Assimilating Atmospheric data into a Terrestrial Biosphere Model: A case study of the seasonal cycle

T. Kaminski[1], W. Knorr and M. Heimann
Max Planck Institut für Biogeochemie, Jena, Germany

P. J. Rayner
CSIRO Atmospheric Research, Aspendale, Victoria, Australia

Short title: ASSIMILATING CO_2 IN BIOSPHERE MODELS

[1]now at Fastopt, Hamburg, Germany

Abstract. This paper demonstrates a new method of assimilating atmospheric concentration data into terrestrial biosphere models. Using a combination of adjoint and tangent linear models of both the underlying biosphere model and atmospheric transport models we directly infer optimal model parameters and their uncertainties. We also compute biospheric fluxes and their uncertainties arising from these parameters. We demonstrate the method using the Simple Diagnostic Biosphere Model (SDBM) and data on the seasonal cycle of CO_2 from 41 observing sites. The light use efficiency for several biomes is well-constrained by concentration observations. Optimal values generally increase with latitude as required to match the seasonal cycle. Q_{10} values are poorly constrained unless local flux measurements are also used. Values also increase with latitude but are less than the commonly assumed value of 2. Fluxes computed from optimal parameters are generally more certain than from direct flux inversions, reflecting the information added by the process model.

1. Introduction

The rising concentration of CO_2 [e.g. *Keeling et al.*, 1995] and its potential impact on climate are topics of substantial scientific and policy interest. Over the last two decades, much effort has been focussed on understanding the processes which control the concentration of this gas in the atmosphere, as a possible prelude to managing the concentration. One tool commonly employed is atmospheric transport inversions [e.g. *Keeling et al.*, 1989; *Enting et al.*, 1995; *Rayner et al.*, 1999]. These studies seek to infer the net carbon exchange with the surface averaged over some time and space resolution, often continental or semi-hemispheric in spatial extent and at its finest at monthly time-scales. *Kaminski et al.* [1999b] is slightly different in that it infers fluxes at the resolution of the underlying transport model. As pointed out by *Kaminski et al.* [2001], this avoids a series of problems associated with trying to infer only large-scale fluxes.

The relative sparsity of observations (maybe 100 monthly mean observations) and the poor sampling of large regions of the globe (e.g. tropical continents) means that these inversions are badly under-determined. This under-determinacy manifests itself as large uncertainty on individual regional fluxes. This is revealed by the high-resolution study of *Kaminski et al.* [1999b] in which the number of estimated spatial flux components increases from approximately 20 to 864. Both because of the number of unknowns, and the diffusive nature of atmospheric transport, the uncertainty generally grows at finer resolution.

It is still true that the $7.8^\circ \times 10^\circ$ resolution of *Kaminski et al.* [1999b] is too coarse for many policy applications, such as national carbon accounting. Hence if progress is to be made we need work on two fronts. First is to open new streams of data, preferably guided so as to optimally improve our flux estimates. This has been approached incrementally by *Gloor et al.* [2000] with suggested expansions to the surface network, or more speculatively by *Rayner and O'Brien* [2001] with a suggested role for remotely-sensed concentration.

The second obvious method is to try to increase the leverage of existing data. At its most abstract, this involves introducing a number of ancillary relationships among the unknown flux components so that a given piece of data gains expanded influence. Naturally such an approach is limited by the veracity of these ancillary relationships. At this level of abstraction the approach is not at all novel, it is embodied in the choice of large (usually contiguous) regions which we describe via a single flux magnitude. This means that a point far from an observing site may have its flux determined by that site even though no signal from the region ever reaches the site. This can be looked upon as a distortion of atmospheric transport and, as pointed out by *Kaminski et al.* [2001], can introduce serious errors into the calculated fluxes. One reason for these errors is that spatial contiguity and, e.g., a uniform flux distribution (chosen, e.g. by *Rayner et al.* [1999]) is an arbitrary choice. A better idea is to choose relationships among flux components consistent with the natural system. Such relationships are part of the task of models of surface processes.

Again, at a sufficient level of abstraction, such models consist of a series of variables describing the state of the system (such as size of carbon pools) and a series of equations describing the evolution of that state. These evolution equations will usually contain parameters and the evolution of the state will depend on the values of these parameters as well as external forcing. Furthermore any model will calculate quantities which may be observed, albeit indirectly. An example of such an observable is a surface flux from a terrestrial biosphere model. If we can use the observables to estimate the state at a particular time and parameters of the system then, subject to the veracity of the model equations themselves, we can describe the physical system.

Such an approach was briefly sketched by *Rayner* [2000]. As mentioned there it has a number of advantages. Firstly the description is likely to be more compact than a high-resolution description of many independent flux components. Hence, depending on the independent observability of the parameters and state variables, such a model is a powerful constraint in the inverse problem so that its drastic under-determinacy may be ameliorated or even avoided.

The second advantage is perhaps more important. In general, there has been a separation between the kinds of inferences made by atmospheric inversions and those from process models. This has been manifest as a so-called scale-gap, where large-scale atmospheric inversions can only make usefully certain inferences at very large scales while process models can only be constrained by fairly small scale observations (flux towers and the like). It has been difficult for the two communities to even disagree meaningfully. The approach outlined here may allow both kinds of estimate to infer the same kinds of quantity.

The third advantage is to expand the list of observables treated by the inversion process. The atmospheric inversions carried out thus far have used two types of input, atmospheric constituent concentrations (including isotopes) and estimates of large-scale fluxes (e.g. from fossil-fuel combustion). These are optimally combined to produce improved estimates, usually for flux magnitudes. This ignores a large number of observables of the underlying processes, e.g. local fluxes, satellite radiances, biophysical variables etc. Combinations of satellite-derived vegetation indices and atmospheric concentrations have been previously used by *Knorr and Heimann* [2001b] to check mutual consistency. The framework outlined here should allow the use of such multiple constraints in a traditional data assimilation mode.

The final advantage of this approach is more ambiguous. In general the estimation of model parameters avoids the distinction between prognostic and diagnostic investigations of the system. The flux estimates of previous inversion studies are purely diagnostic. They do not, of themselves, estimate quantities which will aid prediction. Process model studies, on the other hand, are largely prognostic, only the task of tuning these models involves a diagnosis of the current state. The approach of estimating model parameters and state variables should overcome this difference. Of course, if the underlying model is wrong, the inferred parameters, although producing an optimal match to current data, will yield meaningless predictions.

The approach of estimating underlying parameters in process models from observations is not itself new. *Fung et al.* [1987] used atmospheric CO_2 concentrations to cross-check their model for inferring terrestrial productivity from satellite observations, an approach updated in the study of *Knorr and Heimann* [2001b]. A new element of this study is that the satellite data themselves are used to constrain estimates of model parameters, a method further explored by *Knorr and Schulz* [2001] and *Knorr and Lakshmi* [2001]. *Knorr and Heimann* [1995] used atmospheric CO_2 concentrations to tune parameters in a terrestrial biosphere model while *Balkanski et al.* [1999] used the seasonal cycle of atmospheric O_2/N_2 ratios to tune parameters for an ocean biology model. Most recently, *Randerson et al.* [2001] used both CO_2 and $\delta^{13}C$ observations to constrain their values of isotopic discrimination at high latitudes. In general these studies have adjusted the parameters ad hoc until they got a reasonable match with the concentration data. This is the same development path of flux inversions, the landmark studies by *Keeling et al.* [1989] and *Tans et al.* [1990] used ad hoc parameter adjustment. Later, the studies of *Enting et al.* [1995] introduced systematic methods for optimal parameter estimation. Most importantly, these systematic methods also allowed the calculation of the uncertainty of the flux magnitudes. We intend here to bring a similarly systematic approach to the estimation of model parameters, particularly a calculation, albeit incomplete, of their uncertainties.

The outline of the rest of the paper is as follows: Section 2 describes the methodology in detail. Section 3 describes the underlying terrestrial model and atmospheric transport model, briefly since they are both described elsewhere in the literature. Section 4 describes the results of the estimation of model parameters from a moderate-sized network of concentration observations. Section 5 tests some of the assumptions in our method, particularly the various linearizations for the uncertainty calculations. Finally we sketch the application of this method to more challenging problems.

2. Assimilation methodology

The aim of our calculation is to infer some information, (an estimate of the mean and uncertainty) about a set of parameters $\mathbf{p}$ from some atmospheric concentration observations $\mathbf{c}$. This involves the propagation of information in an inverse sense through rather a chain of models. This information flow is shown in a forward sense in Fig. 1. Starting from the Fig. 1

parameters themselves there is first a terrestrial biosphere model B which produces among other things a space-time distribution of modelled fluxes $\mathbf{f}_M$. Next is some kind of aggregation or interpolation process to take the gridded fluxes and map them onto the relevant grid for the atmospheric transport model. We denote this as $\mathbf{A}$. Although this is not important in the abstract description of the problem, it is an important benefit of the method that small-scale forcing of the biosphere (e.g. the $0.5 \times 0.5°$ grid of SDBM) can propagate through to the atmospheric concentrations. If we were limited to the rather coarse transport model grid (about $8 \times 10°$ in our case) some of this forcing would be lost.

Finally there is the atmospheric transport model which takes fluxes on the transport model grid and maps them into atmospheric concentrations. We denote this operator as $\mathbf{T}$. The use of the adjoint approach described by *Kaminski et al.* [1999a] means that $\mathbf{T}$ can take fluxes at the full resolution of the underlying transport model and map them onto a predetermined set of observed concentrations.

Combining the three steps we have a mapping from biosphere parameters to modelled concentrations $\mathbf{c}_M$ expressed as

$$\mathbf{c}_M = M(\mathbf{p}) = \mathbf{T} \times \mathbf{A} \times B(\mathbf{p}) \tag{1}$$

Strictly, the multiplication in the above expression represents composition of functions. $\mathbf{T}$ and $\mathbf{A}$ are linear however, so we can also consider this as matrix multiplication. We stress again that nowhere in the construction of $\mathbf{c}_M$ do we need to make assumptions about the structure of fluxes etc. apart from those implicit in the biosphere model itself.

We define an optimal vector $\mathbf{p}$ so that it minimizes the overall mismatch with observed concentrations. To this end we construct a cost function

$$J_C(\mathbf{p}) = {}^1\!/_2 \, (M(\mathbf{p}) - \mathbf{c})^T \, \mathbf{C_c}^{-1} \, (M(\mathbf{p}) - \mathbf{c}) \tag{2}$$

where $\mathbf{C_c}$ expresses our uncertainty for the observations $\mathbf{c}$ in the form of a covariance matrix. Several augmented forms of J_c are possible representing information from other points along the chain of models. For example we can add a term embodying a mismatch of $\mathbf{p}$ from some prior estimate $\mathbf{p}_0$. This is the Bayesian approach. It is normally used in flux inversions such as *Rayner et al.* [1999] partly to stabilize the estimates and partly to embody prior knowledge of fluxes. In our case this knowledge may be embodied in the model instead. Also if there are observed fluxes, $\mathbf{f}$, we can add a term such as

$$J_F(\mathbf{p}) = {}^1\!/_2 \left(N(\mathbf{p}) - \mathbf{f}\right)^T \mathbf{C_f}^{-1} \left(N(\mathbf{p}) - \mathbf{f}\right), \qquad (3)$$

where by analogy with (1) we have denoted by N the function that computes modelled fluxes $\mathbf{f_M}$ from $\mathbf{p}$. We can replace fluxes by any other observable calculated by the biosphere model.

