, orbital decay, and incorrect geolocation. These errors are much more serious for AVHRR (designed in the 1970s) than for MODIS (designed in the 1990s with on-board calibration).

Representation error. This is the error associated with the model $\mathbf{z} = \mathbf{h}(\mathbf{y})$ relating the measurements \mathbf{z} (radiances) to the target variables \mathbf{y} (biophysical variables in the terrestrial biosphere model, such as leaf chlorophyll content, leaf or soil water status or leaf area index). In principle, the observation model involves two components. First, biophysical target variables (\mathbf{y}) are related to radiances at the earth's surface (say $\mathbf{z}_{\text{surface}}$), through a model $\mathbf{z}_{\text{surface}} = \mathbf{h}_{\text{surface}}(\mathbf{y})$. Examples are relationships between NDVI and leaf and soil properties (Tucker, 1979; Sellers, 1985; Sellers *et al.*, 1992; Myneni *et al.*, 1995a,b; Lu *et al.*, 2003). Representation errors associated with this component depend on the skill of the model $\mathbf{z}_{\text{surface}} = \mathbf{h}_{\text{surface}}(\mathbf{y})$. Second, the surface radiative properties are related to radiances at the satellite-borne sensor (say $\mathbf{z}_{\text{sensor}}$) through an atmospheric radiative transfer model (\mathbf{g}), which accounts for the effects of clouds, atmospheric absorption and scattering, and the bidirectional reflectance distribution function (BRDF) properties of the surface. Thus, we have $\mathbf{z}_{\text{sensor}} = \mathbf{g}(\mathbf{z}_{\text{surface}}, \text{ancillary data})$, where the ancillary data includes profiles of temperature and radiatively active constituents in the atmosphere. Errors in this component depend on the skill of the model \mathbf{g} and errors in the required ancillary data.

In model-data synthesis, there is a choice about whether to use sensor radiances ($\mathbf{z}_{\text{sensor}}$) or surface radiances ($\mathbf{z}_{\text{surface}}$) as the primary measurements. If $\mathbf{z}_{\text{surface}}$ is used, then the observation model is $\mathbf{z}_{\text{surface}} = \mathbf{h}_{\text{surface}}(\mathbf{y})$ and it is necessary to infer $\mathbf{z}_{\text{surface}} = \mathbf{g}^{-1}(\mathbf{z}_{\text{sensor}}, \text{ancillary data})$ from the at-sensor radiances and the atmospheric transfer model \mathbf{g} . In this case, \mathbf{g} is effectively part of the measurement system itself and its errors appear as measurement errors. On the other hand, if $\mathbf{z}_{\text{sensor}}$ is the primary set of measurements, then the observation model becomes $\mathbf{z}_{\text{sensor}} = \mathbf{h}_{\text{sensor}}(\mathbf{y}) = \mathbf{g}(\mathbf{h}_{\text{surface}}(\mathbf{y}), \text{ancillary data})$. The

model \mathbf{g} is then part of the observation model, and its errors appear as representation errors. New unknown target variables may also appear through the ancillary data, and it is necessary that these be estimated with available observations.

For a semi-quantitative illustration of properties of the error covariance matrix for some remote sensing observations, we consider radiances at the surface ($\mathbf{z}_{\text{surface}}$) to be the observations, so that errors in the radiative transfer model \mathbf{g} form part of the errors in $\mathbf{z}_{\text{surface}}$ and are treated as measurement error. As examples we consider NDVI and surface temperature. Table 1 shows some properties of measurement errors in these quantities. This table takes account of several considerations: first, errors in $\mathbf{z}_{\text{surface}}$ arise from sensor noise, sensor calibration and geolocation (contributing to error in $\mathbf{z}_{\text{sensor}}$), and from inaccurate cloud removal, atmospheric correction or BRDF correction (contributing to error in the radiative transfer model \mathbf{g}). Many of these errors are strongly positively correlated among different spectral bands, especially in the visible and NIR. NDVI is less sensitive to such errors than the band radiances themselves, because it is based on a normalized difference (an example of error cancellation through correlation, and a reason for the popularity of NDVI). Second, correlations between errors in NDVI and surface temperature tend to be positive. For example, unmasked sub-pixel cloud appears as low-NDVI and cold, causing correlated negative outliers and thus a positive error correlation between the two measurements. Finally, in some cases, experimental estimates of the errors are available. For example, comparison of several calibration procedures for the short-wave AVHRR channels on the NOAA-11 satellite revealed calibration discrepancies of around 5% (Mitchell *et al.*, 1996).

Table 1 describes measurement error only, and does not include representation error reflecting uncertainty in the relationship $\mathbf{z}_{\text{surface}} = \mathbf{h}_{\text{surface}}(\mathbf{y})$ (the third of the three error categories mentioned above). Estimates of representation error can be gained, for example, from the scatter in experimental tests of relationships between at-surface radiance properties and biophysical variables (for example, Lu *et al.*, 2003). Usually, these representation errors are comparable with or larger than measurement errors. For example, NDVI saturates at high leaf area index (3-4).

