Tellus (2006), 58B, 279–292
Printed in Singapore. All rights reserved

TELLUS

Regression tools for CO_2 inversions: application of a shrinkage estimator to process attribution

By BENJAMIN A. SHABY* and CHRISTOPHER B. FIELD, *Department of Global Ecology, Carnegie Institution of Washington, 260 Panama Street, Stanford, CA 94305, USA*

(Manuscript received 26 January 2005; in final form 10 March 2006)

ABSTRACT

In this study we perform an atmospheric inversion based on a shrinkage estimator. This method is used to estimate surface fluxes of CO_2, first partitioned according to constituent geographic regions, and then according to constituent processes that are responsible for the total flux. Our approach differs from previous approaches in two important ways. The first is that the technique of linear Bayesian inversion is recast as a regression problem. Seen as such, standard regression tools are employed to analyse and reduce errors in the resultant estimates. A shrinkage estimator, which combines standard ridge regression with the linear 'Bayesian inversion' model, is introduced. This method introduces additional bias into the model with the aim of reducing variance such that errors are decreased overall. Compared with standard linear Bayesian inversion, the ridge technique seems to reduce both flux estimation errors and prediction errors. The second divergence from previous studies is that instead of dividing the world into geographically distinct regions and estimating the CO_2 flux in each region, the flux space is divided conceptually into processes that contribute to the total global flux. Formulating the problem in this manner adds to the interpretability of the resultant estimates and attempts to shed light on the problem of attributing sources and sinks to their underlying mechanisms.

1. Introduction

The technique of inverse modelling is widely used for solving problems in fields ranging from electrocardiography to seismology (e.g. Rudy and Messinger-Rapport, 1988; Yan et al., 1989). Inverse problems arise when one has a known physical model that relates input variables to observations, and one wishes to solve for the inputs. An important special case of inverse modelling occurs when the physical model represents a linear relationship between the inputs and the observations. While the linearity assumption greatly simplifies the formulation, linear inversion problems often suffer from poor conditioning. Errors are likely to be introduced from many sources, including flaws in the physical theory and measurement errors in the observed data. Another potentially large source of error is colinearity. Colinearity occurs when two or more of the inputs, alone or in combination, when propagated through the physical model, have similar effects on the observations. This condition can be pathological because widely varying input sets cannot be readily distinguished from each other through the observations, causing the inverse solutions to suffer from extremely high variance.

In this study we attempt to solve for surface fluxes of carbon dioxide, using atmospheric carbon dioxide concentration measurements and an atmospheric transport model. In so doing, we develop a novel formulation to the atmospheric inversion problem and re-examine linear inverse problems as special cases of linear regression problems, allowing us to borrow techniques from the regression literature to improve our inverse solutions.

2. The problem

Carbon dioxide is continually being both emitted and sequestered at the Earth's surface. Natural biological and chemical mechanisms, as well as anthropogenic processes, contribute to the Earth's total carbon budget. Each of these processes acts as either a source or a sink of carbon dioxide, and the sum of all the sources and sinks (sources will be labelled as positive fluxes and sinks as negative fluxes by convention) ends up in the atmosphere.

Before a flux-contributing element forges its signature on the atmosphere, it is subject to atmospheric circulation, which radically alters its characteristic pattern. What we wish to solve for is the magnitude of the annual surface fluxes associated with each process. What we have at our disposal is a series of CO_2 concentration measurements taken throughout a global network of monitoring stations and a model that calculates the results of atmospheric transport for a given flux distribution.

*Corresponding author.
e-mail: bshaby@globalecology.stanford.edu
DOI: 10.1111/j.1600-0889.2006.00189.x

Past applications of inversion models to global carbon flux quantification have predominantly employed the linear Bayesian inversion methods described in Tarantola (1987) (other approaches include 'mass-balance' inversions and state-space approaches, both described in Enting (2002)). In addition to sharing common methods, most are similar in that they solved for a spatial distribution of carbon sources and sinks, with the spatial regions defined as continuous blocks (Tans et al., 1990; Rayner et al., 1999; Gurney et al., 2002). A smaller number of studies have solved for a 'process' parameter, such as light use efficiency or ^{13}C discrimination (e.g. Knorr and Heimann, 1995; Kaminski et al., 2002; Randerson et al., 2002), and a few include some spatially overlapping components in addition to discrete blocks (e.g. Enting et al. 1995). Here we treat the 'Bayesian inversion' model as a form of shrinkage regression and extend it to yield more accurate flux measurements.

3. Bayesian inversion

In the Bayesian inversion technique (Tarantola, 1987), each flux-contributing unit has a prescribed spatio-temporal distribution of its resultant flux. Since it is assumed that any possible atmospheric concentration distribution is the result of some combination of fluxes due to these units, the set of flux units is called a basis. Normalizing each basis component to a total annual flux of 1 PgC yr^{-1}, we can let the vector $\mathbf{m}$ represent the scaling factor associated with each normalized flux map necessary to reconstruct observed atmospheric concentrations, represented by the vector $\mathbf{d}_{\text{obs}}$.

This normalized basis is subject to atmospheric transport, which is assumed to be a linear function. That is, we assume that the concentration at any point can be expressed as a linear combination of concentrations due to the normalized basis components. Thus, the result of atmospheric transport on the normalized basis can be represented by a matrix $\mathbf{G}$, with $\mathbf{Gm}$ being the concentrations predicted by the coefficient vector $\mathbf{m}$. Instead of solving for $\mathbf{m}$ by minimizing the residual sum of squares

$$\text{RSS} = (\mathbf{d}_{\text{obs}} - \mathbf{Gm})^T(\mathbf{d}_{\text{obs}} - \mathbf{Gm}), \tag{1}$$

the Bayesian strategy is to improve estimates by incorporating prior knowledge of the flux magnitudes. From a non-Bayesian point of view, the RSS solution to the inversion problem is unbiased but suffers from high variance, leading to high MSE (MSE = bias^2 + variance). A reasonable strategy might be to decrease MSE by reducing the variance of the estimate, even at the cost of introducing some bias.

In the Bayesian paradigm, we prescribe a prior distribution $P(\mathbf{m})$ over the magnitudes $\mathbf{m}$ of the normalized basis, which reflects our knowledge and beliefs about $\mathbf{m}$. If this distribution is Gaussian, and errors $\mathbf{C}_t$ due to the transport model and errors $\mathbf{C}_d$ in the atmospheric observations are both Gaussian, Tarantola (1987) shows that we can consider them together as a single covariance matrix $\mathbf{C}_T = \mathbf{C}_t + \mathbf{C}_d$, and that the posterior distribution $P(\mathbf{m}|\mathbf{d}_{\text{obs}})$ is also Gaussian.

The final 'estimates' are the components of the mean vector of the posterior distribution. Since the posterior is normal, the mean is equal to the mode. Thus, we seek to find the maximum over this density function. This is equivalent to finding the minimum of the quadratic function

$$\begin{aligned} S(\mathbf{m}) = {} & (\mathbf{Gm} - \mathbf{d}_{\text{obs}})^T \mathbf{C}_T^{-1}(\mathbf{Gm} - \mathbf{d}_{\text{obs}}) \\ & + (\mathbf{m} - \mathbf{m}_{\text{prior}})^T \mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}). \end{aligned} \tag{2}$$

A potential source of large errors in calculations like these arises from colinearity. Colinearity occurs when two or more geographic regions or processes contribute similarly to the total concentrations at the observations stations. The consequence of collinear regression variables is that some coefficients tend to grow highly positive, while others tend to cancel them out by being highly negative. This condition leads to extremely high variance, and thus high expected error, in the resultant regression model.

4. Another look at colinearity and the bias-variance trade-off

In order to address the errors induced by colinearity among the bases, it is useful to recast the inversion problem in terms of regression theory. The first step is to view 'inversion theory' in the linear case as described by Tarantola (1987) as a Bayesian version of ordinary least squares regression, with aspherical priors on the regression coefficients. In fact, an identical regression model has a long history and has been extensively analysed in the statistical literature (see Lindley and Smith, 1972).

There are many regression models designed to address the issue of colinearity among input variables. It is well known that $\hat{\mathbf{m}}$, the least squares estimate of $\mathbf{m}$, is the minimum variance unbiased linear estimator. Often, however, the least squares estimator has a very high variance, and prediction errors can be reduced by inducing a little bit of bias in order to dramatically decrease the variance. Most variance-reduction methods that might be of interest use either subset selection methods or shrinkage methods. In subset selection, one chooses only a subset of the predictor variables to use as regressors. These techniques are inappropriate in the context of this inversion study because they would require completely discarding one or more of the geographic units or processes of interest. Shrinkage involves shrinking regression parameters along certain dimensions in the input space (Friedman et al., 2001).

Shrinkage methods include ridge regression (Hoerl and Kennard, 1970), truncated principal components regression (e.g. Cooley and Lohnes, 1971; Harman, 1976), partial least squares (Wold, 1975), and the LASSO (Tibshirani, 1996). Frank and Friedman (1993) conclude that ridge regression is slightly preferable to variable selection, principal components, and partial

least squares for reducing prediction error in most contexts. The LASSO also performs well, but requires the additional computational burden of quadratic programming and often shrinks regression coefficients all the way to zero. Truncated principal components regression has appeared in the inversion literature (e.g. Fan et al., 1999), but the results are notoriously difficult to interpret. For these reasons, we now restrict our focus to ridge regression.