We are now left with the computational problem of optimizing $\mathbf{p}$ in the nonlinear composite model M. Note that whatever optimization algorithm we use will involve several, perhaps many, runs of M. This is infeasible if we need to run $\mathbf{T}$ many times, because $\mathbf{T}$ is the most expensive part of our model. Since $\mathbf{T}$ is linear we can precalculate a set of response functions, the approach normally used in previous synthesis inversions. In our application this has a serious drawback. We can only calculate the response functions for a subspace of fluxes. Our choice of subspace represents a choice of a set of flux patterns, usually reflecting structure within a particular region. In our case, as we change $\mathbf{p}$ through the iterative optimization process, these shapes will change. *Kaminski et al.* [2001] showed that incorrect choices of flux shapes could seriously distort an inversion. Fortunately, *Kaminski et al.* [1999a] demonstrated the efficient calculation of the complete response function for a transport model provided we only need modelled concentrations at selected sites and times. So we can include the transport model as a simple matrix multiplication without losing the required generality of flux fields.

The calculation of $\mathbf{T}$ is based on Automatic Differentiation (AD) of the atmospheric transport model in reverse mode (*Griewank* [1989]). It is efficient, because the number of input parameters to the model (fluxes at every month and surface grid cell) considerably exceeds the number of output variables (concentrations at every month at about 50 sites). The same observation holds for the interpolation matrix $\mathbf{A}$. Hence, again, we have applied AD in reverse mode to generate the adjoint code of the interpolation routine. This adjoint code efficiently computes $\mathbf{A}$. The automatic differentiation has been carried out by the Tangent linear and Adjoint Model Compiler (TAMC, R. Giering, 2000, available at http://puddle.mit.edu/$\sim$ralf/tamc))

Powerful minimization algorithms for functions like J, rely on the availability of the gradient of J with respect to the parameters. Hence, again, we apply TAMC in reverse mode to generate a subroutine to evaluate $\nabla_{\mathbf{p}} J(\mathbf{p})$ for any $\mathbf{p}$. Unlike the adjoints of $\mathbf{T}$ and $\mathbf{A}$ which evaluate derivatives of vector valued functions, the adjoint of J evaluates the derivative of a scalar valued function. This derivative is a fairly compact expression, mapping maybe a few dozen parameters onto their impact on the cost function, even though some of the intermediate fields and hence mappings may be large.

The cost function minimization allows us to calculate an optimal parameter set but provides no information on the uncertainty in those parameters. Some information can be obtained from the value of the second derivative or Hessian at the minimum. Geometrically, this is the curvature of the cost function. A subroutine for evaluation of the Hessian can be obtained by reapplying TAMC to the adjoint. It is slightly more efficient to use the forward mode for this second pass, because the function to be differentiated has as many input as output variables (parameters and components of the gradient, respectively).

The way the cost function defined by Eqs. 2 and 3 is set up reflects the assumption of Gaussian probability distributions for the observed concentrations, fluxes and the a priori information about the parameters (the Gaussian assumption). If our model was linear, one could show [see e.g. *Tarantola*, 1987, Eqs. 4.4–4.6] that the posterior probability distribution is Gaussian as well and that the uncertainties in the optimal parameters are quantified by a covariance matrix which is the inverse of the Hessian at the minimum. Since the model is nonlinear, uncertainties derived this way are merely an approximation (the locality approximation). In this work we yet make another approximation (linearity approximation): We apply TAMC in forward mode to compute the linearization or Jacobian matrix $\mathbf{J}$ of the model. Since there are fewer parameters than observations, the forward mode is computationally preferable. In this linear approximation, the posterior covariance matrix $\mathbf{C}_p$ is given by

$$\tilde{\mathbf{C}}_p^{-1} = (\nabla_{\mathbf{p}} M)\mathbf{C}_c^{-1}(\nabla_{\mathbf{p}} M)^T + (\nabla_{\mathbf{p}} N)\mathbf{C}_f^{-1}(\nabla_{\mathbf{p}} N)^T + \mathbf{C}_{0,p}^{-1} \tag{4}$$

Where $\mathbf{C}_{0,p}$ expresses the a priori uncertainty in parameters. The inverse can be computed, e.g., by a Singular Value Decomposition (SVD). We have described the formula and the SVD in *Rayner et al.* [1999] and *Kaminski et al.* [1999b] but all the background is to be found in *Tarantola* [1987]. We shall check the locality approximation and the linearity approximation by comparing the cost function generated by varying the parameters around their optimal values with the cost function implied by our calculated parameter uncertainties.

As a consequence of our approximations, the locality approximation has an important consequence for the calculated uncertainty. As an example consider a cost function which is quadratic in the region of the optimal values but contains a large potential barrier at zero, effectively disallowing negative values. This might occur if we wish to prevent an optimization routine taking logarithms of negative numbers. Our local analysis of uncertainties at the cost function minimum will never, even in principle, see this constraint. Even with this limitation, the calculation of uncertainties still represents a substantial advance on previous methods.

In the final step of the process, the calculation of fluxes and their uncertainties, we use the optimal values from the first step and the covariance matrix from the second. The flux calculation requires only a forward run of the biosphere model, although to reduce storage requirements we can use the aggregation step to output fluxes on the transport model grid. Note that even here, the covariance matrix representing the uncertainty in flux estimates is very large, 10,368$\times$10,368. Fortunately we can store a compact expression for the square root of this matrix and generate relevant fields off-line.

It is worth reiterating the centrality of the automatic differentiation tool TAMC in this whole process. In fact it is used five times. First we use it to calculate the aggregation operator

$\mathbf{A}$. Next we use the matrix representation of the transport model TM2 calculated by *Kaminski et al.* [1999a]. The optimization step uses the derivative of the cost function with respect to input parameters, which is provided by the adjoint of the composite model. The calculation of approximate parameter uncertainties uses the Jacobian of concentrations with respect to biosphere parameters, generated as a tangent linear model to the forward model. Finally, the calculation of flux uncertainties uses the Jacobian of fluxes with respect to parameters, again a tangent linear model but this time leaving out the transport model step.

3. Models and Data

3.1. Biosphere Model

We use the Simple Diagnostic Biosphere Model (SDBM) described by *Knorr and Heimann* [1995]. In brief, the model calculates a pure seasonal cycle of net biosphere-atmosphere fluxes as the difference of two fluxes in any month at each $0.5^\circ \times 0.5^\circ$ land grid cell:

$$\mathrm{F} = \mathrm{H} - \mathrm{NPP} \tag{5}$$

F is the net flux to the atmosphere, H is the heterotrophic respiration and NPP is the net primary productivity.

The calculation is performed on the basis of measured Normalized Differential Vegetation Index (NDVI), incoming solar radiation, and surface temperature. There is an additional time-dependent water-stress term α, which may limit the efficiency either of NPP, or heterotrophic respiration, or both. This factor will be included in various of the calculations in this work in either of the two fluxes H and NPP; in the case described below, it is included in both.

NPP is thus calculated for each month as

$$\mathrm{NPP} = \alpha\epsilon\mathrm{APAR} \tag{6}$$

where APAR is absorbed photosynthetically active radiation, and ϵ is a light-use efficiency parameter. APAR is computed from NDVI data by Gallow [1992] and incoming solar radiation inferred from cloudiness data by *Leemans and Cramer* [1991] using the method of *Linacre* [1986].

Heterotrophic respiration H is calculated as an exponential function of temperature as

$$\mathrm{H} = \alpha\epsilon \mathrm{H}_0 Q_{10}^{T/10} \tag{7}$$

where H_0/ϵ is the heterotrophic respiration rate at $T = 0$ and $\alpha = 1$ (i.e. no water stress) and Q_{10} is the ratio of respiration at $T + 10$ to that at T; T is given in °C. We have absorbed ϵ into the formulation of H for later convenience.

In order to ensure a balanced biosphere (no net flux) we rescale H_0 at each grid cell so that integrated NPP and H are equal over a year. Thus H_0 is not a free parameter.

We can now write the total flux as

$$\mathrm{F}(t) = \alpha\epsilon \left(\mathrm{H}_0 Q_{10}^{T(t)/10} - \mathrm{APAR}(t) \right) \tag{8}$$

Subject to the constraint that

$$\mathrm{H}_0 \int Q_{10}^{T(t)/10} dt = \int \mathrm{APAR}(t) dt \tag{9}$$

where integrals are taken over the annual cycle. Integrals are calculated as sums of monthly values weighted by the length of the month.

3.2. Transport Model

We use the adjoint form of the transport model TM2. The base model is described in *Heimann* [1995]. The model is an off-line tracer transport model with a 7.8°latitude × 10 °longitude grid and nine vertical levels. It is driven by analyzed winds from the European Centre for Medium Range Weather Forecasting (ECMWF) in this case from the year 1986. The model features vertical transport by convection and turbulent eddy transport following the schemes of *Louis* [1979] and *Tiedtke* [1989] respectively.

TM2 was a participant in the Transport Model Comparison (TransCom) project reported by *Rayner and Law* [1995] and *Law et al.* [1996]. Of most importance here was its simulation of the seasonal cycle since the amplitude and phase of the seasonal cycle are the critical parameters in our study. In general, the simulated seasonal amplitudes lie within the range of other model simulations. One different emphasis of this model is the relatively large amplitudes over tropical regions compared to those over higher latitude regions. We should bear this in mind when considering the distribution of NPP between low and high latitudes in our optimizations. We must stress, of course, that the range of model simulations does not necessarily provide guidance for the correct atmospheric transport behaviour.

The adjoint form of the model was constructed by *Kaminski et al.* [1999a] and used in a previous flux inversion by *Kaminski et al.* [1999b]. The adjoint model calculates the derivative of the monthly mean concentration at a specified set of observing sites with respect to emissions in all surface gridpoints and all months. Since simulated concentrations are linear in emissions, this derivative need only be calculated once. The transport model can hence, once we restrict ourselves to a given observing network, be replaced by a single matrix multiplication. As we have already mentioned, this form of the transport model allows the full structure of fluxes upto the resolution of the transport model to be mapped onto concentrations. This is an important advantage in a study like ours since changes in parameters may cause small-scale changes in the structure of fluxes.

3.3. Data

We use the compilations of the seasonal cycle from *GLOBALVIEW-CO_2* [2000]. We use a network of 41 stations. The network we use is shown in Fig. 2 and listed in Table 2. In the table we also show the RMS uncertainty on the data from each station and the RMS error for the fit with our control optimization.