Atmospheric composition measurements

We consider (1) direct CO₂ concentration measurements from the global flask network and WMO Global Atmosphere Watch (http://www.wmo.ch/web/arep/gaw/gaw_home.html) stations with continuous CO₂ monitoring and (2) efforts to measure CO₂ with

Table 1 Indicative properties of the error covariance matrix for remote sensing observations of NDVI and surface temperature, describing measurement error only and omitting representation error

Observation (z_m)	Units	Typical range	Typical error (σ_m)	Error distribution	Error correlations
NDVI = (NIR–Red)/ (NIR + Red)	–	0.1–1	0.1 (AVHRR) 0.05 (MODIS)	Normal with negative outliers because of undetected cloud	Errors in NDVI and surface temperature are probably highly positively correlated because of negative outliers in each associated with undetected cloud
Surface temperature	degree K	250–350	Land: 2 (AVHRR) 1 (MODIS) Ocean: 0.2 (AVHRR) 0.1 (MODIS)		

NDVI, normalized difference vegetation index.

spaceborne remote sensing. Our discussion of measurement and representation errors for these data sources is based on the atmospheric inversion approach for utilizing atmospheric composition data; see ‘Model-data synthesis: methods and examples’.

Direct measurements. The first global flask network developed from work by Keeling (1961) with several nations later commencing sampling in their respective regions. As of 2004, 23 laboratories from 15 nations contribute flask and continuous *in situ* sampling data from around 200 sites to international databases (<http://www.cmdl.noaa.gov/ccgg/globalview/co2/>). Large networks are maintained by the National Oceanic and Atmosphere Administration Climate Monitoring and Diagnostics Laboratory (NOAA-CMDL), Boulder, CO, USA, and CSIRO Atmospheric Research (CSIRO-GasLab), Melbourne, Australia. In the main, flask samples are collected 2–4 times per month from fixed sites, aircraft and ships, and opportunistically during intensive experimental campaigns. Flasks are returned to the central laboratories for automated analyses, which provide concentrations of a number of atmospheric constituents. The analytical procedure includes measurements against an international standard and frequent comparisons with air standards, and actual samples, exchanged between laboratories (Masarie *et al.*, 2001).

Continuous *in situ* analysers provide records at higher precisions and temporal densities than are available from the flask networks and are usually located at baseline stations in remote locations. With both flask and *in situ* monitoring, there has been ongoing improvement in precision and in the accuracy and propagation of the international standards (Francey *et al.*, 2001).

Spaceborne measurements. Remote sensing of atmospheric composition is gradually becoming a

reality. Simulations show that satellite observations improve atmospheric-inverse estimates of carbon fluxes by a factor of up to ten relative to the surface (flask and baseline) networks, because of vastly improved coverage in time and space, albeit with increased error for each measurement (Rayner & O’Brien, 2001; Rayner *et al.*, 2002). Efforts are underway on three fronts: first, proof-of-concept studies have already been undertaken with existing space-based radiometers such as TOVS and AVHRR (Chedin *et al.*, 2002, 2003a,b). Despite limited spectral resolution and unwanted absorption from other atmospheric constituents, signatures of seasonal cycles and trends in CO₂ and other greenhouse gases (N₂O and CO) have been extracted at coarse space-time resolution (15° × 15°, mid-troposphere, monthly). Second, current missions undertaken for other purposes, such as the Advanced Infra-red Sounder (AIRS) and Sciamachy instruments, will likely provide near-term improvements in measuring aspects of the atmospheric CO₂ distribution. Finally, future purpose-built instruments, such as the NASA Orbiting Carbon Observatory (OCO, Crisp *et al.*, 2004), CARBOSAT (European Space Agency) and GOSAT (Japan), should provide dramatic improvements in coverage and precision.

Measurement error. Table 2 gives estimates of measurement error for CO₂ measurements from flasks, continuous *in situ* analysers, and AVHRR (Chedin, 2003b). Errors in international CO₂ databases from near-surface measurements include errors in the assignment and propagation of CO₂-in-air standards on the WMO mole fraction scale, biases associated with different CO₂ measurement methods, flask storage effects, and other factors limiting the repeatability of measurements with the same system. These errors have different temporal characteristics. For example, both flask and continuous measurements from CSIRO share a current calibration bias of -0.1 ± 0.05 ppm

relative to NOAA-CMDL. For some laboratories with smaller networks, WMO round-robin intercomparisons in 1993, 1997 and 2001 indicate a larger, but improving range of calibration uncertainty. There are also differences between flask and *in situ* measurements from within one laboratory that exhibit slowly varying offsets. Such errors are likely to be highly correlated over long time periods. If not included in the inversion procedure they can invalidate aspects of the result, as demonstrated by Rayner *et al.* (2002) in their Observing System Simulation Experiment of remotely-sensed CO₂. Provided the structure of such errors can be predicted, the inversion can cope with them (Law *et al.*, 2003a, b).

Representation error. Current atmospheric inversion studies using atmospheric concentration data generally ascribe much larger errors to the measurements than the aggregate of the measurement errors shown in Table 2. This is partly because of the simple assumption of Gaussian uncorrelated noise used in most existing studies, but also because the contributions of the representation error are large. Representation errors arise in this context from the inability of the atmospheric transport model to simulate point observations in space, either because of systematic errors in model formulation or the implicit averaging in its grid representation. An analogous temporal representation error arises when flask measurements (actually grab samples in time) are interpreted as longer-term means; see 'Scale mismatches between measurements and models' for discussion. A further contribution to representation errors for most atmospheric inversion studies to date has been the projection of possible source distributions to a restricted subspace, usually by dividing the earth into a number of large regions. This is done both for computational reasons and to reduce the error amplification arising from under-determined problems. Errors in the prescription of flux distributions within these regions give rise to a so-called aggregation error, described and quantified by Kaminski *et al.* (2001). This error can be avoided by using adjoint representations of atmospheric transport that do not require aggregation (Rodenbeck *et al.*, 2003a, b).