With ridge regression, one seeks to shrink the regression coefficients by adding a penalty term to the OLS cost function:

$$\hat{\mathbf{m}}^{\text{ridge}} = \underset{\text{m}}{\text{argmin}} \sum_{i=1}^{N} \left[\left(\mathbf{d}_{\text{obs}_i} - \text{m}_0 - \sum_{j=1}^{p} g_{ij} \text{m}_j \right)^2 + \lambda^{\text{ridge}} \sum_{j=1}^{p} \text{m}_j^2 \right]. \quad (3)$$

The parameter λ^{ridge} is called the ridge parameter. The ridge parameter scales the size penalty on the regression coefficients and therefore controls the amount of shrinkage. How to choose a good value for λ^{ridge} will be addressed below.

Centring the inputs by replacing each g_{ij} with $g_{ij} - \bar{g}_j$ and estimating m_0 by $\bar{\mathbf{d}}_{\text{obs}} = \sum_1^N \mathbf{d}_{\text{obs}_i}/N$, the ridge estimator can be rewritten in matrix form as

$$\hat{\mathbf{m}}^{\text{ridge}} = \underset{\text{m}}{\text{argmin}}\, \text{S}_{\text{ridge}}(\mathbf{m}), \quad (4)$$

where $\text{S}_{\text{ridge}}(\mathbf{m}) = (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m})^T (\mathbf{d}_{\text{obs}} - \mathbf{G}\mathbf{m}) + \lambda^{\text{ridge}} \mathbf{m}^T \mathbf{m}$. This expression has the closed-form solution

$$\hat{\mathbf{m}}^{\text{ridge}} = (\mathbf{G}^T\mathbf{G} + \lambda^{\text{ridge}}\mathbf{I})^{-1}\mathbf{G}^T\mathbf{d}_{\text{obs}}. \quad (5)$$

Note that by adding the $\lambda^{\text{ridge}}\mathbf{I}$ term, a constant multiple of the $p \times p$ identity matrix, to $\mathbf{G}^T\mathbf{G}$, a constant term is added to the main diagonal of $\mathbf{G}^T\mathbf{G}$, guaranteeing that the matrix is non-singular. This becomes important if $\mathbf{G}^T\mathbf{G}$ is not of full rank, as would be the case if the number of relevant processes exceeded the number of CO_2 monitoring stations. This fact is equally important when $\mathbf{G}^T\mathbf{G}$ is close to singular in terms of numerical rank (see Golub and Van Loan, 1996).

The amount of coefficient shrinkage in the ridge regression model is not invariant to scaling of the inputs. For this reason, it is common practice to standardize the inputs to be mean zero and variance one before applying the procedure. In this application, however, all the covariates are in the same units and are normalized to unit flux, so such standardization is not necessary.

Ridge regression is equivalent to linear regression with zero-mean normal priors of common variance on the regression coefficients (Lindley and Smith, 1972). The ridge parameter can be seen as a parameter that specifies the covariance matrix (a constant multiple of the identity matrix) of the prior distribution. Thus, ridge regression is a special case of the Bayesian inversion model, with the prior distribution of the fluxes being $N(\mathbf{0},(\sigma^2/\lambda^{\text{ridge}})\mathbf{I})$. This not a surprising result, if one notices the similarities between the loss functions of the ridge regression and the Bayesian inversion models.

Ridge regression has appeared, explicitly and implicitly, in the inversion literature in the past. Mansbridge and Enting (1986) explore ridge regression as an alternative to OLS regression in the context of an ocean tracer inversion. Björkström (2001) advocates the use of ridge regression for a variety of inversion problems. Ridge regression has also previously been used for inversions under the name 'regularization' (e.g. Fan et al., 1999). Krakauer et al. (2004) describe a generalized ridge regression model for inversions for carbon fluxes. It is interesting to note that several recent inversion studies aiming to solve for CO_2 fluxes have used ridge regression. For example, Law et al. (2002) describe an inversion where all prior source magnitudes are assigned a value of zero, with uncertainties of 10 Gt C. This, of course, specifies a ridge regression model. The authors of this paper reason that the large uncertainty of 10 Gt C 'avoids biasing the answer too much'. However, we shall see that because of the one-to-one correspondence between the prior uncertainty in the Bayesian inversion model and the shrinkage parameter in the ridge regression model, it is possible to find prior uncertainty values that bias the resultant estimates optimally.

The ridge estimator does a good job of reducing the variance of regression models, but it does not take into account our prior knowledge about the sign and magnitude of the surface fluxes. Ideally, one would like to take advantage of the variance reduction properties of the ridge estimator and also leverage outside information to obtain the best possible estimate. One possible solution is to combine ridge regression with the Bayesian paradigm by penalizing the magnitude of the regression coefficients, as well as penalizing the weighted deviation from prior estimates. With this in mind, we return to the Bayesian inversion model developed above.

It is possible to view the Bayesian formulation in a non-Bayesian way, as a kind of shrinkage estimator. In fact, comparison of the function $\text{S}(\mathbf{m})$ with $\text{S}_{\text{ridge}}(\mathbf{m})$ shows that they are quite similar. If we consider the matrix $\mathbf{C}_T$ not as a covariance matrix, but as an element-wise collection of shrinkage parameters, we see that we can consider this model to be itself a form of ridge regression, only instead of shrinking towards zero, we are shrinking towards our prior flux estimates. It is with this in mind that we introduce a combined shrinkage estimator.

The resulting model, which we will call a combined ridge, seeks to minimize the loss function

$$\begin{aligned} \text{S}^{C-\text{ridge}}(\mathbf{m}, \lambda^{\text{ridge}}) = {} & (\mathbf{G}\mathbf{m} - \mathbf{d}_{\text{obs}})^T \mathbf{C}_T^{-1} (\mathbf{G}\mathbf{m} - \mathbf{d}_{\text{obs}}) \\ & + (\mathbf{m} - \mathbf{m}_{\text{prior}})^T \mathbf{C}_M^{-1} (\mathbf{m} - \mathbf{m}_{\text{prior}}) \\ & + \lambda^{\text{ridge}} \mathbf{m}^{\text{T}} \mathbf{m}. \end{aligned} \quad (6)$$

The closed form solution is easily obtained as

$$\begin{aligned} \hat{\mathbf{m}}^{C-\text{ridge}}(\lambda^{\text{ridge}}) = {} & \left(\mathbf{G}^{\text{T}}\mathbf{C}_T^{-1}\mathbf{G} + \mathbf{C}_M^{-1} + \lambda^{\text{ridge}}\mathbf{I}\right)^{-1} \\ & \times \left(\mathbf{G}^{\text{T}}\mathbf{C}_T^{-1}\mathbf{d}_{\text{obs}} + \mathbf{C}_M^{-1}\mathbf{m}_{\text{prior}}\right). \end{aligned} \quad (7)$$

Although this model has been presented as a non-Bayesian shrinkage model, it has an equally applicable Bayesian interpretation. The combined shrinkage model is again a Bayesian linear regression, with a prior on the regression coefficient vector being a normal distribution with mean $(\mathbf{C}_M^{-1} + \lambda^{\text{ridge}}\,\mathbf{I})^{-1}\,\mathbf{C}_M^{-1}\mathbf{m}_{\text{prior}}$ and covariance matrix $(\mathbf{C}_M^{-1} + \lambda^{\text{ridge}}\mathbf{I})^{-1}$. Note that this is the similar to the prior in the ordinary Bayesian inversion case, with the prior mean adjusted toward zero according to the relative weights of $\mathbf{C}_M$ and $\lambda^{\text{ridge}}\mathbf{I}$ and the prior covariance matrix shrunken along the main diagonal. The result is that the ridge penalty effectively offsets the effects of priors that are unwarrantedly large in magnitude, yielding improved posterior estimates.

Expressing the model in this way also makes it clear that the combined ridge model is an empirical Bayes method, rather than a fully Bayesian formulation, as the prior distribution is a function of the data (through the determination of λ^{ridge}). Krakauer et al. (2004) propose a similar empirical Bayes model. However, instead of modifying the usual prior distribution by adjusting the mean and the main diagonal of the prior covariance toward zero by an amount controlled by a tunable parameter as we do here, they use the tunable parameter to scale the prior covariance matrix. Viewed as a shrinkage estimator, we differentially shrink flux estimates toward some combination of zero and prior estimates according to an approximately optimal estimation error criterion. They shrink only toward prior estimates and choose a shrinkage parameter by minimizing a generalized cross validation function, a slightly different approximate optimality criterion than the cross validation function described below. These and other similar strategies have long been established in the regression literature (e.g. Golub et al., 1979) as powerful tools for improving both parameter estimates and predictions.

4.1. Choosing the ridge parameter

Much attention has been focused on how best to estimate the optimal λ^{ridge} (e.g. Golub et al., 1979; Gibbons, 1981; Nordberg, 1982; Lee, 1987). In most contexts, optimality is defined as the value of λ^{ridge} that minimizes prediction error. That is, it is the value of λ^{ridge} that minimizes the errors if one were to use the regression model to predict the concentration at a new CO_2 monitoring station.