Fig. 2

Table 2

One obvious property of the network is its oceanic focus. This is clearly undesirable for our current study and one of our aims will be to investigate the knowledge available from such a network for constraining biospheric processes. We have removed the oceanic bias somewhat by eliminating the sampling locations from ship cruises in the Pacific Ocean and South China Sea.

As well as assigning data values we need to assign an uncertainty for weighting the data in Eq. 2. We calculate the data uncertainty from the two uncertainty estimates provided in *GLOBALVIEW-*CO_2 [2000]. These uncertainties are calculated from the residual in fitting the smoothed curves used to construct their timeseries to the raw flask data, and also to the interannual variance in the fitted curves themselves.

While a detailed discussion of the assignment of data uncertainty in synthesis inversions is beyond us here we note that these uncertainty estimates are really a proxy for a range of uncertainties which must be included in our calculation. These include

- Imprecision in the measurements themselves
- The ability to simulate concentrations measured at a point and time with monthly mean model concentrations on a $7.8^\circ \times 10^\circ$ grid.
- Inadequacies in the biosphere or transport model including their driving data and source heterogeneity not captured by the model.
- Uncertainties in the other contributions to the seasonal cycle which are not included in our statistical model.

All these uncertainties are hard to quantify. In particular, uncertainties from source heterogeneity depend critically on the choice of structures which can be resolved by the model, as demonstrated by *Kaminski et al.* [2001]. In general the data uncertainties from *GLOBALVIEW-*CO_2 [2000] have the right structure, with larger errors in the vicinity of large and heterogeneous sources.

In our calculation we quadratically sum the two components of the Globalview uncertainty since we need to take both into account. Finally to deal with potential model inadequacies we require a minimum value of 0.70 ppmv. This minimum value is important since 328 of the 452 values used in our standard case would otherwise have lower uncertainties. Given the contributions to the uncertainty for our need that cannot be captured by the analysis of *GLOBALVIEW-*CO_2 [2000] this represents a reasonable minimum value.

3.4. Background Fluxes

Terrestrial biospheric activity, particularly the seasonal cycle, is focused on the large land masses of the northern hemisphere and tropics. Even here, and particularly in the southern extratropics, the other major carbon fluxes may make a significant contribution to the seasonal cycle of CO_2 concentration. Even seasonality in transport acting on a time-invariant source field may produce substantial seasonality in concentration. We include three such contributions in our study. These extra fluxes should be considered part of the driving data for the problem, that is we do not optimize them in any way. Note that the seasonal cycle is the main data which forces our results, so contributions to seasonality are more important than overall magnitude. We must factor uncertainties in the contribution of these fluxes to seasonality of concentration into uncertainties in the data.

We use the annually invariant flux from fossil fuel combustion and cement production compiled by *Andres et al.* [1996]. The seasonality in concentration arising from this flux is largest at low latitudes, peaking around 1 ppmv at Seychelles in our study. Thus errors in its seasonality are unlikely to make a large impact in predicted seasonal cycles and hence our results.

We use the estimate of ocean CO_2 flux compiled by *Takahashi et al.* [1999] derived from a compilation of air-sea δpCO_2 values and using the air-sea gas exchange formulation of *Wanninkhof* [1992]. This field contains moderate seasonality and in the southern hemisphere is likely to be a major contributor to seasonality in the concentration records. However, in the northern hemisphere and tropics, its maximum values around 1 ppmv are only a small contribution to the observed seasonal cycle. So, while there may be substantial sampling biases in the seasonality of the flux estimates, these probably do not contribute much to the overall seasonality in concentration.

Finally, we use the estimate of flux due to land-use change compiled by *Houghton et al.* [1990]. Strictly this considers only the source due to deforestation rather than land abandonment. It is focused in low latitudes. Most importantly for our study we treat it as annually invariant whereas the burning events which comprise this flux largely occur in the tropical dry season in reality. Thus there may be a significant error in seasonality in this flux and we should consider this when interpreting results for low-latitude biomes.

3.5. Implementation Details

In this work we aim to estimate the two parameters ϵ and Q_{10} in SDBM for each of twelve biomes. The biomes are taken from *DeFries and Townshend* [1994] and their distribution is shown in Fig. 3. Abbreviations used throughout the paper for each biome are listed in Table 1. Fig. 3
This generates 24 unknowns of interest. As well, we need to solve for a mean value at each of Table 1
the stations in our chosen observational network. Although we use the pure seasonal cycles from *GLOBALVIEW-CO*$_2$ [2000], and SDBM generates a pure seasonal cycle, the composite model will compute nonzero mean values for two reasons. Firstly, we must consider the contributions to the seasonal cycle made by fossil-fuel combustion and ocean fluxes, these are input as known fluxes. Second the covariance of transport and seasonal biospheric fluxes will of itself produce a nonzero mean value, as demonstrated by *Keeling et al.* [1989], *Denning et al.* [1995], and *Law et al.* [1996]. To avoid this mismatch producing erroneous corrections to biosphere parameters, we use a free-floating offset to match the mean value. Thus we have one extra unknown per station in the observing network. These offsets are only important in the optimization step and we neglect them later.

Throughout the following calculations we use a Bayesian estimation procedure, i.e we include prior constraints on the biosphere parameters but not on the offsets. Prior uncertainties for the ϵ are set at 50% of their initial value while those for Q_{10} are set at 30% of their prior value, as shown in Figs. 4 and 5.

4. Results

In this section we will discuss two major aspects of the results of optimizations using the framework described above. First we will show some of the estimated parameters and fluxes produced by our control case and some variations of it. Then we will discuss the uncertainties on these parameters, with a view to recovering the information on biospheric processes available from atmospheric concentration observations.

4.1. Control Case

For our control case we choose an inversion with the water-stress parameter α turned on for both the NPP and respiration terms. Fig. 4 and Fig. 5 show the prior and predicted values of ϵ and Q_{10} for each biome in our standard inversion. Several things are immediately apparent from these figures. Firstly the ϵ values are generally shifted further from their prior estimates than the Q_{10} parameters. This would normally be expected because of the generally larger reduction in uncertainty in ϵ, which indicates that the inversion is adding more information about ϵ than about Q_{10}. The proximate cause for this difference is the greater sensitivity of the seasonal cycle of concentrations to changes in ϵ than Q_{10}. This can be seen from the governing equations of SDBM (Eq. 8) and the nature of the forcing functions for the model. As formulated the amplitude of the seasonal cycle of flux in SDBM is governed by the value of ϵ. The phase results from the interaction of seasonal variations in temperature and water-stress with the value of Q_{10}. With the seasonality of temperature and water-stress fixed there is relatively little sensitivity of fluxes to changes in Q_{10} and hence only a weak constraint on this parameter. Fig. 4 Fig. 5

The structure of the changes to ϵ also shows some interesting features. In general, ϵ is increased for high latitude biomes. In particular the optimization suggests the high latitude deciduous forests and tundra require high light use efficiency to match seasonality in concentration. The coniferous forest parameters remain rather close to their prior values. The tundra and deciduous forest values are particularly striking since their optimized confidence interval is shifted outside or nearly outside their original confidence interval. Meanwhile the C4 grassland biome has its light-use efficiency parameter reduced almost to zero.

When considering the above results, it is useful to take into account the relative biological importance of the various biomes. A measure of this importance is the total APAR associated with each biome, or, alternatively, its total NPP. Fig. 6 lists the prior and predicted NPP for each biome. The prior uncertainties reflect the 50% prior uncertainty on ϵ. The global total is hardly changed from the prior estimate (53.8GtC to 53.5GtC). The uncertainty is reduced, from 11 GtC to 7.2 GtC. This reduction in uncertainty suggests only moderate scope for concentration measurements used alone to constrain such large-scale processes. The result is consistent with that of *Knorr and Heimann* [2001b], who found that a reduction in uncertainty of global NPP from 31% (24.1 GtC) to 25% (18.5 GtC) through the inclusion of satellite measurements was consistent with the observed seasonal cycle of CO_2. The size of the uncertainties in their work is larger both because their overall NPP is larger (74.6 GtC), and because the number of parameters that entered the uncertainty calculation is also higher. This is an expected consequence of using a process-based vegetation model, as opposed to a simplified diagnostic model as in the present work. Fig. 6

Fig. 6 shows that some of the largest changes in ϵ occur for biomes with relatively small NPP. In particular the large decrease for the C4 grassland biome and large increase for tundra do not have dramatic effects since their total NPP is relatively small. Uncertainties for ϵ for these biomes also remain close to their prior value. One reason for this is evident from Eq. 6. As the signal in the atmosphere used to constrain the model is created by the seasonal cycle of NEE, and because of the way H_0 is calculated, this signal is scales linearly with ϵ. How sensitive the seasonal cycle is to changes in ϵ, however, depends on the seasonal cycle in α (tropical biomes), in T (northern biomes), and APAR (both). As noted by *Knorr and Heimann* [1995], the seasonal cycle in the tropics is very small when α is included in both NPP and

H. This means that the seasonal cycle of concentrations that we observe is not sensitive to ϵ from the C4 grassland biome. Such a lack of sensitivity means the parameter is only weakly constrained. In fact, *Knorr and Heimann* [2001b] found that tropical trees and C4 grasses make a sizeable contribution to the seasonal cycle only south of the equator (15% and 25%, respectively), where this seasonal cycle is small.

We can convolve our estimates of ϵ with spatial maps of APAR and α to produce spatial distributions of NPP. These can be compared with the NPP fields produced by other calculations. In the Potsdam Intercomparison (*Cramer et al.* [1999]), many terrestrial modelling groups produced such estimates. Fig. 7 shows the zonal mean NPP per unit land area for four optimized cases along with the median and extreme deciles from *Kicklighter et al.* [1999]. The four optimization cases show the impact of including water-stress for either or both of NPP and respiration. We see that our standard case (on for both) lies close to the median value in the tropics but almost at the upper decile in high northern latitudes. This is consistent with the observation of *Nemry et al.* [1999] that most of the models in the Potsdam intercomparison underestimated seasonality of concentrations at high latitudes. This is also true for initial parameter settings of SDBM and the optimization increases ϵ in high latitudes to improve the match. NPP is increased as a consequence. Fig. 7

Low latitude NPP shows great sensitivity to the treatment of water-stress. While our standard case lies close to but above the median Potsdam result, all the other cases lie at or below the lowest decile. There is less sensitivity elsewhere, suggesting either that NPP is so low that the range is tightly constrained (subtropics), or that water-stress is not a serious factor in controlling NPP (high latitudes). These results are consistent with *Knorr and Heimann* [1995] who showed that the impact of water-stress on seasonal cycles of concentration could only be noticed from tropical regions.