There are few experiments where representation errors can be evaluated, since this requires simultaneous knowledge of sources and atmospheric transport. However, one can use the range of model simulations as a guide (e.g. Law *et al.*, 1996; Gurney *et al.*, 2003). Since representation errors are completely dependent on the inversion process (especially the atmospheric transport model) rather than the measurements themselves, we do not attempt to quantify them in Table 2.

Table 2 Indicative properties of the error covariance matrix for observations of atmospheric CO₂ concentrations, describing measurement error only and omitting representation error

Observation (z_m)	Units	Typical range	Typical error (σ_m)	Error distribution	Error correlations
[CO ₂] from flask	ppmv	350–500	0.01–1	Normal with slowly varying bias	Measurement errors: correlated offsets which vary over months to years
[CO ₂] from baseline station	ppmv	350–500	0.05 (CSIRO-GasLab, NOAA-CMDL) 0.5 (all laboratories)		
[CO ₂] from AVHRR (Chedin <i>et al.</i> 2003b) (15° × 15°, mid-troposphere)	ppmv	360–400	4	Normal?	Uncorrelated between different observations in space and time?

Queries indicate speculation.

Direct flux measurements at intensive study sites

The global FluxNet network (Baldocchi *et al.*, 2001, Falge *et al.*, 2002a,b) includes over 200 sites where tower-based eddy covariance measurements are made of the land-air fluxes of sensible heat, latent heat, CO₂ and other entities, at sub-diurnal (typically half-hourly) temporal resolution. Many other meteorological variables are measured, including solar and net radiation, ground heat flux and precipitation. Many sites include measurements of other biogeochemical processes, such as soil respiration by chamber methods. Most measurements at these sites are at the patch scale, applying to a nominally homogeneous land unit. The extent of horizontal averaging varies with the measurement (for example, a narrow along-wind ellipse of order $1 \times 0.1 \text{ km}^2$ for eddy-covariance fluxes, a circle with radius of order 10 m for a net radiometer mounted 10 m above a forest, and a chamber footprint of order 1 m^2 for soil respiration). A great strength of intensive flux study sites is that independent checks on the uncertainty of the measurements are possible through checks on the closure of the energy and water balances.

Measurement error. Table 3 shows some estimated error properties for the above measurements. Several systematic errors are known to influence eddy-flux measurements, including high-frequency and low-frequency flux losses, storage in the air column below the measurement height, intermittency of nocturnal turbulence, and ‘rectification’ because of nocturnal cold-air drainage of air rich in respired CO₂ towards low-lying areas, causing systematic advection errors (Wofsy *et al.*, 1993). Most of these problems are much more severe at night than by day. Experimental techniques and analysis procedures (especially the use of longer averaging periods of up to several hours to include fluxes transported by large eddies of the scale of the entire atmospheric boundary layer) are progressively resolving these issues, with more success to date for daytime than nocturnal fluxes (Finnigan *et al.*, 2003). Relative hourly accuracies of order 10% represent the state of the art for daytime fluxes, attainable with excellent technique; however, errors of up to 30% are common. For nocturnal fluxes, eddy covariance measurements become unreliable in light wind conditions, that is, much of the time at night (Goulden *et al.*, 1996). Errors include biases associated mainly with the above systematic nocturnal effects.

Carbon flux measurements from intensive study sites are often aggregated to time-averaged (typically monthly to multi-annual) measures of primary production for the patch under study: GPP = [net assimilation]; net primary productivity (NPP) =

[GPP–autotrophic respiration]; net ecosystem productivity (NEP) = [NPP–heterotrophic respiration]; net biome productivity (NBP) = [NEP–disturbance flux]. All these have units $\text{gC m}^{-2} \text{yr}^{-1}$ and are positive for carbon uptake into the biosphere. The total exchange between the terrestrial biosphere and the atmosphere is NBP. The disturbance fluxes in NBP include grazing, harvest, and catastrophic events (fire, windthrow, clearing). If these processes do not occur on an intensive study site, an eddy-covariance CO₂ flux (when aggregated in time) yields an estimate of NEP. Table 3 also shows estimated properties of error covariances for these time-aggregated productivity estimates. The uncertainty in NEP estimates is much higher than for hourly CO₂ flux estimates, because of the difficulties mentioned above. Comparisons between eddy-covariance and other means of measuring NEP suggest typical uncertainties of order 20 to 50%, with the lower end of this range being attainable over actively growing ecosystems where NPP is much larger than heterotrophic respiration and NEP is relatively large. This range is also likely to be representative of GPP and NPP estimates. Uncertainties in NBP at patch scale are very much higher (in fact it is arguable that NBP cannot be defined at patch scale).

Representation error. Fluxes are almost always directly represented in terrestrial biosphere models, so representation errors are not an issue for direct flux measurements at the patch scale. However, the spatial aggregation issue (upscaling flux measurements from intensive study sites to yield estimates of fluxes or productivities over grid cells, large regions or continents) is a significant source of representation error. It is not usually possible to upscale by area weighting of patch-scale estimates without conditioning from other measurements and models, because of heterogeneity in landscapes and disturbance patterns; see ‘Scale mismatches between measurements and models’ for further discussion. A particular example of this issue is that patch-scale measurements (including fluxes) are often made at unmanaged or undisturbed sites, creating a bias with respect to the landscape as a whole.