The most common approach to minimizing prediction error is K-fold cross-validation. The idea behind cross-validation is to divide the input data into K equally sized partitions, repeatedly estimating $\hat{\mathbf{m}}^{\text{C-ridge}}(\lambda^{\text{ridge}})$ K times, each time omitting one of the partitions to use for validation. More concretely, let $\hat{\mathbf{m}}^{\text{C-ridge}(k)}(\lambda^{\text{ridge}})$ be the ridge estimator with partition k removed from $\mathbf{d}_{\text{obs}}$ and $\mathbf{G}$. The argument is that if λ^{ridge} is chosen well, then the k^{th} partition of $\mathbf{G}\hat{\mathbf{m}}^{\text{C-ridge}(k)}(\lambda^{\text{ridge}})$ should be a good predictor of the omitted portion of $\mathbf{d}_{\text{obs}}$. Averaging the squared error of $\mathbf{G}\hat{\mathbf{m}}^{\text{C-ridge}(k)}(\lambda^{\text{ridge}})$ over all k, one arrives at the cross-validation estimate of prediction error.

1. Randomly divide the rows of $\mathbf{G}$ and the corresponding entries of $\mathbf{d}_{\text{obs}}$ into K groups of roughly equal size.
2. Set aside the k^{th} group of data, and use the remaining $K-1$ groups of data to calculate $\hat{\mathbf{m}}^{\text{C-ridge}(k)}(\lambda^{\text{ridge}})$.
3. Repeat step 2 for $k = 1, 2, \ldots, K$ and calculate

$$\text{CV}(\lambda^{\text{ridge}}) = \frac{1}{K}\sum_{k=1}^{K} \times \left\{ \sum_{i \in k\text{th group}} \left[\left[\mathbf{G}\hat{\mathbf{m}}^{\text{C-ridge}(k)}(\lambda^{\text{ridge}})\right]_i - \mathbf{d}_{\text{obs}_i}\right]^2 \right\}$$

4. Repeat steps 1–3 for many values of λ^{ridge}, to find

$$\hat{\lambda}^{\text{ridge}}_{CV} = \underset{\lambda^{\text{ridge}}}{\text{argmin}}(\text{CV}(\lambda^{\text{ridge}})) \tag{8}$$

(Golub et al., 1979).

An important consideration, however, is whether the above view of the 'optimality' of $\hat{\lambda}^{\text{ridge}}{}_{CV}$ is appropriate in the present context. As the goal here is to estimate fluxes, we are in the unusual situation of having the regression coefficients themselves, rather than predictions they induce, as the primary quantities of interest. In light of this unusual perspective, it might be advantageous to employ a strategy other than minimizing prediction error in choosing a ridge parameter. One possible approach is to simply make up hypothetical regression coefficients, simulate observations from these coefficients and the transport matrix, and perform the shrinkage procedure, choosing the value of λ^{ridge} that produces posterior estimates that are as close as possible to the original coefficients. Conducting this simulation many times yields a ridge parameter that, under a particular set of assumptions, is optimized for producing the most accurate posterior flux estimates.

More concretely, we choose λ^{ridge} using the following algorithm:

(a) Let $\mathbf{m}_{\text{coef}}$ be a random deviate from the distribution $N(\mathbf{m}_{\text{prior}}, \mathbf{C}_{\text{m}})$, and let $\mathbf{G}_{\text{coef}}$ be the original transport matrix $\mathbf{G}$.

(b) Calculate $\mathbf{d}_{\text{coef}_{\text{ideal}}} = \mathbf{G}_{\text{coef}}\mathbf{m}_{\text{coef}}$, the vector of observations if $\mathbf{m}_{\text{coef}}$ were the true fluxes, $\mathbf{G}_{\text{coef}}$ described atmospheric transport exactly, and there were no observational or processing errors in the CO_2 data.

(c) Let $\mathbf{d}_{\text{coef}}$ be a random deviate from $N(\mathbf{d}_{\text{coef}_{\text{ideal}}}, \mathbf{C}_{\text{D}})$ to simulate observational noise in the typical range of the combined uncertainty arising from $\mathbf{d}_{\text{obs}}$ and $\mathbf{G}$.

(d) Using the simulated values $\mathbf{G}_{\text{coef}}$ and $\mathbf{d}_{\text{coef}}$, calculate

$$\hat{\mathbf{m}}_{\text{coef}}(\lambda^{\text{ridge}}) = \text{argmin}[(\mathbf{G}_{\text{coef}}\mathbf{m} - \mathbf{d}_{\text{coef}})^T\mathbf{C}_T^{-1}(\mathbf{G}_{\text{coef}}\mathbf{m} - \mathbf{d}_{\text{coef}}) + (\mathbf{m} - \mathbf{m}_{\text{prior}})^T\mathbf{C}_M^{-1}(\mathbf{m} - \mathbf{m}_{\text{prior}}) + \lambda^{\text{ridge}}\mathbf{m}^{\text{T}}\mathbf{m}]_{\mathbf{m}}$$

(e) Repeat step 4 for many values of λ^{ridge} to find

$$\hat{\lambda}^{\text{ridge}}_{\text{coef}} = \underset{\lambda^{\text{ridge}}}{\text{argmin}} \left\| \mathbf{m}_{\text{coef}} - \hat{\mathbf{m}}_{\text{coef}}(\lambda^{\text{ridge}}) \right\|, \tag{9}$$

where $||\cdot||$ indicates the Euclidian norm.

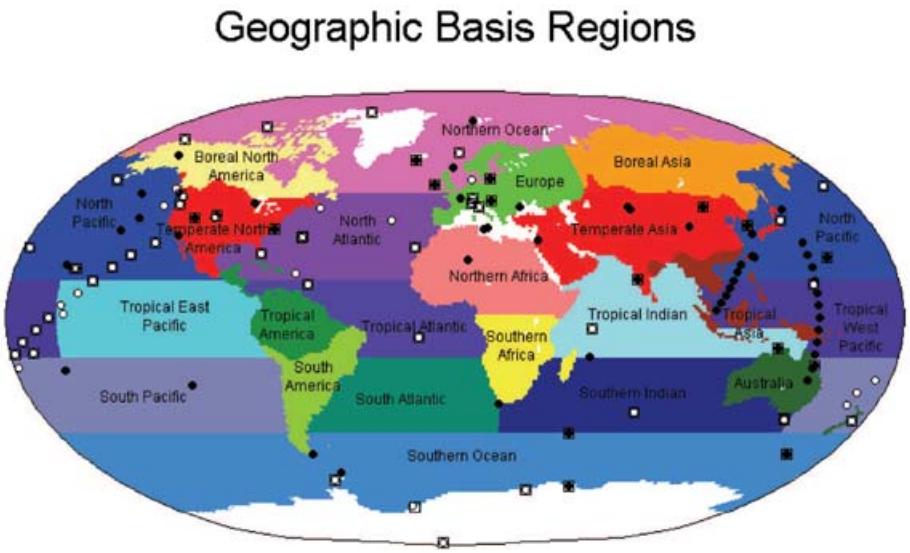

Fig. 1. The basis regions for the 'baseline' inversion. Markers on the map indicate the locations of data collection sites in the GlobalView database. White dots indicate sites that were used in the combined ridge inversion, solid dots indicate sites that were not used in the combined ridge inversion, and boxes indicate sites that were used in the 'baseline' inversion.

When performed many times, on average the above procedure will estimate the optimal value for the ridge parameter, where optimality is defined in terms of proximity of posterior flux estimates to actual flux values, in situations similar to the CO_2 scenario in the reference period.

5. The combined ridge model applied to a 'baseline' inversion problem

To examine the behaviour of model described above, it was applied to a 'baseline' inversion problem. The 'baseline' setup is described in Gurney et al. (2003) and consists of a collection of 22 geographically discrete bases, 11 terrestrial and 11 oceanic regions (Fig. 1). In addition to these regional fluxes, 4 background fluxes (2 for fossil fuel emissions, one for oceanic fluxes, and one for neutral biosphere fluxes) were also run through the transport model and subtracted from observed concentrations. This setup was chosen for comparison purely because of its simplicity, which makes it amenable to comparisons between the standard Bayesian and combined ridge inversion models.

The GISS-UCB tracer transport model (Fung et al., 1983) was used for the forward simulations. This model takes as inputs spatio-temporal surface flux distributions and outputs the results of atmospheric circulation acting on the inputs after a prescribed length of time. The model operates on a 4×5 degree grid and uses hourly integrated winds derived from the GISS AGCM Model II (Hansen et al., 1983), which was run for 1 yr. Each source distribution was run through the GISS-UCB model for a total of 4 yr, approximating a steady state. Observational data used in this experiment is an average over 5 yr from 76 sites in the Glovalview dataset (Globalview, 2004). The Bayesian framework requires us to specify prior distributions of all process fluxes. As described above, all prior distributions are assumed to be Gaussian. See Gurney et al. (2003) for details and justification for the prior flux scenario.

Flux estimates obtained from the combined shrinkage model are shown next to estimates obtained from the standard Bayesian inversion model. The coefficient error method of calculating $\hat{\lambda}^{\text{ridge}}$ was used, with 1,000 iterations. The average value 0.322 of $\hat{\lambda}^{\text{ridge}}_{\text{coef}}$ was used in the inversion (Fig. 2). As expected, estimates are shrunken towards zero, relative to estimates from the standard Bayesian inversion, in many of the regions, some of them quite substantially. For example, the estimated source in tropical America decreased by over 25%, and the estimated sink for southern Africa decreased by over 18%. It should be noted that in these cases the combined ridge estimates are within one standard deviation of the posterior mean of the standard Bayesian inversion. Interestingly though, imposing the ridge penalty had the opposite effect on a few of the estimated fluxes. Estimates for boreal North America, boreal Asia, and 7 of the 11 oceanic fluxes increased in magnitude, albeit some of them very slightly. We see that the addition of the ridge penalty does not blindly reduce the magnitudes of flux estimates, but rather redistributes the flux in a way that lowers overall error on the estimated flux vector (Fig. 3).