As we have already noted, the most striking facet of Fig. 5 is the large predicted uncertainties on this value. This weak constraint occurs despite the use of biomes representing large regions; it would be worse at finer spatial resolution. Hence any comment we make on the predicted estimates must be rather cautious. One general feature that does emerge is the tendency for high values for high latitude biomes and lower values for tropical biomes. All biomes suggest a positive sensitivity of heterotrophic respiration to air temperature ($Q_{10} > 1$) consistent with the meassurements compiled by *Raich and Potter* [1995], which also refer to air temperature. When soil respiration is related to soil temperature, however, Q_{10} values are mostly higher, closer to the commonly assumed value of 2, as reported by *Raich and Schlesinger* [1992]. However, the median of the values found experimentally by *Holland et al.* [2000] for tropical soils (1.74) still lies within the confidence interval for the tropical biomes.

To study the impact of extra measurements in this calculation we repeated the standard case with the addition of some pseudoflux measurements. These fake measurements were created by archiving the fluxes for one point in the Broadleaf Evergreen biome, then recalculating the propagated uncertainties. Since the pseudoflux data is that from the optimized model itself the estimated values do not change. We assign the monthly mean pseudoflux data an uncertainty of $10gm^{-2}month^{-1}$, corresponding to an error of approximately 5% of the maximum monthly flux. This uncertainty should be regarded as an example, particularly as we have not included the representativeness of such a flux for the entire biome. Overall uncertainties for both ϵ and Q_{10} in this biome are much reduced. Fig. 8 shows the results for $Q_{10}$for five selected biomes. The reduction is particularly striking for those Q_{10} values that were not well-constrained by concentration measurements. This complimentary role suggests combined use of both types of measurements. Fig. 8

SDBM is a very simple model of seasonal carbon exchange. It is worthwhile asking, then, how well it can fit the concentration observations when its parameters are optimized. As examples, Figure Fig. 9 shows the predicted and observed seasonal cycle for four stations: Point Barrow, Alaska; Ulaan Uul, Mongolia; Mahe Island, Seychelles; and Cape Grim Tasmania. The fine lines either side of the observed line show the uncertainty used on each monthly datapoint. Barrow is one of the worst simulated stations. It's large seasonal cycle, of course, provides some context for this error. The general phase and amplitude of the seasonal Fig. 9

cycle are well simulated. The clearest error is an over-estimate of the source during Spring. Notwithstanding this, the optimized and even unoptimized model produces more accurate seasonality at high latitudes than many models with more sophisticated biogeochemistry but less diagnostic information [see e.g. *Nemry et al.*, 1999].

At Ulaan Uul, again, the general form of the seasonal cycle is well-simulated, the only error being a slight over-estimate of the amplitude. The much smaller seasonal cycle at Seychelles is also fairly well simulated. Errors are larger as a proportion of the seasonal cycle but the data uncertainties are also larger. Seasonality at Seychelles is driven by combinations of seasonality from nearby land and complex modulations of large-scale transport such as monsoon circulations and the movement of the intertropical convergence zone (ITCZ). Simulations with only the background (fossil and ocean) fluxes suggest these may contribute as much to the seasonal cycle at Seychelles as the biosphere, which suggests both that our conservative choice of data uncertainties is wise and that this is not a good site to try to constrain tropical biomes.

Finally, the even smaller seasonality evident at Cape Grim is fairly well simulated by the model. The amplitude is roughly correct but the model leads the observations by perhaps two months. Here there might be a case for much tighter uncertainty bounds on the observations but the uncertain contribution of the background fluxes (the ocean in particular here) would support our more conservative choice. *Law* [1996] also showed considerable sensitivity of the phase of the seasonal cycle at Cape Grim to methods used to select clean or base-line air, already pointed out in the original analysis with SDBM by *Knorr and Heimann* [1995].

We can calculate a global measure of the quality of the optimal simulation as the mean-square mismatch or χ^2 value

$$\chi^2 = \frac{1}{N}\Sigma_{i=1}^{N}(\frac{C_{M,i} - C_{O,i}}{\sigma_i})^2 = J_c\frac{2}{N} \tag{10}$$

where $C_{M,i}$ is modelled concentration, $C_{O,i}$ is observed concentration and σ_i is the prior uncertainty on the observed concentration. N is the total number of concentration observations. In general, if our statistical assumptions are valid χ^2 will be around 1. At least it should not be greater than 1 or we should suspect that our prior uncertainties on data are too tight.

The χ^2 value gives us a convenient summary measure of the quality of a given optimization. Table 3 shows this mismatch along with the global NPP and NPP for selected biomes from the four optimization cases from Fig. 7. The table reinforces the behaviour seen in the figure, notably that NPP is much lower for the tropical biomes in our alternative cases than the control. This lowers the global NPP also. It is interesting that applying the water-stress term to respiration (either with or without considering its impact on NPP) has a substantial effect on optimized NPP values. This occurs despite the assumption in the model that, at every point, annual respiration is adjusted to balance NPP. A similar result was previously noted by *Lloyd and Farquhar* [1994] and *Knorr and Heimann* [1995] for Q_{10}. A larger Q_{10} value, they noted, increases the proportion of respiration in summer and hence requires a larger NPP to produce the large draw–down in summer concentrations. In our case, water-stress reduces respiration in the dry season so requires a smaller value of ϵ to avoid a large draw-down. The mismatch is not very sensitive to the presence of α although we see that, as suggested by *Knorr and Heimann* [1995], the match is worst when water-stress is considered to impact NPP but not respiration. Table 3

As a final example, Fig. 10 and Fig. 11 show the terrestrial biospheric fluxes for July and their uncertainties as calculated from the optimal parameter values and their uncertainties. The overall structure of the fluxes is as one might expect for July, with large uptake over the high latitude continents decreasing at lower latitudes and, generally, sources in the southern hemisphere. The uncertainty in fluxes generally scales with the flux, suggesting that uncertainties in ϵ are the main contributor to flux uncertainty, at least for this month. There is also a role for correlations between parameters in reducing uncertainties on net flux. This role is only possible because of the propagation of the complete covariance structure through the calculation. Fig. 10 Fig. 11

5. Discussion

Given the extreme simplicity of SDBM, it is questionable how much weight to put on the particular results of our optimization. The ability to simulate the observed data, while a necessary condition for the use of the model, should not be taken as a validation of the optimized values. We have not performed a detailed cross-validation study here but this would be required before we placed great weight on the results. As we have already stated, the technique is an attempt to build process understanding explicitly into atmospheric inversions; if the process understanding is over-simplified the conclusions will be correspondingly erroneous.

All this having been said, some of the general structure of the results is interesting. The overall value of Q_{10} is lower than studies relating soil respiration to soil temperature [e.g. *Raich and Schlesinger*, 1992], but similar to those that use air temperature [e.g. *Raich and Potter*, 1995]. This difference has also been noted by other studies using atmospheric forcing to drive terrestrial models [e.g. *Maisongrande et al.*, 1997; *Randerson et al.*, 2001]. It most likely results from the difference between soil and air temperatures. The dominant process in heterotrophic respiration, respiration from within soils, presumably reacts to soil temperature. Soil tends to damp the large variations in air temperature, either diurnally or seasonally. Snow has a particularly strong effect on the seasonal variations. Thus the apparent sensitivity of respiration to air temperature is lower than it would be for soil temperature.

The latitudinal structure of Q_{10} is also interesting, although the uncertainties make strong conclusions unwise. In general, the above argument would suggest a closer match between soil-temperature and air-temperature derived values of Q_{10} at low latitudes than at high latitudes, since air temperature variations are so much higher at high latitudes. We, however, observe higher values at high latitudes than low in agreement with *Maisongrande et al.* [1997]. Our result also agrees with the finding by *Holland et al.* [2000] that many tropical soils show values below 2, less than the typical values for northern latitudes cited by *Raich and Schlesinger* [1992].

We have already noted a similar gradient in optimized values of ϵ with latitude. The structure is clearly at variance with the consensus understanding of terrestrial biological models as shown in Fig. 7. There are several potential explanations for this. One possibility is the impact of processes that may affect the seasonal cycle but are not included in the model. These include seasonality in disturbance fluxes. Most of these (e.g. fires) would be expected to counteract the strong summer draw-down of atmospheric CO_2 and hence require even larger values of ϵ to match the observed seasonal cycle of concentrations. More likely is the model's assumption about the relationship between seasonal and annual mean production. Here these are governed by the one parameter ϵ. Increasing ϵ to match the seasonal cycle of concentration necessarily increases annual NPP. Possibilities like reducing minimum NPP to achieve the same increase in seasonality are not available to the model. Similarly, annual mean NPP depends on the relationship between annual mean APAR and its seasonality. For example, the seasonality in NPP is unaffected by adding a constant value to every monthly APAR value but the annual mean NPP will change. The effect on the monthly net flux is complicated by the constraint of a balanced biosphere.

There is a more fundamental issue facing the approach outlined in this paper which transcends any particular results. It concerns the propagation of constraints from one series of measurements across large and spatially separate regions. This is a common concern in terrestrial biological modelling often (though confusingly) referred to as the scaling problem. In our calculations the issue is most starkly demonstrated by the influence of flux measurements taken at a specific point throughout an entire biome. Recall that a fictitious measurement on a single $0.5^\circ \times 0.5^\circ$ gridcell produced a substantial reduction in uncertainty in productivity for an entire large biome. This is inherent in the use of single parameters to describe entire biomes. The idea is challenged by a recent study by *Wang and Leuning* [2001] who optimized parameters in their biophysical model against campaign-mode flux observations taken at several apparently similar sites. They calculated large variations in photosynthetic capacity challenging the very idea of applying parameters on large scales. We should note that the model used in that study did not consider differences in nutrient status between the various sites so that in some sense their model was not complete.

This objection is certainly serious but has implications far beyond inversion studies like this. However it is stated, the aim of local measurements is to garner process understanding and apply it more widely. If this generalization is in principle impossible then a major motivation for such measurements is invalid. In its broadest sense the inability to translate understanding from one place to another represents model failure. Whether this failure is inevitable and insoluble remains an open question. The framework outlined here provides a method for testing such questions since local or process-based estimates can be generalized and rigorously compared with integral constraints such as atmospheric concentration observations.

Finally there are some aspects of the uncertainty calculations shown throughout this paper which require further comment. These concern the propagation of uncertainty through our linearized model to yield multivariate normal distributions for our inferred parameters from multivariate normal distributions of our observables. This is an inherent limitation of our analysis. As noted in section 2, we make the linearity approximation of calculating the derivitave of the cost function via the derivative of concentrations with respect to parameters rather than by direct calculation of the Hessian. We can test this approximation by comparing the cost function as a function of one parameter to the ideal cost function assuming the normal distribution computed by linear error propagation. Fig. 12 shows an example of such a comparison, for Q_{10} for the broadleaf evergreen biome. We note good agreement in the curvature around the minimum. This agreement suggests that our uncertainty estimate is locally correct, i.e. that our linearity approximation is a good one. Fig. 12 also provides a test of our locality approximation. Were the real cost function purely quadratic (corresponding to a Gaussian posterior distribution) this agreement would not only hold near the minimum, but across all values. However we see a more rapid increase in cost for low Q_{10} values in the real than the ideal cost function, reflecting that our ideal probability distribution overestimates the probability of low Q_{10} values. This disagreement might partly explain the relatively low Q_{10}

Fig. 12

values seen in this study since the *mean* Q_{10}value from the real probability distribution would be larger than from the ideal.