Measurements of carbon stores in vegetation and soils

Terrestrial biospheric carbon pools include leaves, wood above and below ground, fine roots, coarse litter, fine litter and soil carbon; the soil carbon is often partitioned into fractions with different biochemical properties or turnover times, such as microbial, humic and inert. *In situ* measurements or estimates of these pools are available from vegetation and soil surveys,

Table 3 Indicative properties of the error covariance matrix for some direct flux measurements, describing measurement error only and omitting representation error

Observation (z_m)	Units	Typical range	Typical relative error (σ_m/z_m) (%)	Error distribution	Error correlations
Hourly net irradiance	$W m^{-2}$	0–600	5–15	Normal	Errors in radiation, eddy covariance and soil respiration measurements on hourly time scales are uncorrelated to first approximation. However, errors in hourly eddy covariance measurements of sensible heat, latent heat and CO_2 fluxes are highly correlated.
Hourly sensible heat flux (from eddy covariance)	$W m^{-2}$	0–600	10–30	Normal	
Hourly latent heat flux (from eddy covariance)	$W m^{-2}$	0–600	10–30	Normal	
Hourly CO_2 flux (from eddy covariance)	$gC m^{-2} d^{-1}$	0–10 (+ down)	20–50	Normal with nocturnal bias	
Hourly soil-respired CO_2 flux	$gC m^{-2} d^{-1}$	0–5 (+ up)	20–50	Normal	
Annual GPP	$gC m^{-2} yr^{-1}$	0–4000 (+ down)	20–50	Normal with bias	
Annual NPP	$gC m^{-2} yr^{-1}$	0 to 2000 (+ down)	20–50	Normal with bias	
Annual NEE	$gC m^{-2} yr^{-1}$	–500–1500 (+ down)	20–50	Normal with bias	
Long-term average NBP	$gC m^{-2} yr^{-1}$	–100–100 (+ down)	$\gg 100$	Normal with bias	

GPP, gross primary production; NPP, net primary productivity; NEE, net ecosystem exchange; NBP, net biome productivity.

Table 4 Indicative properties of the error covariance matrix for some measurements of carbon stores in vegetation and soils, describing measurement error only and omitting representation error

Observation (z_m)	Units	Typical range	Typical relative error (σ_m/z_m) (%)	Error distribution	Error correlations
Leaf carbon	kgC m ⁻²	0–1	10–30	Log-normal (normal if many measurements are aggregated)	At an individual site, errors are uncorrelated because measurements are made with different techniques
Wood carbon (above and below ground)	kgC m ⁻²	0–50	15–50		
Fine root carbon	kgC m ⁻²	0–1	30–100		
Coarse litter (including standing dead)	kgC m ⁻²	0–10	30–100		
Fine litter	kgC m ⁻²	0–0.5	10–30		
Soil carbon (to 1 m)	kgC m ⁻²	0 to >100	30–100		

long-term observations at ecological study sites, and national forest inventories.

Measurement errors. Table 4 shows some estimated error properties for these measurements. Large errors are associated with the below-ground pools. As for our error estimates for flux measurements at intensive study sites (Table 3), these estimates apply to measurement errors at the spatial scale of individual patches.

Representation errors. At first sight, representation errors for measurements of carbon stores are not a major issue at patch scales because carbon stores are directly represented in terrestrial biosphere models (as for flux measurements). However, representation errors arise in assigning measurements of carbon pools to their model counterparts (Barrett, 2002). For instance, litter is often measured in physically defined compartments such as fine litter and coarse woody debris, but represented in models in biogeochemical compartments such as metabolic and structural litter. Spatial aggregation of measurements for upscaling and incorporation into spatially coarse models is also a representation issue, as for direct flux measurements—for example, the bias introduced by the tendency for patch-scale measurements to be made at undisturbed sites.

Scale mismatches between measurements and models

The examples in the above four subsections all highlight the problem that measurements are almost always made at a different scale (that is, with a different spatial and temporal averaging operator) to that used in the dynamic model. In its spatial form, this problem arises when the measurements apply to particular points or homogeneous land elements (for example, eddy flux measurements or *in situ* measurements of carbon stores) while the model represents an aggregated set

of land elements by a single set of state variables (for example, for a grid cell in a coarse spatial model). A temporal version of the same problem arises when the measurements are intermittent (for example, satellite measurements at time of overpass, flask sampling of atmospheric composition, or occasional *in situ* measurements of carbon stores).

This ‘scaling problem’ is so prevalent that it is worth a brief discussion in the context of model–data synthesis. Suppose that the model is defined at ‘coarse’ space-time scales by physical parameterizations that demand some (usually implicit) space-time averaging of all variables, and by choices of space-time grids, which are compatible with these averaging requirements. Since it is necessary that all variables in both the dynamic and observation models be averaged in a consistent way throughout, the observation model can be written as $\mathbf{z}_{\text{coarse}} = \mathbf{h}(\mathbf{y}_{\text{coarse}}) + \text{noise}$, where the measurements (\mathbf{z}) and target variables (\mathbf{y}) are identified as applying at coarse (model) scale. However, the actual measurements are often available at some much finer scale \mathbf{z}_{fine} , and may also be sparsely sampled. The variability in an ensemble of measurements \mathbf{z}_{fine} (which cover the coarse scale) is defined by a covariance \mathbf{R}_{fine} , the integral of the cospectrum of \mathbf{z} between the fine and coarse scales. The problem is to find a relationship between \mathbf{z}_{fine} and $\mathbf{z}_{\text{coarse}}$ so that \mathbf{z}_{fine} can be used as an observable in the model–data synthesis process. There are several generic ways to do this.