6. Process inversion

We now turn our attention to a slightly different formulation of the inversion problem where the basis is not a set of spatially discrete geographic regions, but instead consists of a set of spatially overlapping distributions that are conceptually discrete. In

the past, a few CO_2 inversion studies have used the phrase process inversion to characterize an approach that solves for one or more parameters in a model that scales CO_2 uptake and release on the basis of a satellite vegetation index, temperature, or other environmental characteristics (e.g. Knorr and Heimann, 1995; Kaminski et al., 2002; Randerson et al. 2002). We have a very different concept, defining a process inversion as one that partitions flux among a number of fundamentally different mechanisms. Of the inversions designed to partition flux among non-overlapping geographic regions, many have worked with a small number of regions, typically entire or large parts of continents and oceans (Pacala et al., 2001). Other approaches begin with a larger number of regions (Kaminski et al., 1999). Still, the common element among studies has been the definition of source/sink regions that are simple, non-overlapping, and defined on the basis of a geographical scheme. This approach to defining spatial domains has both advantages and limitations. The main advantages are (1) the (potential) spatial independence of source/sink regions and (2) the (potential) simplicity of connecting inferred sources and sinks to political boundaries. The main disadvantage of this approach is the difficulty of interpreting fluxes defined only spatially.

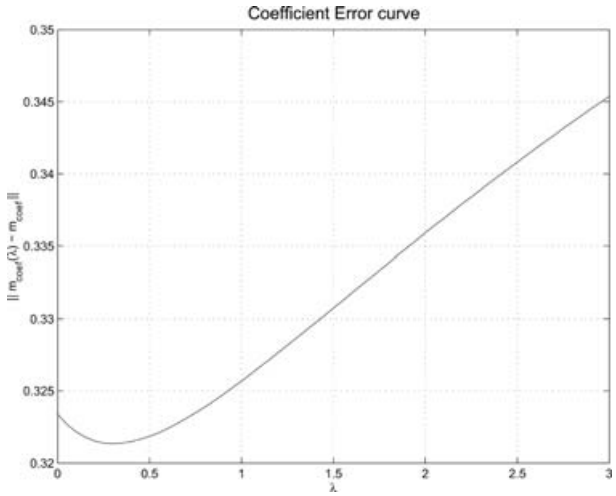

Fig. 2. A typical coefficient error curve estimated by the algorithm described in 9:1–5. Here, the minimum of the function lies near the value 0.322 that was used for the combined ridge inversion.

With linear inversion models, there is no intrinsic reason that the source/sink regions need to be simple in shape, uniform in flux density, or non-overlapping. The fluxes that contribute to the sources and/or sinks need only to be functionally independent. For the inversion to discriminate effectively, they also need to be spatially distinctive. This does not, however, mean that the contributing fluxes cannot overlap spatially.

Using a process-based basis instead of a purely geographical basis does not require changes to the mathematical formulation of the inversion, but it greatly alters its interpretability. Specifically, it allows us to directly address a problem that has attracted great attention recently, the problem of attributing sources and sinks to underlying mechanisms (Schimel et al., 2001). Here, we explore the idea that Bayesian inversion can be used as a tool for addressing the attribution problem.

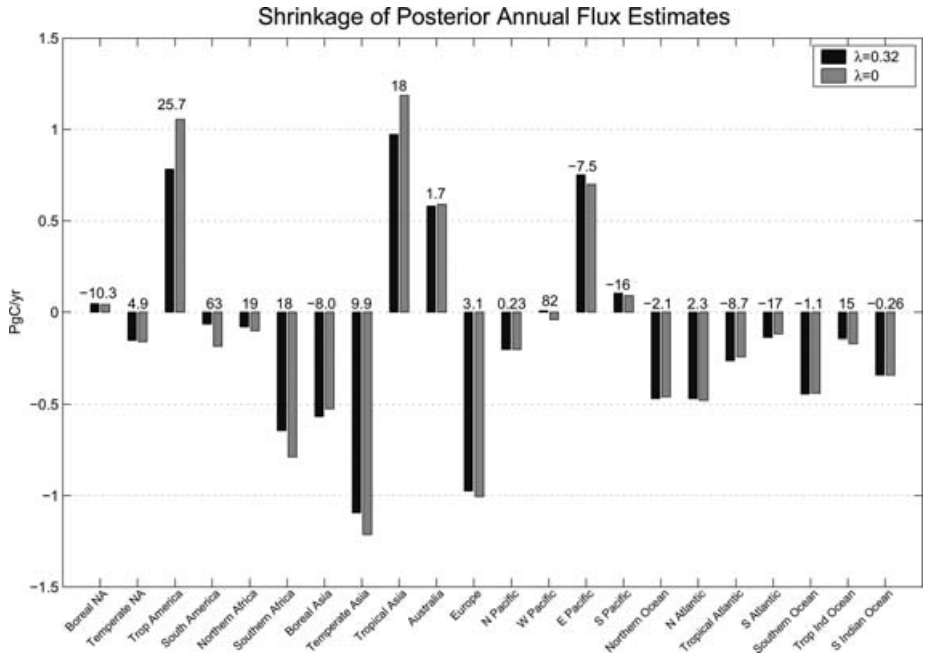

Fig. 3. Flux estimates obtained from the combined ridge inversion compared with estimates obtained from the standard Bayesian inversion. The numbers above the bars indicate how much the ridge estimate 'shrunk' relative to the standard estimate as a percentage of the standard estimate.

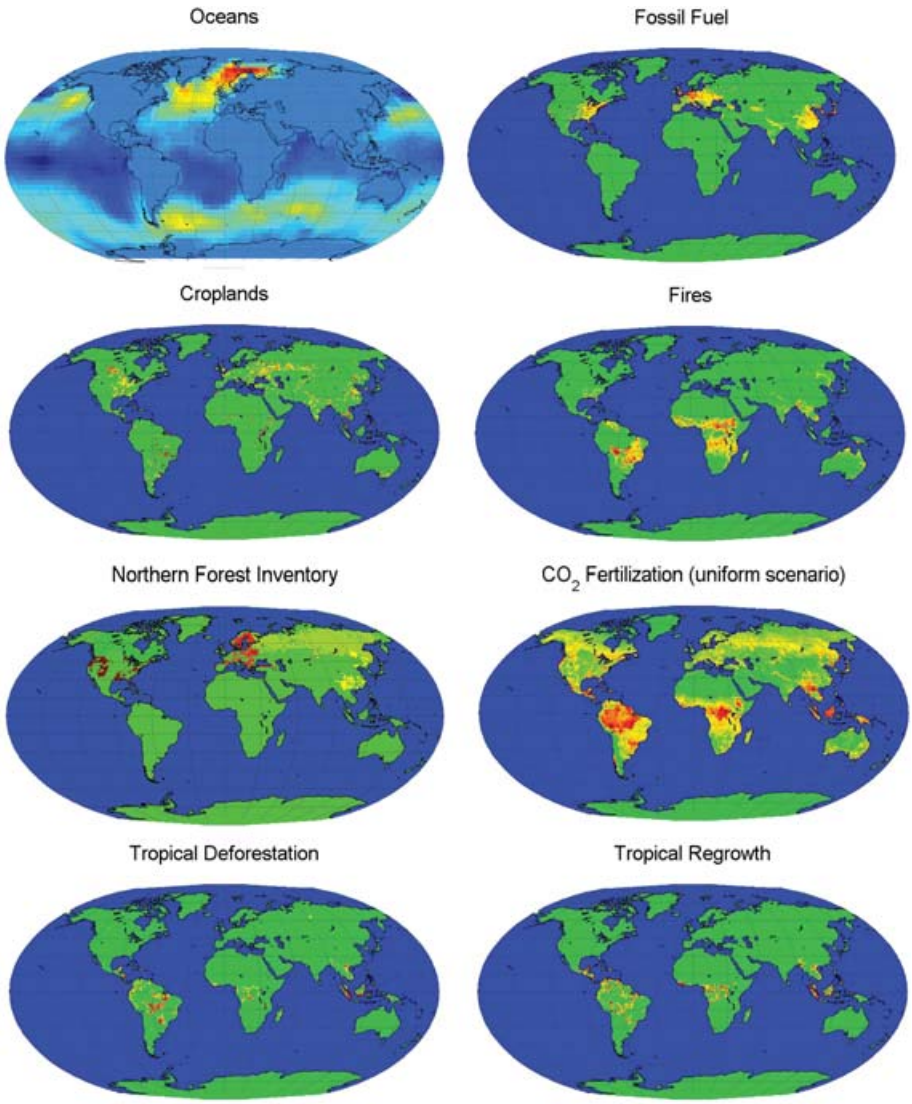

Fig. 4. The basis maps used in the 'process' inversion

The starting point for our 'process' inversion is a pair of important postulates. First, this approach depends on the idea that we can, with reasonable confidence, create a map of the spatial signature of each of the candidate mechanisms. Second, the spatial map needs to be stable to the variation in the total flux due to a given mechanism. With these two postulates, the effect of the inversion is to scale the flux maps associated with each mechanism and to identify the combination of scaled maps that best explains the global distribution of CO_2 concentrations.