The simplicity of the problem studied in this paper is mainly driven by the simplicity of the terrestrial biosphere model we employ. The simplicity manifests itself in two important aspects. Firstly the model calculates only the seasonal cycle of fluxes. With the dominant focus of research in the carbon cycle now shifting to understanding the climate sensitivity of the underlying source processes, there is an obvious opening to repeat this work with a model capable of simulating interannual variability of fluxes. In particular it would be interesting to see how much confidence in predictions of terrestrial biosphere responses to climate change can be improved by parameter calibrations against the historical record. Another consequence of the simplicity of the biosphere model is that prior uncertainties only enter two parameters, although in a more realistic, process-based model, far more parameters would have to be specified. This has been explored in detail by *Knorr and Heimann* [2001a], and the result is an approximate doubling of uncertainties in global annual NPP compared to this study.

The other important limitation of this model is its limited set of observables. This meant we could not do a real comparison of local and integral measures. The work of *Knorr and Heimann* [2001b] suggests that other forms of data may constrain terrestrial biosphere models more strongly than the atmospheric concentrations used here. Models which simulated, rather than were driven by, absorbed photosynthetically active radiation as well as fluxes and stock changes might yield a much more stringent test of whether the framework outlined here is useful in real situations. The use of such satellite-derived data has been investigated by *Knorr and Schulz* [2001] and *Knorr and Lakshmi* [2001] and we intend to combine that approach with the concentration assimilation we have used here.

6. Conclusions

We have outlined and demonstrated an assimilation framework for combining atmospheric concentration measurements with local measurements in constraining a terrestrial biosphere model. Using the Simple Diagnostic Biosphere Model and the techniques of automatic differentiation, we have been able to constrain the Net Primary Productivity globally and in most major biomes using the seasonal cycle of atmospheric concentration. The results generally show higher light use efficiencies in high latitudes than low latitudes, a consequence of the unoptimized model's failure to match the seasonal cycle in concentration at high latitudes. Predicted NPP for the base case is close to the median of the Potsdam intercomparison at low latitudes and higher at high latitudes. Low latitude NPP is sensitive to the treatment of water-stress in SDBM with 3 variations that we tested showing substantially lower NPP than our base case. The sensitivity of heterotrophic respiration to temperature is poorly-constrained by concentration measurements but better constrained by local flux pseudodata. Finally the calculated uncertainties on fluxes for July suggest that the method may be useful in constraining tropical terrestrial fluxes from the sparse observing network. This constraint relies on using the process model to extrapolate information gained in one region throughout large biomes. This is a more soundly based approach than the previous flux-based inversions but is subject to the veracity of the terrestrial model.

Acknowledgments.

This study was carried out with the support of the Bundesministerium für Bildung und Forschung (BMBF) contract 01LA9898/9.

The authors wish to thank Ying-Ping Wang and Ian Enting for helpful comments on the manuscript.

References

Andres, R. J., G. Marland, I. Fung, and E. Matthews, A $1^\circ \times 1^\circ$ distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950-1990, *Global Biogeochem. Cyc.*, *10*, 419–429, 1996.

Balkanski, Y., P. Monfray, M. Battle, and M. Heimann, Ocean primary production derived from satellite data: An evaluation with atmospheric oxygen measurements, *Global Biogeochem. Cycles*, *13*, 257–271, 1999.

Cramer, W., D. W. Kicklighter, A. Bondeau, B. I. Moore, G. Churkina, A. Ruimy, A. L. Schloss, J. Kaduk, and the participants of the Potsdam NPP Model Intercomparison, Comparing global models of terrestrial net primary productivity (npp): overview and key results, *Glob. Change Biol.*, *5*, 1–15, 1999.

DeFries, R. S., and J. R. G. Townshend, NDVI-derived land cover classifications at a global scale, *Int. J. Remote Sens.*, *15*, 3567–3586, 1994.

Denning, A. S., I. Y. Fung, and D. A. Randall, Gradient of atmospheric CO_2 due to seasonal exchange with land biota, *Nature*, *376*, 240–243, 1995.

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

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

Gallow, K. P., Experimental global vegetation index from AVHRR utilizing pre-launch calibration, cloud and sun-angle screening, Digital Data, National Geophysical Data Center, National Oceanic and Atmospheric Administration, Boulder Colorado, 1992.

GLOBALVIEW-CO_2, Cooperative Atmospheric Data Integration Project - Carbon Dioxide, CD-ROM, NOAA CMDL, Boulder, Colorado, 2000, [Also available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW].

Gloor, M., S. M. Fan, S. Pacala, and J. Sarmiento, Optimal sampling of the atmosphere for purpose of inverse modeling: A model study, *Global Biogeochem. Cycles*, *14*, 407–428, 2000.

Griewank, A., on automatic differentiation, in *Mathematical programming: recent developments and applications*, edited by M. Iri and K. Tanabe, pp. 83–108, Kluwer Academic Publishers, 1989.

Heimann, M., The global atmospheric tracer model tm2, *Technical report no. 10, deutsches klimarechenzentrum, hamburg, germany (issn 0940-9327)*, 1995.

Holland, E. A., J. C. Neff, A. R. Townsend, and B. McKeown, Uncertainties in the temperature sensitivity of decomposition in tropical and subtropical ecosystems: Implications for models, *Global Biogeochem. Cycles*, *14*, 1137-1151, 2000.

Houghton, J. T., G. J. Jenkins, and J. J. Ephraums (Eds.), *Climate change: The IPCC assessment*, Cambridge University Press, 1990.

Kaminski, T., M. Heimann, and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport, 1, Adjoint model and Jacobian matrix, *J. Geophys. Res.*, *104*, 18,535–18,553, 1999a.

Kaminski, T., M. Heimann, and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport, 2, Inversion of the transport of CO_2 in the 1980s, *J. Geophys. Res.*, *104*, 18,555–18,581, 1999b.

Kaminski, T., P. J. Rayner, M. Heimann, and I. G. Enting, On aggregation errors in atmospheric transport inversions, *J. Geophys. Res.*, *106*, 4703–4732, 2001.

Keeling, C. D., S. C. Piper, and M. Heimann, A three-dimensional model of atmospheric CO_2 transport based on observed winds: 4. Mean annual gradients and interannual variations, in *Aspects of Climate Variability in the Pacific and the Western Americas, Geophysical Monograph 55*, edited by D. H. Peterson, pp. 305–363, AGU, Washington, (USA), 1989.

Keeling, C. D., T. P. Whorf, M. Wahlen, and J. van der Plicht, Interannual extremes in the rate of rise of atmospheric carbon dioxide since 1980, *Nature*, *375*, 666–670, 1995.

Kicklighter, D. W., A. Bondeau, A. L. Schloss, J. Kaduk, A. D. McGuire, and the participants of the Potsdam NPP Model Intercomparison, Comparing global models of terrestrial net primary productivity (npp): global pattern and differentiation by major biomes, *Glob. Change Biol.*, *5*, 16–24, 1999.

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

Knorr, W., and M. Heimann, Uncertainties in global terrestrial biosphere modeling, part I: A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, *Glob. Biogeochem. Cyc.*, *15*, 207–225, 2001a.

Knorr, W., and M. Heimann, Uncertainties in global terrestrial biosphere modeling, part II: Global constraints for a process-based vegetation model, *Glob. Biogeochem. Cyc.*, *15*, 227–246, 2001b.

Knorr, W., and J.-P. Schulz, Using satellite data assimilation to infer global soil moisture status and vegetation feedback to climate, in *Remote Sensing and Climate Modeling: Synergies and Limitations*, Advances in Global Change Research, Kluwer Academic Publishers, Dordrecht, the Netherlands, pp. 273-306, 2001.

Knorr, W., and V. Lakshmi, Assimilation of fAPAR and surface temperature into a land surface and vegetation model, in *Land Surface Hydrology, Meteorology, and Climate Observations and Modeling*, Water Science and Application, Vol. 3, American Geophysical Union, Washington, pp. 177-200, 2001.

Law, R., The selection of model-generated CO_2 data: a case study with seasonal biospheric sources, *Tellus*, *48B*, 474–486, 1996.

Law, R. M., et al., Variations in modelled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, *Global Biogeochem. Cycles*, *10*, 783–796, 1996.

Leemans, R., and W. Cramer, The IIASA climate database for mean monthly values of temperature, precipitation and cloudiness on a terrestrial grid, *Tech. Rep. 91-18*, Institute of Applied Systems Analysis, Laxenburg/Austria, 1991.

Linacre, E. T., Estimating the net-radiation flux, *Agr. Meteorol.*, 1986.

Lloyd, J., and G. D. Farquhar, ^{13}C discrimination during CO_2 assimilation by the terrestrial biosphere, *Oecologia*, *99*, 201–215, 1994.

Louis, J. F., A parameteric model of vertical eddy fluxes in the atmosphere, *Bound.-Layer Meteor.*, *17*, 187–202, 1979.

Maisongrande, P., G. Dedieu, A. Ruimy, L. Kergoat, B. Saugier, and B. B. P. Ciais, Diagnostic and prognostic modeling of the terrestrial biosphere with remotely sensed measurements, in *The global carbon cycle and the terrestrial biosphere*, edited by G. Dedieu and J. L. Probst, vol. 50, pp. 33–58, (unknown), 1997.

Nemry, B., L. Fran cois, J. C. Gérard, M. Heimann, and the participants of the Potsdam NPP Model Intercomparison, Comparing global models of terrestrial net primary productivity (npp): analysis of the seasonal atmospheric CO_2 signal, *Glob. Change Biol.*, *5*, 65–76, 1999.

Raich, J. W., and W. H. Schlesinger, The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate, *Tellus*, *44B*, 81–99, 1992.

Raich, J. W., and C. S. Potter, Global patterns of carbon dioxide emisions from soils, *Global Biogeochem. Cycles*, *9*, 23–36, 1995.

Randerson, J. T., et al., The ^{13}c discrimination of arctic and boreal biome net co_2 exchange inferred from remote atmospheric measurements and a biosphere-atmosphere model, *Global Biogeochem. Cycles*, 2001, (submitted).

Rayner, P., Atmospheric perspectives on the ocean carbon cycle, in *global biogeochemical cycles in the climate system*, edited by E. D. Schulze, S. P. Harrison, M. Heimann, E. A. Holland, J. Lloyd, I. C. Prentice, and D. Schimel, Academic Press, San Diego, 2000, (in press).

Rayner, P. J., and R. M. Law, A comparison of modelled responses to prescribed CO_2 sources, *Tech. paper no 36*, CSIRO Div. of Atmos. Res., 1995.

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

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

Takahashi, T., R. H. Wanninkhof, R. A. Feely, R. F. Weiss, D. W. Chipman, N. Bates, J. Olafsson, C. Sabine, and S. C. Sutherland, Net sea-air CO_2 flux over the global oceans: An improved estimate based on the sea-air pCO_2 difference, in *Extended abstracts of the 2nd International CO_2 in the Oceans Symposium, Tsukuba, Japan, January 18-22, 1999*, pp. 18–01, 1999.