A first, very simple option is to take the fine scale measurements (\mathbf{z}_{fine}) as a noisy sample of $\mathbf{z}_{\text{coarse}}$, with the variability in \mathbf{z}_{fine} treated as a contribution to the representation error. In this case, the observation model becomes $\mathbf{z}_{\text{fine}} = \mathbf{h}(\mathbf{y}_{\text{coarse}}) + [\text{measurement error in } \mathbf{z}_{\text{fine}}] + [\text{representation error in } \mathbf{h}] + [\text{noise with covariance } \mathbf{R}_{\text{fine}}]$, where the last noise term accounts for unresolved space-time variability in \mathbf{z}_{fine} . This is often the best option if there is no other information

available about the sources of variability in \mathbf{z}_{fine} , as for example with flask measurements of atmospheric composition or soil samples taken from imprecisely known locations.

A second option is that the fine scale measurements (\mathbf{z}_{fine}) can be aggregated directly to find $\mathbf{z}_{\text{coarse}} = \Sigma w(\mathbf{a}_{\text{fine}})\mathbf{z}_{\text{fine}}$, where w is a set of weights determined by ancillary fine-scale data \mathbf{a}_{fine} . For example, reflected radiances can be aggregated with area weighting, and *in situ* soil or biomass carbon measurements can be aggregated approximately using a weighting based on soil and microclimate attributes that can be inferred from fine-scale digital elevation maps. The inferred $\mathbf{z}_{\text{coarse}}$ becomes a pseudo-observation to be utilized through an observation model $\mathbf{z}_{\text{coarse}} = \mathbf{h}(\mathbf{y}_{\text{coarse}}) + \text{noise}$, where the noise term includes errors in \mathbf{z}_{fine} , errors in the weighting model for $\mathbf{z}_{\text{coarse}}$, and errors in \mathbf{a}_{fine} . This is a good option when plenty of fine-scale measurements are available within each coarse-scale element, for example from remote sensing.

A third option is to introduce a fine-scale process or empirical model $\mathbf{z}_{\text{fine}} = \mathbf{g}(\mathbf{x}_{\text{coarse}}, \mathbf{a}_{\text{fine}})$ which relates the fine-scale observations to coarse-scale state variables (which here act as boundary conditions) and additional fine-scale ancillary data such as topography and land surface attributes from remote sensing. The model \mathbf{g} then becomes the observation model for model–data synthesis. This is often the best option if the fine-scale data are sparse, expensive and of high information content. For example, for eddy flux measurements, \mathbf{g} may be a fine-scale model of land–air exchanges and microclimates in inhomogeneous or hilly terrain, or a neural-network-based model of flux distributions conditioned with available ancillary data \mathbf{a}_{fine} . It is likely that the development of \mathbf{g} involves substantial effort, so this is not an option to be taken lightly.

This brief discussion has focussed on the observational issue of scale mismatches between measurements and models, and their implications for the observation model $\mathbf{z} = \mathbf{h}(\mathbf{y}) + \text{noise}$, Eqn (2). Another form of the ‘scaling problem’ arises for the dynamic model $\text{d}\mathbf{x}/\text{d}t = \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p}) + \text{noise}$, Eqn (1), in translating process knowledge about the model function \mathbf{f} between scales. Typically \mathbf{f} is a set of phenomenological equations for fluxes contributing to changes in stores \mathbf{x} , which are valid only for certain (often implicit) spatial and temporal resolutions. For example, the Darcy law describes water movement in soil columns but not catchments or heterogeneous regions described by aggregated state variables. Translation of these phenomenological equations from fine to coarse scales is possible by treating the fine-scale variability statistically, but one result is that the fine-scale and coarse-scale equations are different (for instance, biased with

respect to each other) because of interactions between fine-scale variability and nonlinearity in the fine-scale function $\mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{p})$.

Summary and conclusions

In the context of terrestrial carbon observation, we have focussed on model–data synthesis and its implications for data, especially the specification of data uncertainty. Our analysis has been framed by an initial statement of the purposes and attributes of a TCOS. The purposes for a TCOS are congruent with those of a ‘Global Earth Observation System of Systems’, with specific contributions occurring in the areas of weather and climate prediction, water resource management, ecosystem management, agricultural sustainability, combating desertification and monitoring biodiversity. Among the major attributes of a TCOS are scientific credibility, consistency with global budgets, adequate spatial and temporal resolution, observation of sufficient ranges of entities and processes, and the requirement to quantify uncertainty. These attributes demand a model–data synthesis approach because of the need to combine a range of observations and models to determine the terrestrial stores and fluxes of carbon and related entities (water, nutrients, energy), and the ways that they are influenced by human management.