6.1. Flux distributions

In principle, the set of process distributions should account for all possible CO_2 fluxes. For this study, 9 candidate processes were included, plus an additional pseudo-process that is spatially uniform and accounts for the CO_2 already present in the atmosphere at the beginning of the reference period. One may think of this pseudo-process as the y-intercept of a linear regression model. The processes that were included are: fossil fuel emissions (Andres et al., 1996), ocean carbon exchange (Takahashi et al., 2002), the equilibrated biosphere (modelled using CASA—see Potter et al., 1993), plant fertilization resulting from elevated levels of atmospheric CO_2 (Field and Shaby, in preparation), net changes in temperate cropland (Ramankutty and Foley, 1999), tropical deforestation (Defries et al., 2002), regrowth of disturbed tropical forest (DeFries et al., 2002), changes in the forest inventory in northern nations (Goodale et al., 2002), and fire suppression (Mouillot and Field, 2005). All processes are represented by spatially and temporally explicit flux distributions, which are then input to an GISS-UCB atmospheric transport model and run for 5 yr (Fig. 4).

The processes included in this study are not completely functionally independent, as they would be in an ideal setup. Instead, they are the best reasonable approximation to an orthogonal basis that we could construct given the available data. There are two immediate consequences of the bases not being functionally independent. The first is additional estimation error, again in the form of colinearity. The ridge penalty will mitigate this effect. The second is ambiguity in the interpretation of results. For example, when looking at the estimated flux due to fire suppression, one should keep in mind that part of this flux has been accounted for in the forest inventory basis.

6.2. Priors

Prior flux means were derived from the same sources as their corresponding spatio-temporal distributions. Standard deviations were set at 1 PgC for each process, with the exception of the neutral biosphere and tropical regrowth, whose fluxes are known to be extremely small (DeFries et al., 2002), and fossil fuel emissions and ocean fluxes, whose magnitudes are fairly well constrained.

6.3. Treatment of observational data

What is treated here as observational data was taken from the Globalview-CO_2 database. Globalview itself contains no actual observed measurements, but consists of smoothed, interpolated, and extended time series derived from measurements contained in the CMDL and WDCGG data archives. Smoothing, interpolation, and extension procedures are described fully by Masarie and Tans (1995). In addition to weekly concentration estimates, Globalview provides weight figures $w(i, j)$ at annual intervals for year i at monitoring station j. $w(i, j)$ is computed as follows:

$$w(i, j) = \alpha(i, j) \cdot \frac{\sqrt{N(i, j)}}{\mathrm{RSD}(i, j)},$$

where RSD(i,j) is the residual standard deviation of the actual flask measurements about the fitted smooth curve $\hat{f}\,j(i)$, $N(i,j)$ is the number of residuals used in the RSD calculation, and $\alpha(i,j)$ is a scaling factor.

For this study each $\mathbf{d}_{\mathrm{obs}(j)}$ was calculated as the average Glovalview concentration value over the reference period of 1985–1990 at station j. The average was taken over several years rather than just 1 yr to ensure that $\mathbf{d}_{\mathrm{obs}}$ is representative of the entire reference period, not just 1 yr, which could be atypical. Sites whose weights were less than 1.5 were not considered because we regarded them to be unreliable, as their corresponding concentration records were almost entirely 'extended', meaning they were not derived from actual observations during the reference period.

Diagonal entries $\mathbf{C}_{\mathrm{D}(j,j)}$ of the data covariance matrix consist of a constant times $w(i,j)^2$, again averaged over the reference period $i = 1985\ldots1990$. This is a reasonable estimate of measurement variance σ^2, as long as we assume that the smooth curve $\hat{f}_j(i)$ is a good estimate of $f_j(i)$, where flask measurements $y_j = f_j(i) + \varepsilon_j$ and $\varepsilon_j \sim N\,(0,\sigma_2^j)$. Off-diagonal entries of $\mathbf{C}_{\mathrm{D}}$ are left as zero, and errors $\mathbf{C}_{\mathrm{T}}$ due to transport are ignored for simplicity.

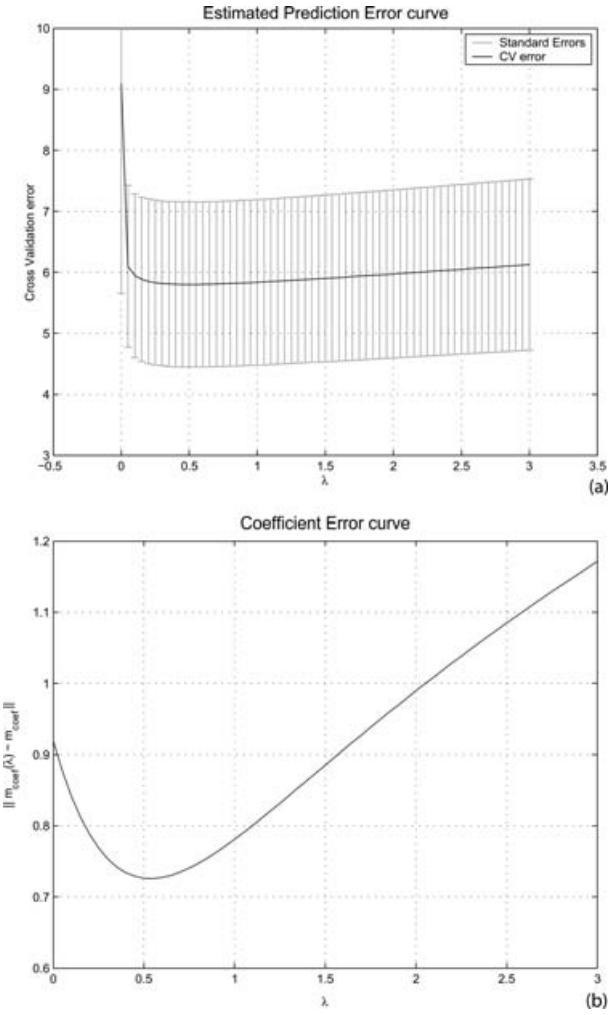

Fig. 5. A typical EPE curve estimated by cross-validation error (a) and a typical coefficient error curve (b). In both of these examples, the minimum lies near 0.467, the value used in the combined ridge 'process' inversion.

7. Results

Values of $\hat{\lambda}^{\mathrm{ridge}}{}_{CV}$ and $\hat{\lambda}^{\mathrm{ridge}}{}_{\mathrm{coef}}$ were each calculated 1,000 times. The respective means of the 1,000 values of $\hat{\lambda}^{\mathrm{ridge}}{}_{CV}$ and $\hat{\lambda}^{\mathrm{ridge}}{}_{\mathrm{coef}}$ were 0.429, and 0.467 (Fig. 5). These two estimates of the optimal ridge parameter are remarkably close. That the two methods of calculating $\hat{\lambda}^{\mathrm{ridge}}$, one which seeks to minimize prediction error, the other which seeks to minimize coefficient error, yield results that differ by only about 8% should inspire confidence in the application of the Bayesian ridge to this inversion problem;

introducing a small ridge penalty improves upon the standard Bayesian technique in terms of both estimating fluxes and making predictions about CO_2 concentrations.

The combined ridge inversion procedure was performed using with $\lambda^{\text{ridge}} = \bar{\hat{\lambda}}^{\text{ridge}}_{\text{coef}}$, using the same bases and concentration inputs as the standard inversion. Variances of the fluxes estimated using the combined ridge model are markedly smaller than those obtained using the standard Bayesian inversion model. This expected, as the posterior covariance matrix for the combined ridge model is

$$\mathbf{C}^{C-\text{ridge}}_{\text{posterior}} = \left(\mathbf{G}^{\mathrm{T}}{\mathbf{C}_T}^{-1}\mathbf{G} + {\mathbf{C}_M}^{-1} + \lambda^{\text{ridge}}\mathbf{I}\right)^{-1} \tag{10}$$

Comparing $\mathbf{C}^{C-\text{ridge}}{}_{\text{posterior}}$ with the posterior covariance matrix for the standard Bayesian model,

$$\mathbf{C}^{\text{Bayes}}_{\text{posterior}} = \left(\mathbf{G}^{\mathrm{T}}{\mathbf{C}_T}^{-1}\mathbf{G} + {\mathbf{C}_M}^{-1}\right)^{-1}, \tag{11}$$

we see that former is a shrunken version of the latter.

Also expectedly, annual flux estimates decreased in absolute value with the addition of the ridge penalty, some dramatically (Fig. 6). Fluxes due to the fire suppression and tropical deforestation processes, which were suspected to be highly colinear upon visual inspection, shrunk by 33.7 and 29.5 percent, respectively, relative to fluxes estimated by the standard Bayesian procedure. Somewhat unexpectedly, the croplands flux decreased by 42.4%. Though the croplands basis did not seem to have obvious colinearities with any other single basis, it is probable that it is highly correlated with a linear combination of two or more other bases. The tropical deforestation and the northern forest inventory bases, for example, seem to be likely candidates. As a means of comparison, the same calculation was done with $\lambda^{\text{ridge}} = \bar{\hat{\lambda}}^{\text{ridge}}_{\text{CV}}$, yielding similar results, as expected.

The amount of shrinkage in the process inversion is much greater than the baseline inversion. The enhanced shrinkage is a result of algorithm (9) selecting a larger shrinkage parameter. This is probably the case because colinearity among the process bases is more pronounced than among the geographically discrete bases. The more colinear the bases, the more the model will shrink the estimates toward zero to control the error.