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

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

Tiedtke, M., A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Mon. Weather Rev.*, *117*, 1779–1800, 1989.

Wang, Y. P., and R. Leuning, Parameter estimation in surface exchange models using non-linear inversion: How many parameters can we estimate and which measurements are most useful?, *Glob. Change Biol.*, 2001, (submitted).

Wanninkhof, R., Relationship between wind speed and gas exchange over the ocean, *J. Geophys. Res.*, *97*, 7373–7383, 1992.

T. Kaminski, FastOpt, Martinistr. 21, D-20251 Hamburg, Germany. (thomas@fastopt.de)

W. Knorr and M. Heimann, Max-Planck-Institut für Biogeochemie, Postfach 100164, D-07701 Jena, Germany. (wknorr@bgc-jena.mpg.de;mheimann@bgc-jena.mpg.de)

P. Rayner, CSIRO-DAR, PMB 1 Aspendale, Victoria 3195, Australia. (e-mail: peter.rayner@dar.csiro.au)

Received ______________________

draft July 2001

This manuscript was prepared with AGU's LaTeX macros v4, with the extension package 'AGU^{++}' by P. W. Daly, version 1.6 from 1999/02/24.

Table 1.
Table 2.

Table 3.

Figure 1.

Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.
Figure 7.
Figure 8.
Figure 9.
Figure 10.
Figure 11.
Figure 12.

Figure Captions

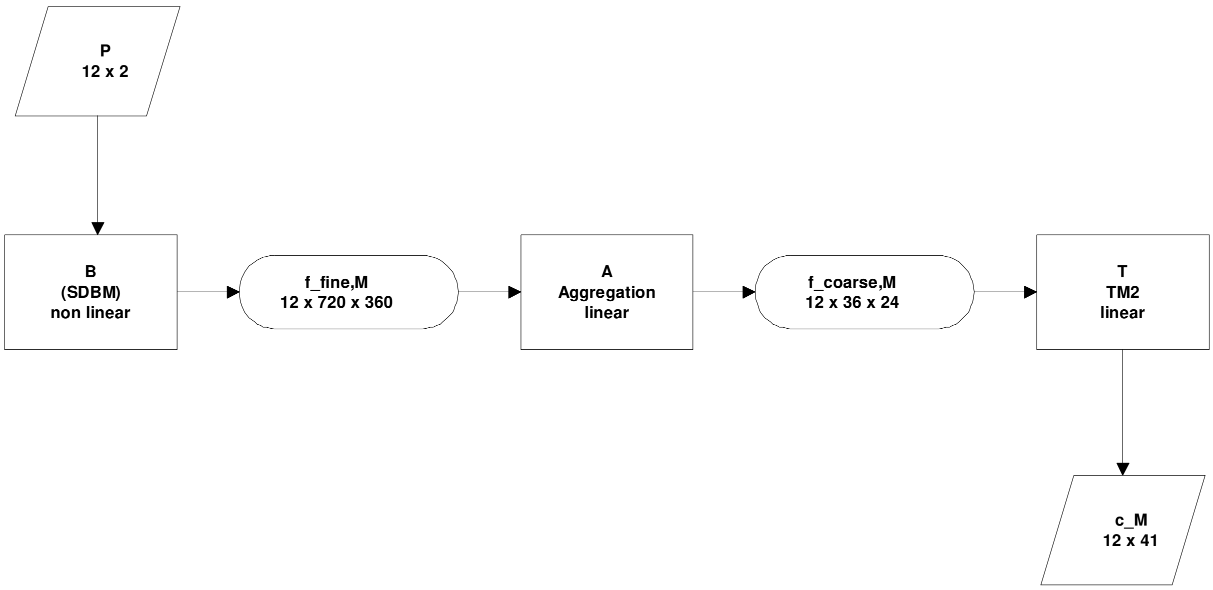

Figure 1. Flow diagram for the forward direction of information in our composite model.

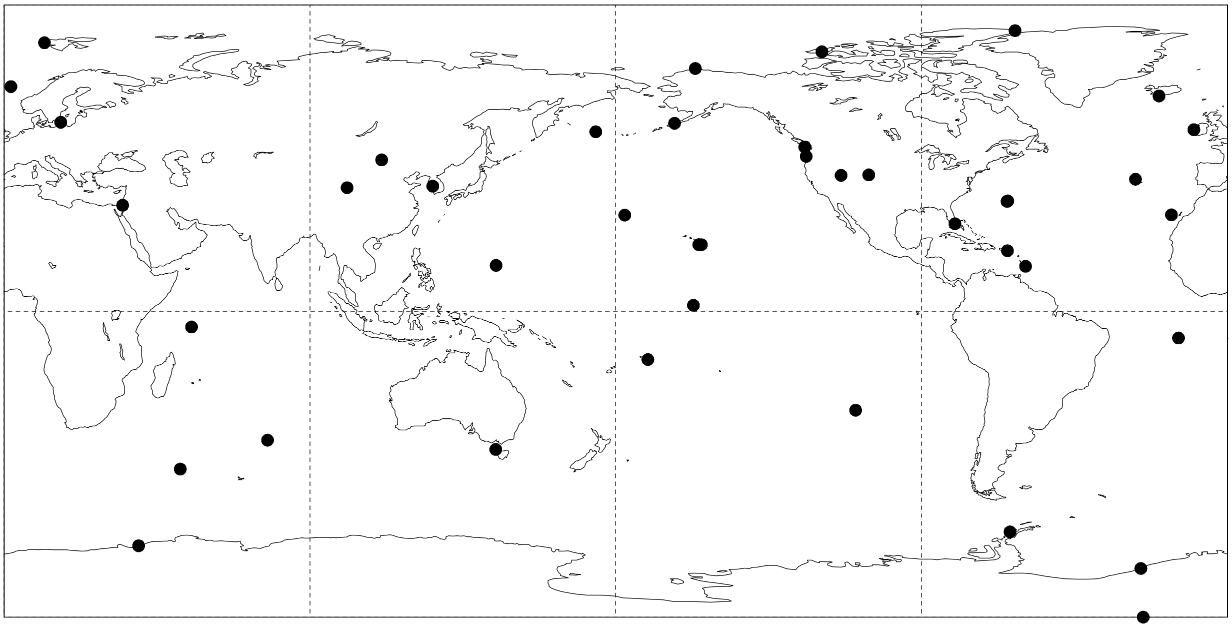

Figure 2. CO_2 observing network used in the study.

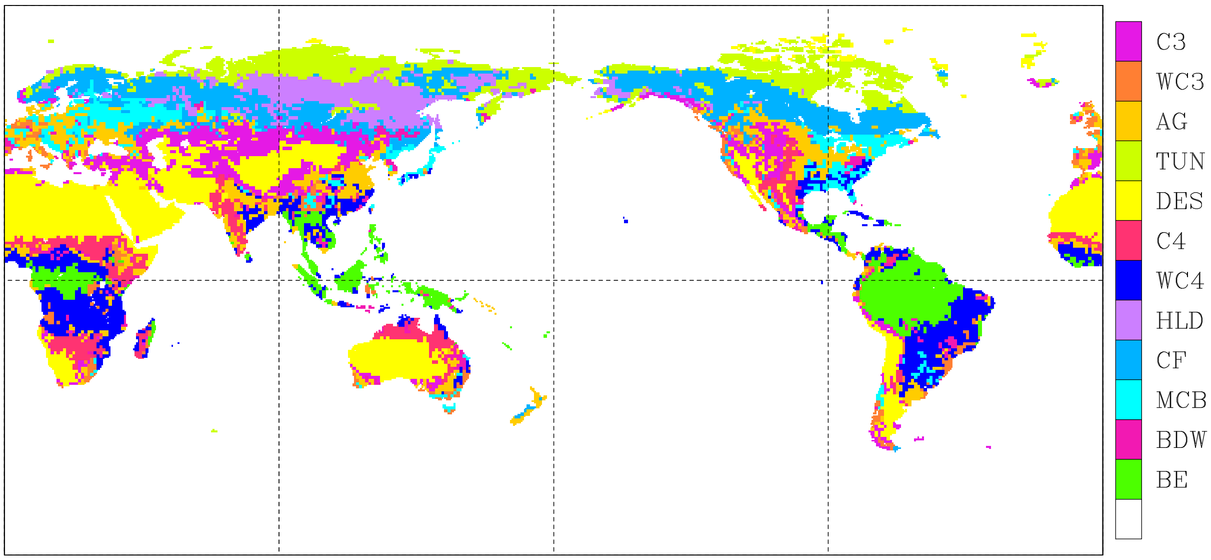

Figure 3. Distribution of biomes used in the study, taken from *DeFries and Townshend* [1994]. Biome labels are described in Table 1

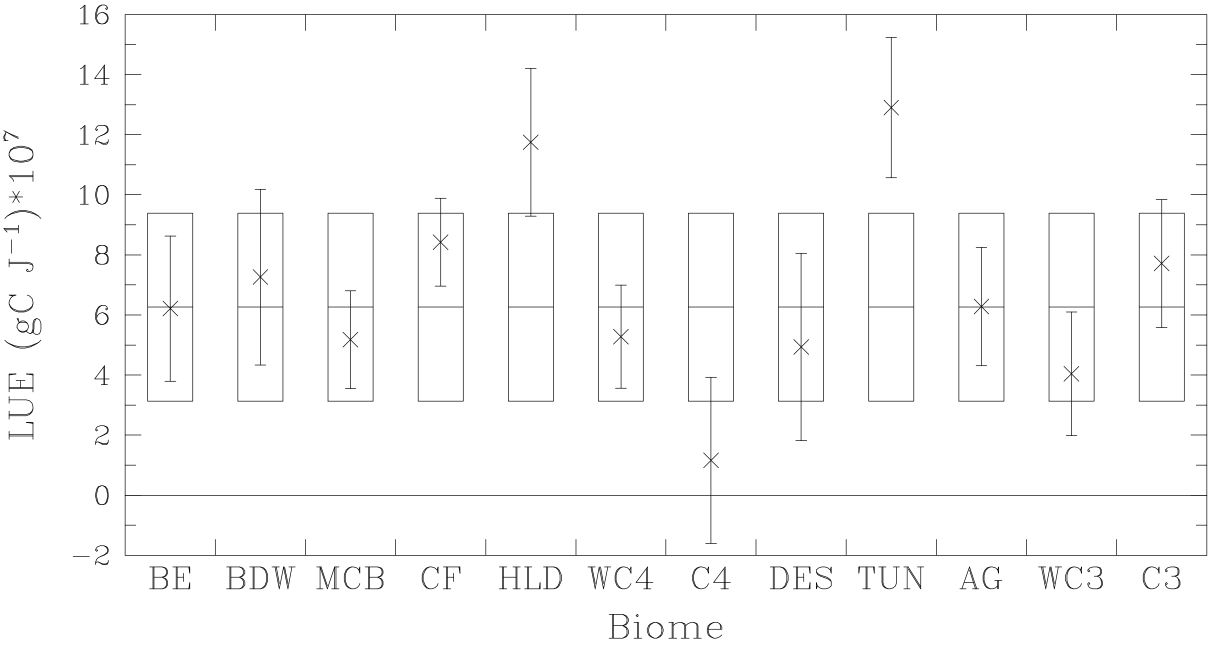

Figure 4. Light Use Efficiency ϵ for 12 biomes. The horizontal line shows the prior estimate, the box the 1 standard deviation of prior uncertainty, the cross the optimized value and the error bar the one standard eviation of predicted uncertainty. Biome labels are described in Table 1.