Data for model–data synthesis approaches come in two forms, observations and prior knowledge (for instance constraints on model parameters). For both, uncertainty estimates have an influence on the outcome of the synthesis process comparable with that of the data values themselves. Data uncertainties affect not only the predicted uncertainty of the eventual result, but also the predicted best estimate. Therefore, there is an urgent need for soundly based uncertainty specification, based initially on an error covariance matrix.

In this paper, we have made semi-quantitative estimates of some of the main properties of the covariance matrix for measurement error, for four kinds of data central to terrestrial carbon observation: remote sensing of land surface properties, atmospheric composition measurements, direct flux measurements, and measurements of carbon stores. Critical error properties include (1) the diagonal elements $[\text{Cov } \mathbf{z}]_{mm} = \sigma_m^2$ of the measurement error covariance matrix (where σ_m is the error magnitude for an observation z_m); (2) the correlations between different observations, quantified by the off-diagonal elements of the covariance matrix; (3) the temporal and (4) the spatial structure of errors, (5) the error distribution, (6) possible scale mismatches between measurements and models and (7) the representation of the observations in the model.

Finally, we note that there is a need not only to quantify uncertainty, but also to reduce uncertainty in our estimates. The approach outlined here contributes to this goal, by providing a framework to formalize the manner in which we constrain uncertainty. The critical step is to better understand the error structures of both the priors and the observations, leading to improved focus on the major sources of uncertainty.

Acknowledgments

This paper is the result of a workshop held in Sheffield, UK, 3–6 June 2003, organized jointly by the Terrestrial Carbon Observation (TCO) initiative of the Global Terrestrial Observing System (GTOS), and the Global Carbon Project (GCP; www.globalcarbonproject.org). We thank GTOS/TCO, NASA, IGBP, and the GCP Canberra Office (funded by Australian Greenhouse Science Program) for financial and organizational support, and all workshop participants for their contributions and enthusiasm. We thank several colleagues, particularly John Norton, John Finnigan and Roger Francey, for comments on a draft of the paper. This paper contributes to the implementation of the GCP research agenda on ‘Development of model–data fusion techniques’ (Activity 1.2), as GCP Publication No.8.

References

- Annan JD, Hargreaves JC (2004) Efficient parameter estimation for a highly chaotic system. *Tellus*, **56A**, 520–526.
- Baldocchi D, Falge E, Gu LH *et al.* (2001) FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, **82**, 2415–2434.
- Barrett DJ (2002) Steady state turnover time of carbon in the Australian terrestrial biosphere. *Global Biogeochemical Cycles*, **16**, 1108, doi:10.1029/2002GB001860.
- Chedin A, Serrar S, Armante R *et al.* (2002) Signatures of annual and seasonal variations of CO₂ and other greenhouse gases from comparisons between NOAA TOVS observations and radiation model simulations. *Journal of Climate*, **15**, 95–116.
- Chedin A, Saunders R, Hollingsworth A *et al.* (2003a) The feasibility of monitoring CO₂ from high-resolution infrared sounders. *Journal of Geophysical Research-Atmospheres*, **108**, 4064, doi:10.1029/2001JD001443.
- Chedin A, Serrar S, Scott NA *et al.* (2003b) First global measurement of midtropospheric CO₂ from NOAA polar satellites: TROPICAL zone. *Journal of Geophysical Research-Atmospheres*, **108**, 4581, doi:10.1029/2003JD003439.
- Chen CR, Lamb PJ (2000) Improved treatment of surface evapotranspiration in a mesoscale numerical model part II: via the assimilation of satellite measurements. *Terrestrial Atmospheric and Oceanic Sciences*, **11**, 789–832.
- Ciais P, Meijer HAJ (1998) The 18O/16O isotope ratio of atmospheric CO₂ and its role in global carbon cycle research. In: *Stable Isotopes: Integration of Biological, Ecological and Geochemical Processes* (ed. Griffiths H), pp. 409–431. Bios Scientific Publishers Ltd., Oxford.
- Ciais P, Moore BI, Steffen WL *et al.* (2004) *Integrated Global Carbon Observation Theme: A Strategy to Realise a Coordinated System of Integrated Global Carbon Cycle Observations*. Integrated Global Observing Strategy, Stockholm.
- Cihlar J, Denning S, eds (2002b) *Terrestrial Carbon Observation: The Frascati Report on In Situ Carbon Data and Information*. Environment and Natural Resources Series 5, Food and Agriculture Organization, Rome.
- Cihlar J, Denning S, eds (2002c) *Terrestrial Carbon Observation: The Rio de Janeiro Recommendations for Terrestrial And Atmospheric Measurements*. Environment and Natural Resources Series 3, Food and Agriculture Organization, Rome.
- Cihlar J, Denning AS, Gosz J, eds (2002a) *Terrestrial Carbon Observation: The Ottawa assessment of requirements, status and next steps*. Environment and Natural Resources Series 2, Food and Agriculture Organization, Rome.
- Crisp D, Atlas RM, Breon F-M *et al.* (2004) *The orbiting carbon observatory (OCO) mission*. http://oco.jpl.nasa.gov/pubs/OCO_Cospar_2002_final.pdf
- Drécourt J-P (2003) *Kalman filtering in hydrological modelling*. DAIHM Technical Report 2003-1, DHI Water & Environment, Hørsholm, Denmark.
- Enting IG (1999a) *Characterising the temporal validity of the global carbon cycle*. Technical Paper No. 40, CSIRO Atmospheric Research, Melbourne, Australia.
- Enting IG (1999b) Green’s function methods of tracer inversions. In: *Inverse Methods in Global Biogeochemical Cycles* (ed. Kasibhatla P), American Geophysical Union, Washington, DC, pp. 19–31.
- Enting IG (2002) *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press, Cambridge.
- Enting IG, Trudinger CM, Francey RJ (1995) A synthesis inversion of the concentration and delta-C-13 of atmospheric CO₂. *Tellus Series B-Chemical and Physical Meteorology*, **47**, 35–52.
- Evans SN, Stark PB (2002) Inverse problems as statistics. *Inverse Problems*, **18**, R55–R97.
- Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *Journal of Geophysical Research-Oceans*, **99**, 10143–10162.
- Evensen G (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.
- Falge E, Baldocchi D, Tenhunen J *et al.* (2002a) Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. *Agricultural and Forest Meteorology*, **113**, 53–74.
- Falge E, Tenhunen J, Baldocchi D *et al.* (2002b) Phase and amplitude of ecosystem carbon release and uptake potentials as derived from FLUXNET measurements. *Agricultural and Forest Meteorology*, **113**, 75–95.
- Field CB, Raupach MR, eds (2004) *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World*. Island Press, Washington.
- Finnigan JJ, Clement R, Malhi Y *et al.* (2003) A re-evaluation of long-term flux measurement techniques—Part I: averaging and coordinate rotation. *Boundary-Layer Meteorology*, **107**, 1–48.