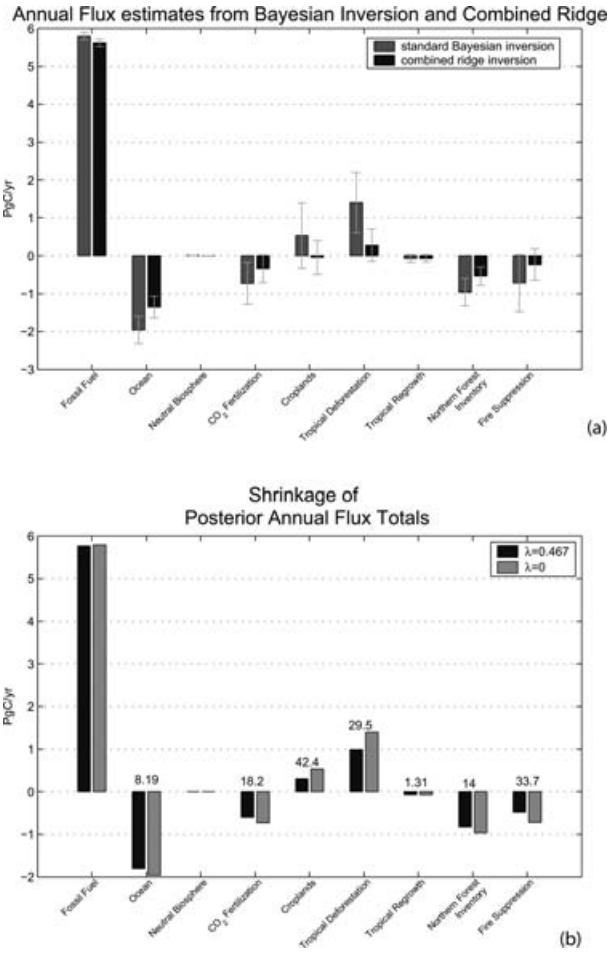

Fig. 6. Results of the combined ridge 'process' inversion. Note that the error bars around the ridge estimates are smaller than those around the standard Bayesian estimates (a). The amount of shrinkage resulting from the addition of the ridge penalty is shown (b). Again, numbers above the bars indicate the percentage change (relative to the Bayesian inversion) in the estimated flux

8. Error analysis

Assessing the quality of these results is not a clear-cut problem. Uncertainty in inversion studies arises from many sources. Colinearity among the basis functions, the basis functions themselves, the forward transport model, and the treatment of observation data all contribute errors that are difficult or impossible to quantify analytically.

Examination of the basis maps suggests how colinearity may come into play. For example, the distributions of the fire suppression and tropical deforestation bases overlap considerably. The results of the standard Bayesian inversion show that the magnitudes of the posterior flux estimates for these two processes grew markedly in magnitude relative to their respective prior estimates. $\hat{\mathbf{m}}_{\text{fire}}$ became more negative, and $\hat{\mathbf{m}}_{\text{tropDF}}$ became more positive, effectively cancelling each other out. It is likely that at least part of this effect is attributable to the spatial covariation between the two inputs. The addition of the ridge penalty can limit, but not eliminate, problems caused by colinearity in the bases.

We have presented our results visually (Figs 3 and 6) as point estimates with 'error bars' defined as ± 1 standard deviation, as is consistent with standard practice. This representation, however, is extremely misleading. In Bayesian analyses, inference is carried out based on the posterior distribution. These 'error bars' do *not* represent the posterior distribution in an especially meaningful way. It is standard in Bayesian analyses to report 'credible regions'. The most intuitive form of credible regions is the highest posterior density ('hpd') region. This region is

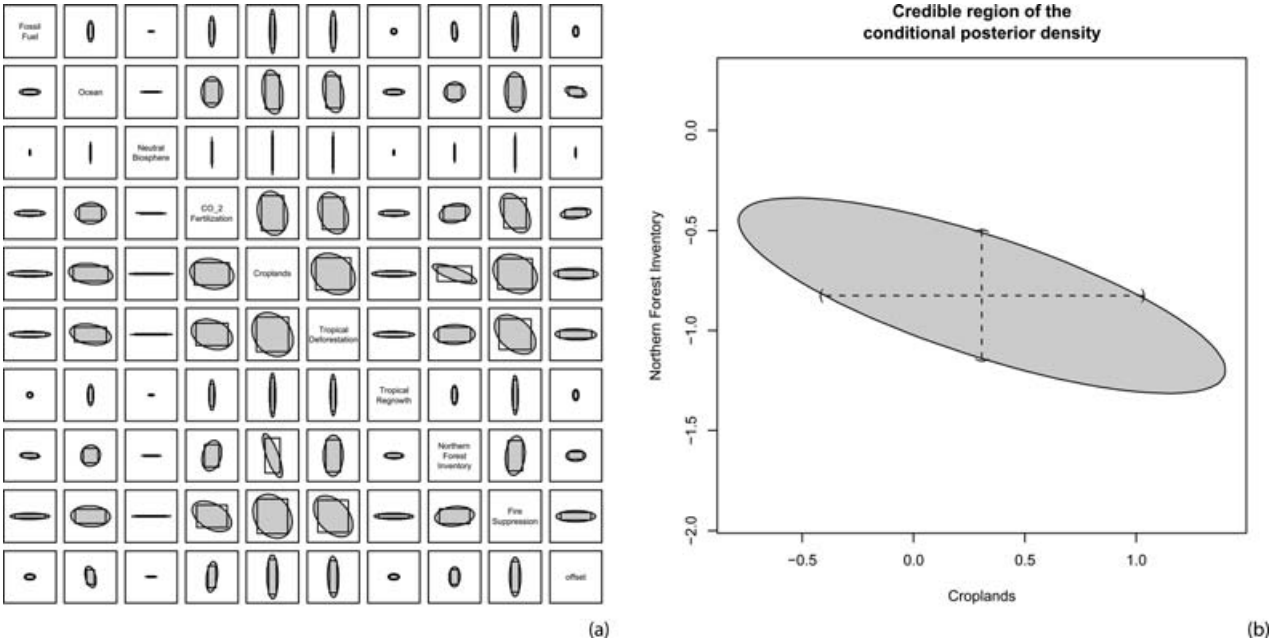

Fig. 7. 68% hpd ellipses of the joint density of components *i* and *j* conditional on all the remaining components being equal to their posterior means. (a) Shows these ellipses for all pairs *i* and *j*. The boxes represent the region implied by ±1 standard deviation. The ellipse for croplands and northern forest inventory is blown up for illustration. The dotted lines are the ±1 standard deviation intervals.

defined by fixing a level α and choosing the smallest region that contains a fraction of the posterior probability mass equal to α. One tends to think of the mean ±1 standard deviation as containing approximately 68% of the mass of a normal density. In the case of a multivariate normal, this is absolutely not the case. What ±1 standard deviation defines is referred to as the full conditional density of component j ($j = 1..p$). This is the density of component j, if we assume that the other components i ($i \neq j$) are known and equal to $\hat{\mathbf{m}}_{i \neq j}$. This is obviously not the case. In fact, the region implied by the ±1 standard deviation construction conveys little information about the hpd region of the multivariate normal.

The hpd region for a multivariate normal is an ellipsoid in p dimensions. The relevant region then is the ellipsoid that contains the fraction α of them posterior probability mass. The form of this ellipsoid is given by the vectors $\mathbf{x}$ that satisfy

$$(\mathbf{x} - \hat{\mathbf{m}}^{C-\text{ridge}})^{\mathrm{T}} \mathbf{C}_{\text{posterior}}^{C-\text{ridge}^{-1}} (\mathbf{x} - \hat{\mathbf{m}}^{C-\text{ridge}}) = k, \tag{12}$$

for the appropriate constant k. The quadratic form $(\mathbf{x} - \hat{\mathbf{m}}^{C-\text{ridge}})^{\mathrm{T}} \mathbf{C}_{\text{posterior}}^{C-\text{ridge}^{-1}} (\mathbf{x} - \hat{\mathbf{m}}^{C-\text{ridge}})$ is distributed as a non-central χ^2 with p degrees of freedom. Hence, letting α be .68 for consistency, and when $p = 10$, we get $k \approx 11.5$. The axes of the hpd ellipse can be obtained by finding the eigenvectors of the posterior covariance matrix, with the longest axis being the eigenvector corresponding to the largest eigenvalue. Note that inspection of the posterior covariance matrix alone *cannot* tell us much beyond the level of correlation between pairs of variables.

Since we cannot plot ellipsoids in 10 dimensions, we illustrate the problem in 2 dimensions. Figure 7(a) shows the joint posterior densities of components i and j conditional on all the other components being equal to their posterior means, for all pairs i and j. The boxes represent the regions implied by the ±1 standard deviation intervals, and the ellipses are the 68% hpd regions. Notice that these regions are *not* the same. Figure 7(b) shows a clear example of how the uncertainty reported as in Figs 3 and 6 is extremely misleading.

Conditional bivariate densities were chosen for illustrative purposes only—they represent something other than the posterior density, just as the full conditional densities do. While the hpd region is defined precisely mathematically, it is difficult to interpret when it exists in high dimensions. It simply does not make sense to report intervals for each individual component without considering all of the others at the same time, especially when strong correlations exist. Table 1 shows the extreme values of each component in the full 68% hpd ellipse. This is probably not a good way to quantify uncertainty, but it is interesting to compare the intervals generated in this way with the ±1 standard deviation intervals. In most cases, the sizes of the intervals differ by several fold.