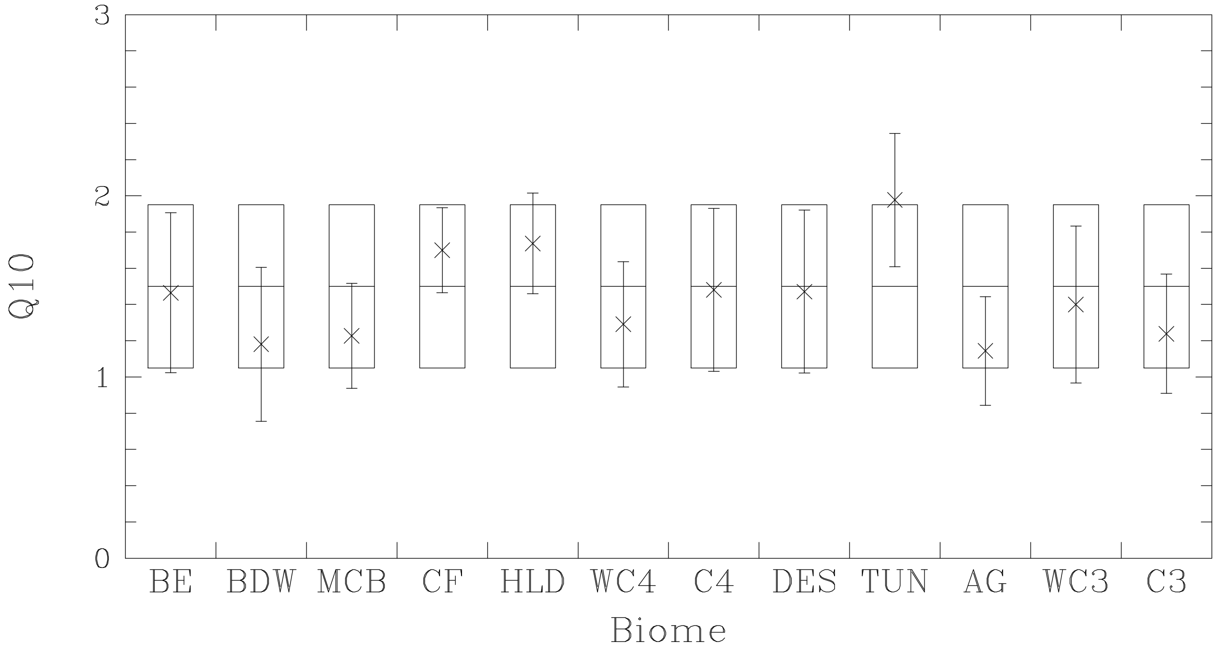

Figure 5. Q_{10} for twelve biomes. The horizontal line shows the prior estimate, the box the 1 standard deviation of prior uncertainty, the cross the optimized value and the error bar the one standard eviation of predicted uncertainty. Biome labels are described in Table 1.

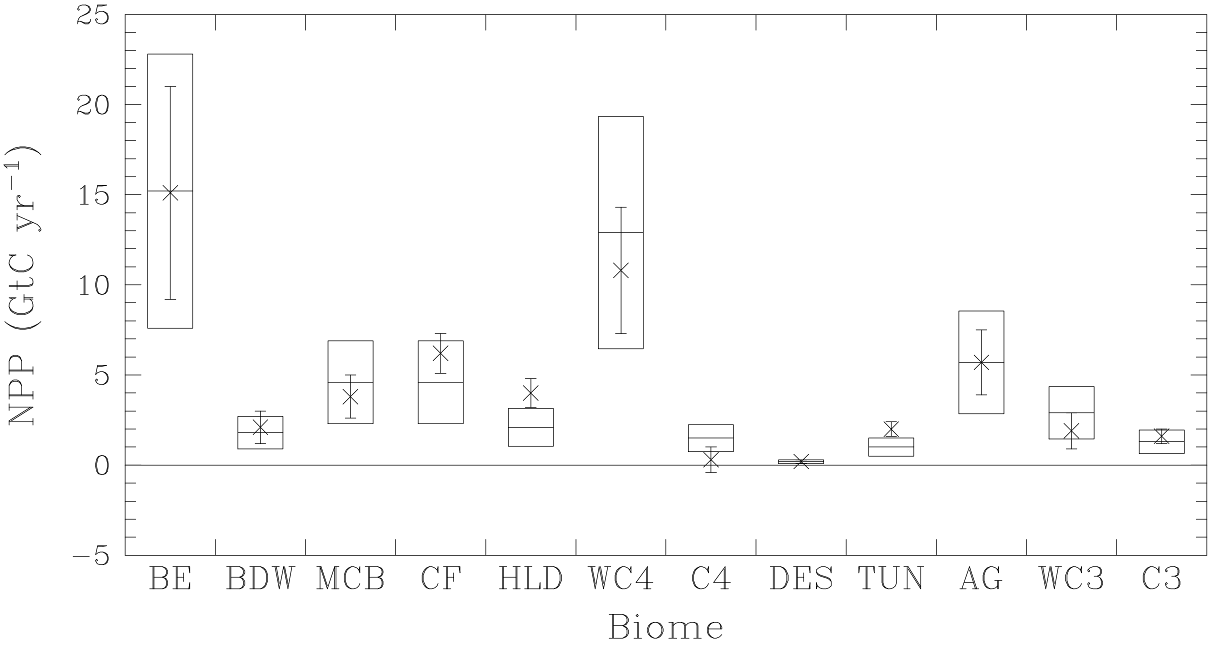

Figure 6. NPP for 12 biomes. The horizontal line shows the prior estimate, the box the 1 standard deviation of prior uncertainty, the cross the optimized value and the error bar the one standard eviation of predicted uncertainty. Biome labels are described in Table 1.

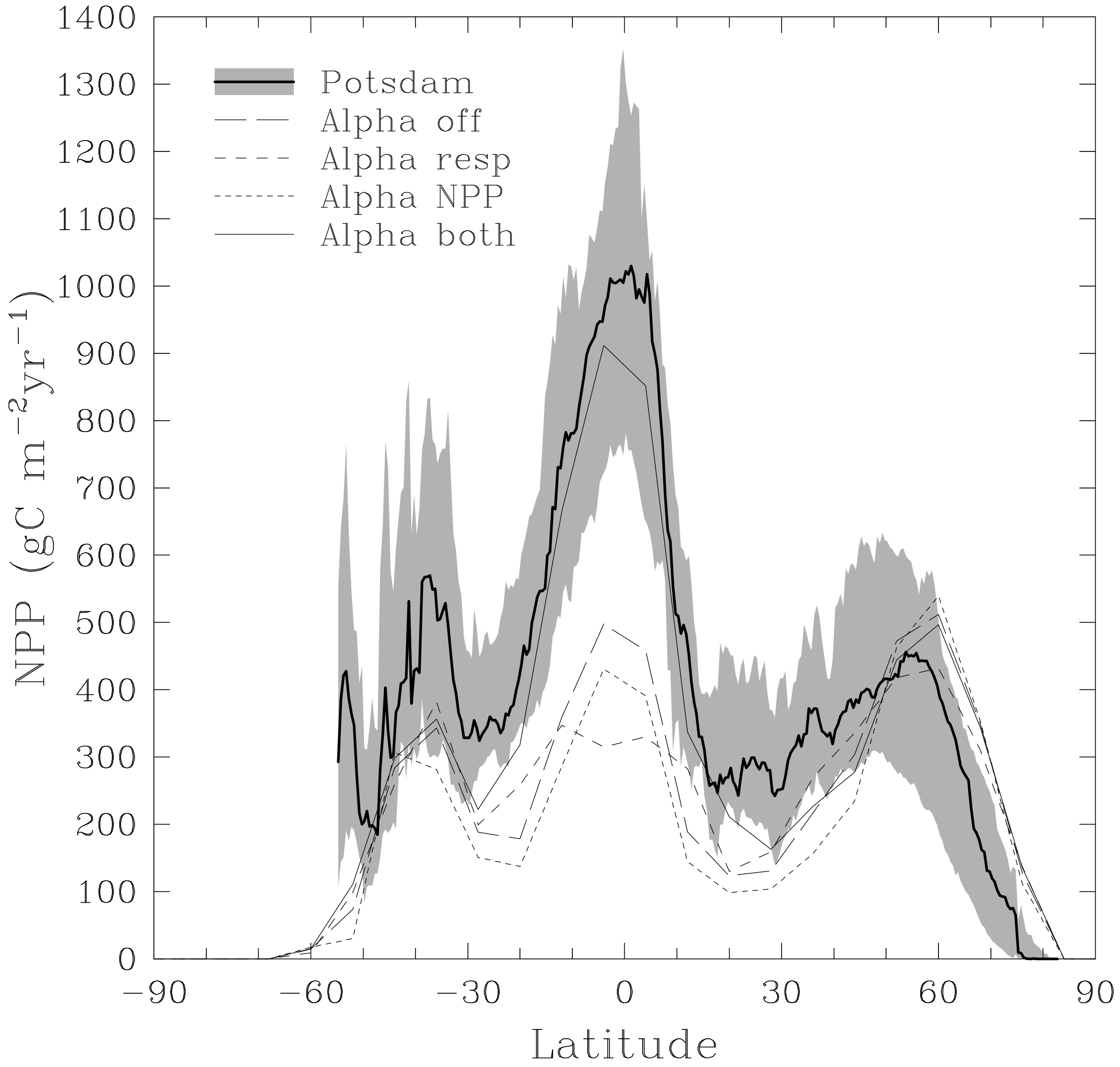

Figure 7. Zonal mean NPP per unit land area for four cases. The solid line shows our standard case while the three dashed lines show cases where the effect of water-stress is turned off for either or both of NPP and respiration. Line styles are as shown in the key. The shading indicates the interdecile range of model estimates from the Potsdam intercomparison.