- Francey RJ, Rayner PJ, Allison CE (2001) Constraining the global carbon budget from global to regional scales—the measurement challenge. In: *Global Biogeochemical Cycles in the Climate System* (eds Schulze E-D, Heimann M, Harrison S, Holland E, Lloyd J, Prentice IC, Schimel DS), pp. 245–252. Academic Press, San Diego.
- Gershensfeld NA (1999) *The Nature of Mathematical Modeling*. Cambridge University Press, Cambridge.
- Giering R (2000) Tangent linear and adjoint biogeochemical models. In: *Inverse Methods in Global Biogeochemical Cycles* (eds Kasibhatla P, Heimann M, Rayner P, Mahowald N, Prinn RG, Hartley DE), pp. 33–48. American Geophysical Union, Washington, DC.
- Global Carbon Project (2003) *Science Framework and Implementation*, Earth System Science Partnership (IGBP, IHDP, WCRP, Diversitas) Report No. 1; GCP Report No. 1, Global Carbon Project, Canberra.
- Gloor M, Bakwin P, Hurst D *et al.* (2001) What is the concentration footprint of a tall tower? *Journal of Geophysical Research—Atmospheres*, **106**, 17831–17840.
- Goulden ML, Munger JW, Fan SM *et al.* (1996) Exchange of carbon dioxide by a deciduous forest: response to interannual climate variability. *Science*, **271**, 1576–1578.
- Grewal MS, Andrews AP (1993) *Kalman Filtering: Theory and Practice*. Prentice-Hall, Englewood Cliffs, NJ.
- Gurney KR, Law RM, Denning AS *et al.* (2002) Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature*, **415**, 626–630.
- Gurney KR, Law RM, Denning AS (2003) TransCom 3 CO₂ inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information. *Tellus Series B—Chemical and Physical Meteorology*, **55**, 555–579.
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *ASME Journal of Basic Engineering, Series D*, **82**, 34–45.
- Kaminski T, Knorr W, Rayner PJ *et al.* (2002) Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. *Global Biogeochemical Cycles*, **16**, doi:10.1029/2001GB001463.
- Kaminski T, Rayner PJ, Heimann M *et al.* (2001) On aggregation errors in atmospheric transport inversions. *Journal of Geophysical Research—Atmospheres*, **106**, 4703–4715.
- Kasibhatla P, Heimann M, Rayner PJ *et al.*, eds (2000) *Inverse Methods in Global Biogeochemical Cycles*. American Geophysical Union, Washington DC.
- Keeling CD (1961) The concentration and isotopic abundance of carbon dioxide in rural and marine air. *Geochimica et Cosmochimica Acta*, **24**, 277–298.
- Law RM, Chen YH, Gurney KR (2003a) TransCom 3 CO₂ inversion intercomparison: 2. Sensitivity of annual mean results to data choices. *Tellus Series B—Chemical and Physical Meteorology*, **55**, 580–595.
- Law RM, Rayner PJ, Denning AS *et al.* (1996) Variations in modeled atmospheric transport of carbon dioxide and the consequences for CO₂ inversions. *Global Biogeochemical Cycles*, **10**, 783–796.
- Law RM, Rayner PJ, Steele LP *et al.* (2003b) Data and modelling requirements for CO₂ inversions using high-frequency data. *Tellus Series B—Chemical and Physical Meteorology*, **55**, 512–521.
- Le Dimet F-X, Talagrand O (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus Series A—Dynamic Meteorology and Oceanography*, **38**, 97–110.
- Lloyd J, Kruijt B, Hollinger DY *et al.* (1996) Vegetation effects on the isotopic composition of atmospheric CO₂ at local and regional scales: theoretical aspects and a comparison between rain forest in Amazonia and a Boreal Forest in Siberia. *Australian Journal of Plant Physiology*, **23**, 371–399.
- Lu H, Raupach MR, McVicar TR *et al.* (2003) Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment*, **86**, 1–18.
- Masarie KA, Langenfelds RL, Allison CE *et al.* (2001) NOAA/CSIRO flask air intercomparison experiment: A strategy for directly assessing consistency among atmospheric measurements made by independent laboratories. *Journal of Geophysical Research—Atmospheres*, **106**, 20445–20464.
- Mitchell RM, O'Brien DM, Forgan BW (1996) Calibration of the AVHRR shortwave channels. 2. Application to NOAA 11 during early 1991. *Remote Sensing of Environment*, **55**, 139–152.
- Myneni RB, Hall FG, Sellers PJ *et al.* (1995a) The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, **33**, 481–486.
- Myneni RB, Maggion S, Jaquinto J *et al.* (1995b) Optical remote-sensing of vegetation—modeling, caveats, and algorithms. *Remote Sensing of Environment*, **51**, 169–188.
- Park SK, Zupanski A (2003) Four-dimensional variational data assimilation for mesoscale and storm-scale applications. *Meteorology and Atmospheric Physics*, **82**, 173–208.
- Press WH, Teukolsky SA, Vetterling WT *et al.* (1992) *Numerical Recipes in Fortran 77: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK.
- Raupach MR (2001) Inferring biogeochemical sources and sinks from atmospheric concentrations: general considerations and applications in vegetation canopies. In: *Global Biogeochemical Cycles in the Climate System* (eds Schulze E-D, Heimann M, Harrison S, Holland E, Lloyd J, Prentice IC, Schimel DS), pp. 41–60. Academic Press, San Diego.
- Raupach MR, Canadell JG, Bakker DCE *et al.* (2004) Interactions between CO₂ stabilization pathways and requirements for a sustainable earth system. In: *The Global Carbon Cycle: Integrating Humans, Climate, and the Natural World* (eds Field CB, Raupach MR), pp. 131–162. Island Press, Washington.
- Rayner PJ (2001) Atmospheric perspectives on the ocean carbon cycle. In: *Global Biogeochemical Cycles in the Climate System* (eds Schulze E-D, Heimann M, Harrison S, Holland E, Lloyd J, Prentice IC, Schimel DS), pp. 285–294. Academic Press, San Diego.
- Rayner PJ, Enting IG, Francey RJ *et al.* (1999) Reconstructing the recent carbon cycle from atmospheric CO₂, delta C-13 and O₂/N₂ observations. *Tellus Series B—Chemical and Physical Meteorology*, **51**, 213–232.