Results of an inversion of this kind are also, of course, subject to random uncertainties. A useful tool for assessing the variability of the posterior means due to the randomness associated with the data collection sites from which concentration data was used is the bootstrap (Efron and Tibshirani, 1994). This randomness includes both the choice of the locations of these sampling sites,

Table 1. Intervals defined by ±1 S.D. and extreme values of the 0.68 credible ellipse for each component of the posterior.

Component	Interval defined by ± 1 S.D.	Extreme values of the 0.68 credible ellipse
Fossil fuel	±0.10	±0.29
Ocean	±0.34	±1.16
Neutral biosphere	±0.01	±0.03
CO_2 fertilization	±0.50	±1.68
Croplands	±0.72	±2.45
Tropical deforestation	±0.68	±2.28
Tropical regrowth	±0.10	±0.29
Northern forest inventory	±0.32	±1.09
Fire suppression	±0.64	±2.17
Offset	±0.18	±0.60

as well as the measurement error and errors related to interpolation and extension. Assuming that the site locations and their measurements were randomly sampled from an arbitrary multivariate probability distribution over the entire globe and the space of concentration values, the bootstrap can be used to estimate the distribution of combined ridge posterior means that arises from the randomness associated with data collection. The bootstrap procedure proceeds as follows, letting N be the number of monitoring sites under consideration:

1. Randomly choose, with replacement, N observations and call this vector $\mathbf{d}_{\mathrm{obs}(j)}{}^{\mathrm{bootstrap}}$
2. Construct $\mathbf{C}_{\mathrm{D}}{}^{\mathrm{bootstrap}}$ and $\mathbf{G}^{\mathrm{bootstrap}}$ by selecting the respective rows of $\mathbf{C}_{\mathrm{D}}$ and $\mathbf{G}$ corresponding to the observations selected for $\mathbf{d}_{\mathrm{obs}(j)}{}^{\mathrm{bootstrap}}$
3. Perform the inversion described above, estimating the minimum $\mathbf{m}^{\mathrm{bootstrap}}$ from $\mathbf{d}_{\mathrm{obs}(j)}{}^{\mathrm{bootstrap}}$, $\mathbf{C}_{\mathrm{D}}{}^{\mathrm{bootstrap}}$, and $\mathbf{G}^{\mathrm{bootstrap}}$
4. Repeat steps 1–3 many times

The distribution of the $\hat{\mathbf{m}}^{\mathrm{bootstrap}}$ estimates approximates the distribution of $\hat{\mathbf{m}}^{C-\mathrm{ridge}}$ due to the randomness of the observation stations.

The results of the resampling simulations using the data described above show that posterior flux estimates are within a standard deviation of the mean of the bootstrap-estimated distribution of $\hat{\mathbf{m}}^{C-\mathrm{ridge}}$. All individual posterior estimates are shown to be relatively insensitive to randomness present in selection and treatment of CO_2 concentration data, with the possible exception of the CO_2-fertilization process. In this case, while the combined ridge estimate falls within a standard deviation of the mean of the bootstrap distribution, it predicts a sink that is nearly twice as great as the bootstrap mean. In all cases, bootstrap standard deviations are of similar magnitude to standard deviations of the combined ridge flux estimates (Fig. 8).

When the same bootstrap procedure is performed using all available Globalview CO_2 time series instead of only those

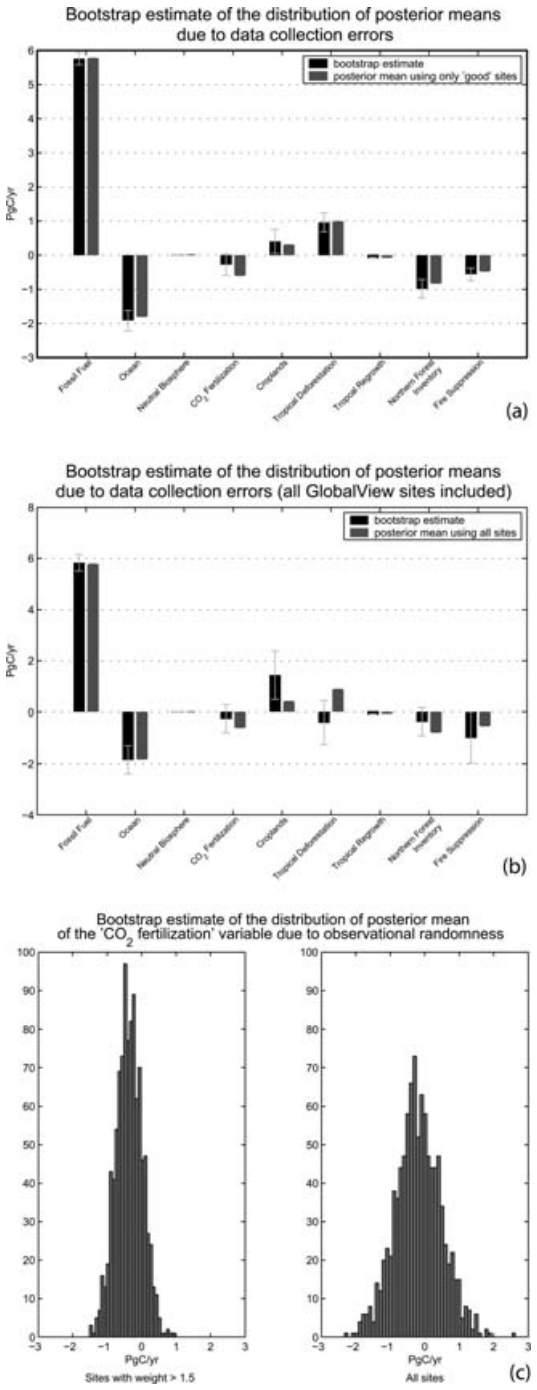

Fig. 8. When only 'good' data series are used, bootstrap means of the distribution of combined ridge estimates are within one standard deviation of the flux estimates in all cases (a). If all data series are used, the bootstrap shows extremely high variance for combined ridge flux estimates (b). In some cases, flux estimates do not even fall within these large standard deviations. One may compare the histograms of the bootstrap-sampled posterior means of the CO_2 fertilization variable when only 'good' data are used with the same histogram when all series are used (c). In the second case, the mean of the distribution is shifted toward zero, and the variance increases markedly.

whose weights were greater than 1.5, the means of the bootstrap distributions of $\hat{\mathbf{m}}^{C-\text{ridge}}$ differ substantially from combined ridge estimates of $\hat{\mathbf{m}}^{C-\text{ridge}}$, and the variances of bootstrap distributions of $\hat{\mathbf{m}}^{C-\text{ridge}}$ increase several fold. In one extreme case, that of the tropical deforestation basis, the bootstrap procedure estimates a sink of 0.40 Pg, compared to a source of 0.88 Pg estimated by the combined ridge inversion. The large variances and shifting of the means might be interpreted in several ways. One possibility is that the particular subset of sites whose data weights exceed 1.5 is not a representative sample from the distribution described above, and the addition of the remainder of the CO_2 data corrects this bias. If this were the case, then the flux estimates presented above must be regarded with some scepticism. Another, more likely, interpretation is that the original subset is a good representative sample, and adding the remainder of the data introduces bias and high variance because the additional data itself suffers from high bias and variance. This additional data was derived from few or no actual flask measurements taken during the reference period. It is possible that the high bias and variance exhibited in the bootstrap simulations indicates that the extension procedure used to generate these low-weighted time series does not accurately reconstruct the true concentrations.

9. Discussion

This study puts forth two major ideas that are significant diversions from the standard linear Bayesian inversion techniques that have been applied to problems like estimating surface CO_2 fluxes from CO_2 concentration data. The first is viewing inversion as a form of regression, and in so doing drawing from the wealth of available regression techniques to diagnose and reduce errors. The second is the idea that the flux basis may consist of maps representing fluxes that arise from conceptually discrete mechanisms, rather than from geographically discrete regions.

By re-examining the linear Bayesian inversion technique and applying tools that are commonly employed in the regression context, it is possible to improve the results of studies like the one presented here. The application of the bootstrap gives us insights into the nature of the error distribution that arises from the treatment of the observational data. This is valuable in itself, but another, more powerful interpretation of the bootstrap results is possible. One might view the bootstrap means as a more robust estimate of the fluxes under consideration. Viewed in this light, the bootstrap provides a separate set of solutions to the inverse problem that can be compared with the Bayesian and combined ridge solutions.

The combined ridge model presents an improvement over the standard linear Bayesian method. By biasing the model with a ridge penalty, the resultant decrease in variance improves the overall error rate. It can be seen from the error curves presented above that whether we are concerned with prediction error or coefficient error (i.e. errors in our flux estimates), adding a non-zero ridge parameter improves the results. Further, simulation tests show that not only is the optimal ridge parameter non-zero in both of the contexts under consideration, but also that, for this problem, the optimal ridge parameters under both interpretations of optimality are strikingly similar. This suggests that the combined ridge model is an improvement upon the procedure that is often employed for calculating flux estimates from atmospheric concentrations of CO_2.

Of course, this study only presents a tiny fraction of the tools available in the regression literature that may be useful in the inversion context. Further inquiry into these methods may prove valuable to the atmospheric inverse modelling community.

Starting with a basis consisting of flux distributions due to specific processes lends important advantages in interpretability and utility of the results. Insights gained toward solving the attribution problem have the potential to be extremely valuable. The ability to effectively partition carbon balance into components due to different processes is a fundamental prerequisite for credible models of the future of the carbon cycle. Interest in this partitioning or attribution problem has increased recently, with the maturation of the accounting standards for the Kyoto Protocol (UNFCCC). The report of the seventh conference of the parties (October–November 2001) of the United Nations Framework Convention on Climate Change decided that credits for carbon sequestration would be limited to direct effects of active management since 1990, after correcting for indirect effects from past practices, climate change, elevated atmospheric CO_2, and other factors. With this decision, the attribution problem assumes direct policy relevance as well as fundamental scientific importance.

To what extent can our process-based inversion augment the results from geographically based analyses? As an approach to estimating regional fluxes, the process-based inversion is not yet at a level of maturity to add quantitative constraints. Prerequisites for this will include both improvements in spatial map associated with each flux and additional data constraints that can be built into the inversion. While the process-based inversion cannot yet improve regional flux estimates, it creates a number of potentially important opportunities. Many of these involve improved connections between bottom-up process studies and top-down constraints (Schimel et al., 2001). Others involve contributions to unravelling the contributions of several spatially overlapping mechanisms. Ultimately, this kind of approach may facilitate the task of separating effects on sources and sinks from direct human actions, indirect human actions, and past practices, as required by the Marrakech Accords to the Kyoto Protocol (UNFCCC).

10. Acknowledgments

We would like to thank N. Ramankutty and J. Foley for making their landcover change data available to us. We are much obliged to J. Randerson and N. Krakauer for providing helpful comments and suggestions. We also wish to thank I. Fung and J. John for their advice and help with the tracer transport

model. We are grateful to K. Gurney for his help in explaining the details of the TransCom project. Thanks to NASA for support from the EOS-IDS program and to the Office of Science, Biological and Environmental Research Program (BER), US Department of Energy, through the Western Regional Center of the National Institute of Global Environmental Change (NIGEC) under Cooperative Agreement Nos. DE-FC03-90ER61010 and DE-FC02-03ER63613.

Financial support does not constitute an endorsement by DOE of the views expressed in this article/report.

References

Andres, R. J., Marland, G., Fung, I. Y. and Matthews, E. 1996. A 1 degrees x 1 degrees distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990. *Global Biogeochemical Cycles* **10**, 419–429.

Björkström, A. 2001. Ridge regression and inverse problems. Technical Report in Mathematical Statistics 2000:5. Stockholm University, Sweden.

Cooley, W. and Lohnes, P. 1971. *Multivariate Data Analysis*. Wiley, New York, 364.

DeFries, R., Houghton, R. A., Hansen, M., Field, C. B., Skole, D. and Townshend, J. 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 14 256–14 261.

Efron, B. and Tibshirani, R. 1994. *An Introduction to the Bootstrap*. Chapman and Hall, London, 436.

Enting, I. G. 2002. *Inverse Problems in Atmospheric Constituent Trnasport*. Cambridge University Press, London, 394.

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

Fan, S., Sarmiento, J. L., Gloor, M. and Pacala, S. W. 1999. On the use of regularization techniques in the inverse modeling of atmospheric carbon dioxide. *Journal of Geophysical Research* **104**(D17), 21 503–21 512.

Frank, I. and Friedman, J. 1993. A statistical view of some chemometrics regression tools (with discussion). *Technometrics* **35**(2), 109–148.

Friedman, J., Hastie, T. and Tibshirani, R. 2001. *The Elements of Statistical Learning*. Springer-Verlag, New York, 533.

Fung, I. Y., Prentice, K., Matthews, E., Lerner, J. and Russell, G. 1983. Three-dimensional tracer model study of atmospheric CO_2: response to seasonal exchanges with the terrestrial biosphere. *Journal of Geophysical Research* **88**, 1281–1294.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. 2004. *Bayesian Data Analysis*. Chapman and Hall, London, 668.

Gibbons, D. 1981. A simulation study of some ridge estimators. *Journal of the American Statistical Association* **76**, 131–139.

GLOBALVIEW-CO_2: Cooperative Atmospheric Data Integration Project—Carbon Dioxide. 2004. CD-ROM, NOAA CMDL, Boulder, Colorado [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].

Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 694.

Golub, G. H., Heath, M. and Wahba, G. 1979. Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–224.

Goodale, C., Apps, M., Birdsey, R., Field, C. B., Heath, L. and co-authors. 2002. Forest carbon sinks in the Northern Hemisphere. *Ecological Applications* **12**, 891–899.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D. and co-authors. 2002. Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models. *Nature* **415**, 626–630.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D. and co-authors. 2003. TransCom 3 CO_2 inversion intercomparison: 1. Annual mean control results and sensitivity to transport and prior flux information. *Tellus* **55B**, 555–579.

Hansen, J., Russel, G., Rind, D., Stone, P., Lacis and co-authors. 1983. Efficient three-dimensional global models for climate studies: Models I and II. *M. Weather Review* **111**, 609–662.

Harman, H. 1976. *Modern Factor Analysis*. University of Chicago Press, Chicago, 487.

Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–67.

Kaminski, T., Heimann, M. and Giering, R. 1999. A coarse grid three-dimensional global inverse model of the atmospheric transport—2. Inversion of the transport of CO_2 in the 1980s. *Journal of Geophysical Research—Atmospheres* **104**, 18 555–19 581.

Kaminski, T., Knorr, W., Rayner, P. J. and Heimann, M. 2002. Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. *Global Biogeochemical Cycles* **16**, 1066–1083.

Knorr, W. and Heimann, M. 1995. Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model. *Tellus* **47B**, 461–489.

Krakauer, N. Y., Schneider, T., Randerson, J. T. and Olsen, S. C. 2004. Using generalized cross-validation to select parameters in inversions for regional carbon fluxes. *Geophysical Research Letters* **31**, L19 108–L19 111.

Law, R. M., Rayner, P. J., Steele, L. and Enting, I. G. 2002. Using high temporal frequency data for CO_2 inversions. *Global Biogeochemical Cycles* **16**, 1053–1070.

Lee, T. 1987. Optimum ridge parameter selection. *Applied Statistics Series C* **36**, 112–117.

Lindley, D. and Smith, A. F. 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society B* **34**, 1–41.

Mansbridge, J. and Enting, I. G. 1986. A study of linear inversion schemes for an ocean tracer model. *Tellus* **38B**, 11–26.

Masarie, K. and Tans, P. P. 1995. Extension and integration of atmospheric carbon dioxide data into a globally consistent measurement record. *Journal of Geophysical Research* **100**, 11,593–11,610.

Mouillot, F. and Field, C. B. 2005. Fire history and the global carbon budget: A 1x1 degree fire history reconstruction for the 20th century. *Global Change Biology* **11**, 398–420.

Nordberg, L. 1982. A procedure for determination of a good ridge parameter in linear regression. *Communication in Statistics Simulation and Computation* **B11**, 285–309.

Pacala, S., Hurtt, G., Baker, D., Peylin, P., Houghton, R. A. and co-authors. 2001. Consistent land- and atmosphere-based US carbon sink estimates. *Science* **292**, 2316–2320.

Potter, S. C., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M. and co-authors. 1993. Terrestrial ecosystem production—a process model—based on global satellite and surface data. *Global Biogeochemical Cycles* **7**, 811–841.

Prentice, I. C., Heimann, M. and Stich, S. 2000. The carbon balance of the terrestrial biosphere: Ecosystem models and atmospheric observations. *Ecological Applications* **10**, 1553–1573.

Ramankutty, N. and Foley, J. 1999. Estimating historical changes in global land cover: Croplands from 1700–1992. *Global Biogeochemical Cycles* **13**, 997–1027.

Randerson, J. T., Still, C. J., Balle, J. J., Fung, I. Y., Doney, S. C. and co-authors. 2002. Carbon isotope discrimination of arctic and boreal biomes inferred from remote atmospheric measurements and a biosphere-atmosphere model. *Global Biogeochemical Cycles* **16**, 1028–1028.

Rayner, P. J., Enting, I. G., Francey, R. J. and Lagenfelds, R. 1999. Reconstructing the recent carbon cycle from atmospheric CO2, $\delta^{13}C$ and O_2/N_2 observations. *Tellus* **51B**(2), 213–232.

Rudy, Y. and Messinger-Rapport, B. 1988. The inverse problem in electrocardiography—solutions in terms of epicardial potentials. *CRC Critical Reviews in Biomedical Engineering* **16**, 215–268.

Schimel, D., House, J., Hibbard, K., Bousquet, P., Ciais, P. and co-authors. 2001. Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**, 169–172.

Takahashi, T., Sutherland, S., Sweeney, C., Poisson, A., Metzl, N. and co-authors. 2002. Global sea-air CO_2 flux based on climatological surface ocean pCO_2, and seasonal biological and temperature effects. *Deep-sea Research Part II—Topical Studies in Oceanography* **49**, 1601–1622.

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

Tarantola, A. 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, New York, 613.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B* **58**, 267–288.

UNFCCC. 1992. *Kyoto Protocol*. http://unfccc.int/resource/docs/convkp/kpeng.html

Wold, H. 1975. Soft modeling by latent variables: the nonlinear iterative partial least squares (NIPALS) approach. *Perspectives in Probability and Statistics, In Honor of M. S. Bartlett*, 117–144.

Yan, G., Graham, E. and Furlong, K. 1989. Lateral variations inferred from 3-dimensional seismic inversion models. *Geophysical Research Letters* **16**, 449–452.