- Rayner PJ, Law RM, O'Brien DM *et al.* (2002) Global observations of the carbon budget-3. Initial assessment of the impact of satellite orbit, scan geometry, and cloud on measuring CO₂ from space. *Journal of Geophysical Research–Atmospheres*, **107**, 4557, doi:10.1029/2001JD000618.
- Rayner PJ, O'Brien DM (2001) The utility of remotely sensed CO₂ concentration data in surface source inversions. *Geophysical Research Letters*, **28**, 175–178.
- Reichle RH, McLaughlin DB, Entekhabi D (2002) Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, **130**, 103–114.
- Rodenbeck C, Houweling S, Gloor M *et al.* (2003a) CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmospheric Chemistry and Physics*, **3**, 1919–1964.
- Rodenbeck C, Houweling S, Gloor M *et al.* (2003b) Time-dependent atmospheric CO₂ inversions based on interannually varying tracer transport. *Tellus Series B–Chemical and Physical Meteorology*, **55**, 488–497.
- Running SW, Baldocchi DD, Turner DP *et al.* (1999) A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of the Environment*, **70**, 108–127.
- Schimel DS, House JL, Hibbard KA *et al.* (2001) Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature*, **414**, 169–172.
- Sellers PJ (1985) Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, **6**, 1335–1372.
- Sellers PJ, Berry JA, Collatz GJ *et al.* (1992) Canopy reflectance, photosynthesis, and transpiration 3: a reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sensing of the Environment*, **42**, 187–216.
- Styles JM, Raupach MR, Farquhar GD *et al.* (2002) Soil and canopy CO₂, (CO₂)-C-13, H₂O and sensible heat flux partitions in a forest canopy inferred from concentration measurements. *Tellus Series B–Chemical and Physical Meteorology*, **54**, 655–676.
- Tarantola T (1987) *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, Amsterdam.
- Todling R (2000) Estimation theory and atmospheric data assimilation. In: *Inverse Methods in Global Biogeochemical Cycles* (eds Kasibhatla P, Heimann M, Rayner P, Mahowald N, Prinn RG, Hartley DE), pp. 49–66. American Geophysical Union, Washington, DC.
- Tucker CJ (1979) Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, **8**, 127–150.
- Wang YP, Barrett DJ (2003) Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach I. Using remotely sensed data and ecological observations of net primary production. *Tellus Series B–Chemical and Physical Meteorology*, **55**, 270–289.
- Wang YP, Bentley ST (2002) Development of a spatially explicit inventory of methane emissions from Australia and its verification using atmospheric concentration data. *Atmospheric Environment*, **36**, 4965–4975.
- Wang YP, McGregor JL (2003) Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach II. The Atmospheric constraint. *Tellus Series B–Chemical and Physical Meteorology*, **55**, 290–304.
- Wofsy SC, Goulden ML, Munger JW *et al.* (1993) Net exchange of CO₂ in a midlatitude forest. *Science*, **260**, 1314–1317.