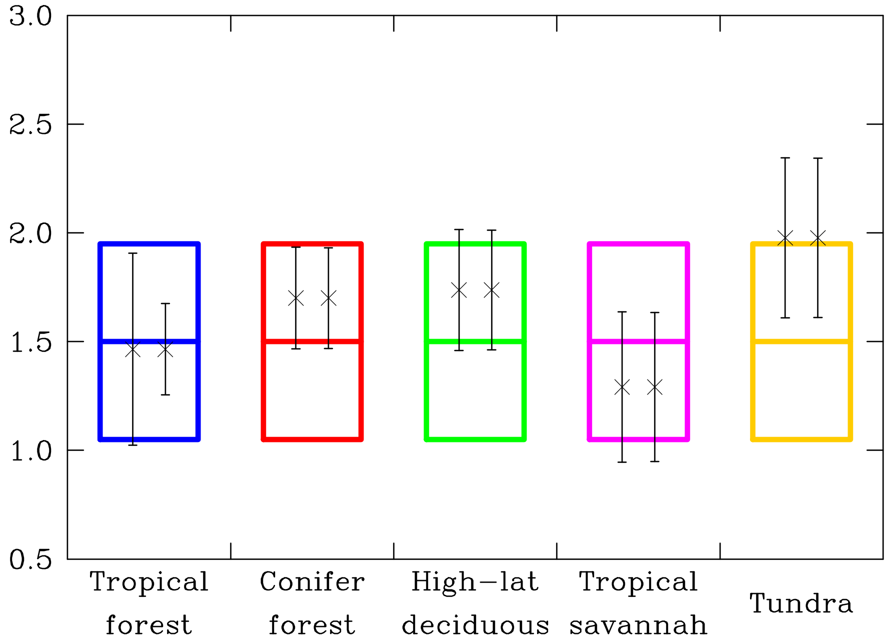

Figure 8. Predicted Q_{10} for five selected biomes. The horizontal line shows the prior estimate, the box the 1 standard deviation of prior uncertainty, and the cross the optimized value. The left vertical bar is as in Fig. 5 while the right error bar shows the one standard deviation uncertainty when pseudodata for flux measurements at one point in the Broadleaf Evergren biome is included. The flux pseudodata is that predicted from optimized values for that point and the flux measurement uncertainty is 10 $gm^{-2}y^{-1}$. Biome labels are described in Table 1.

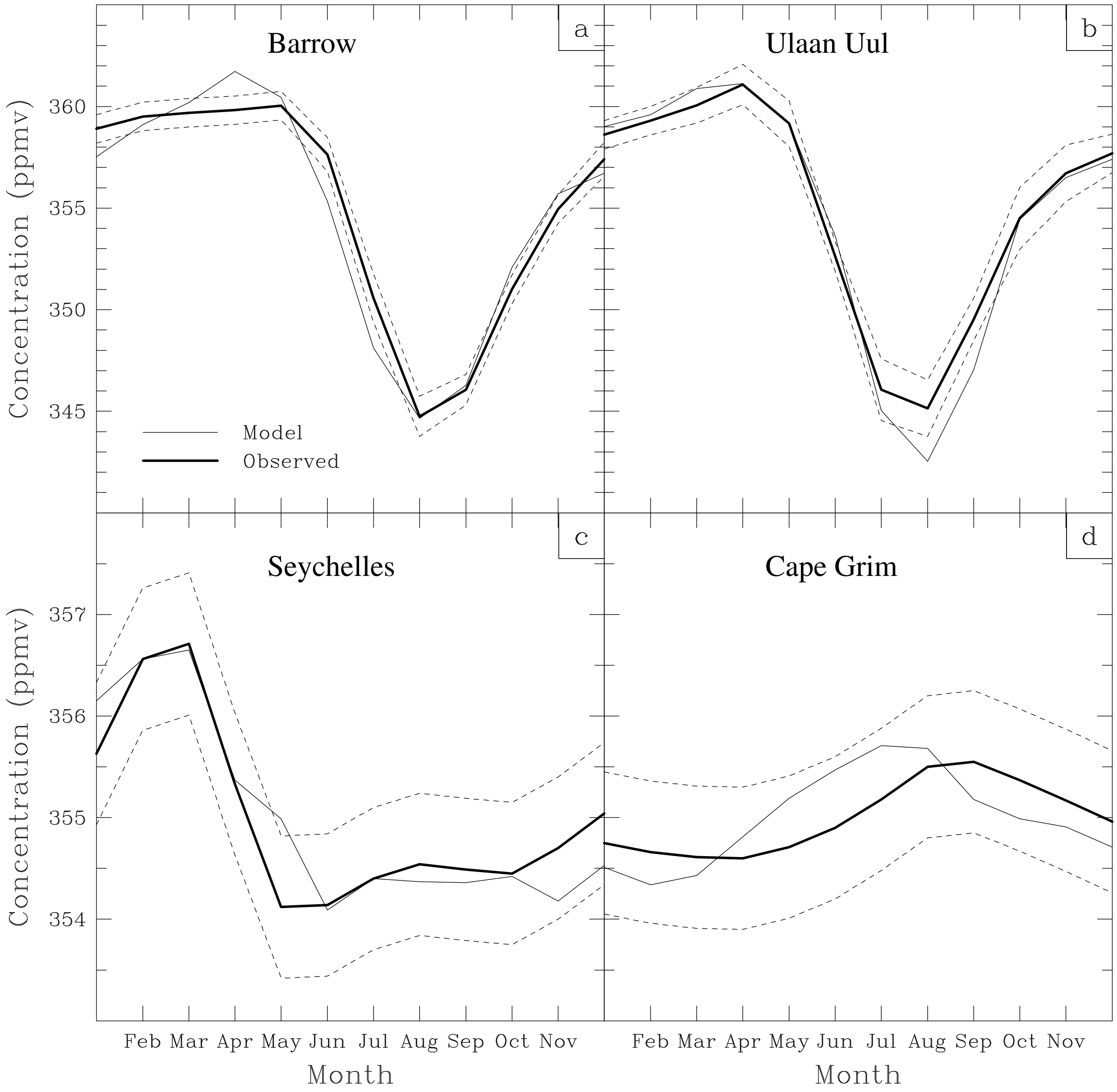

Figure 9. Simulated and observed seasonal cycle of concentration (ppmv)for four observing sites: (A) Point Barrow Alaska, (B) Ulaan Uul, Mongolia, (C) Mahe Island, Seychelles and (D) Cape Grim, Tasmania. The dashed line indicates the ± 1 standard deviation confidence interval about the observations.

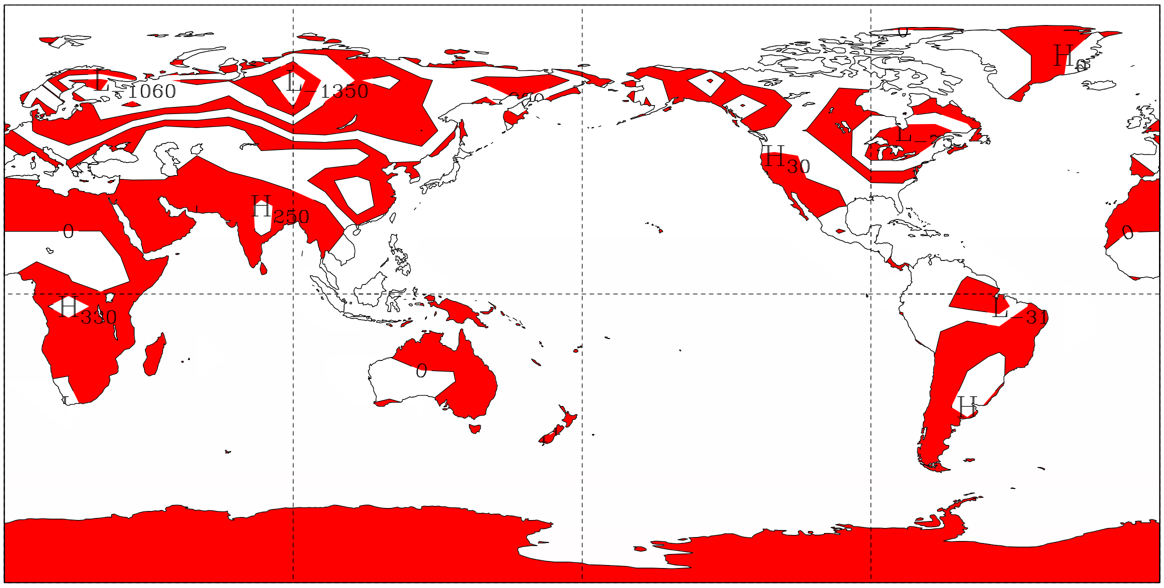

Figure 10. Predicted monthly mean net flux ($gm^{-2}y^{-1}$) from land for July in the control case. The contour interval is 200 and positive values are shaded.

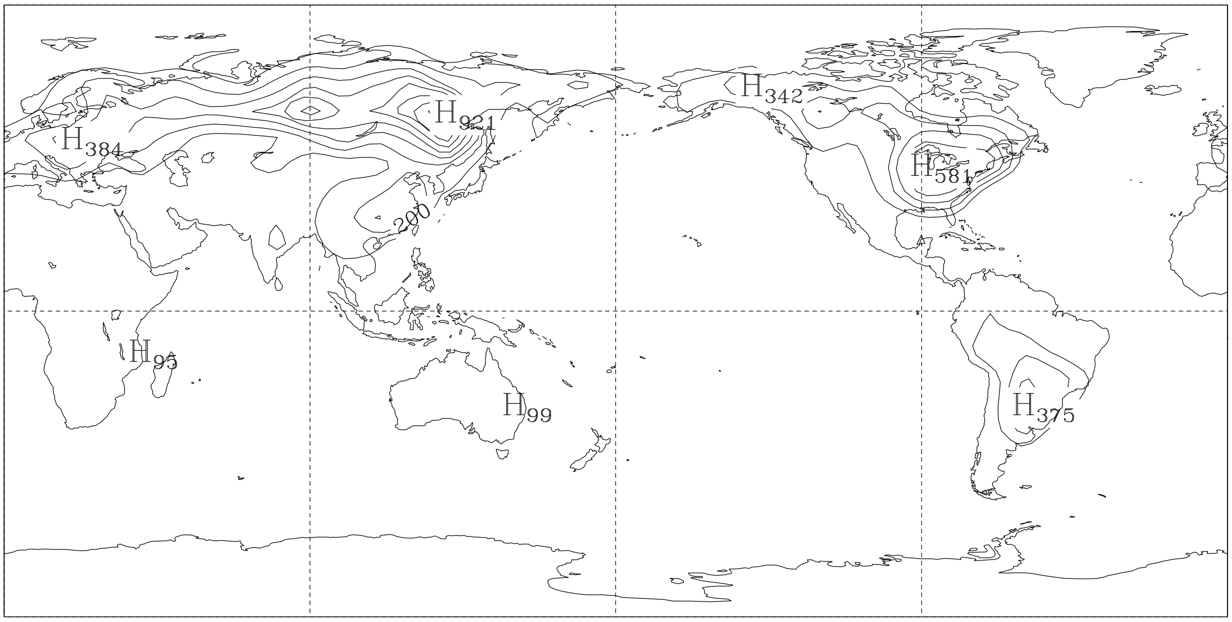

Figure 11. Predicted standard deviation for net flux ($gm^{-2}y^{-1}$) from land for July. in the control case. The contour interval is 100.

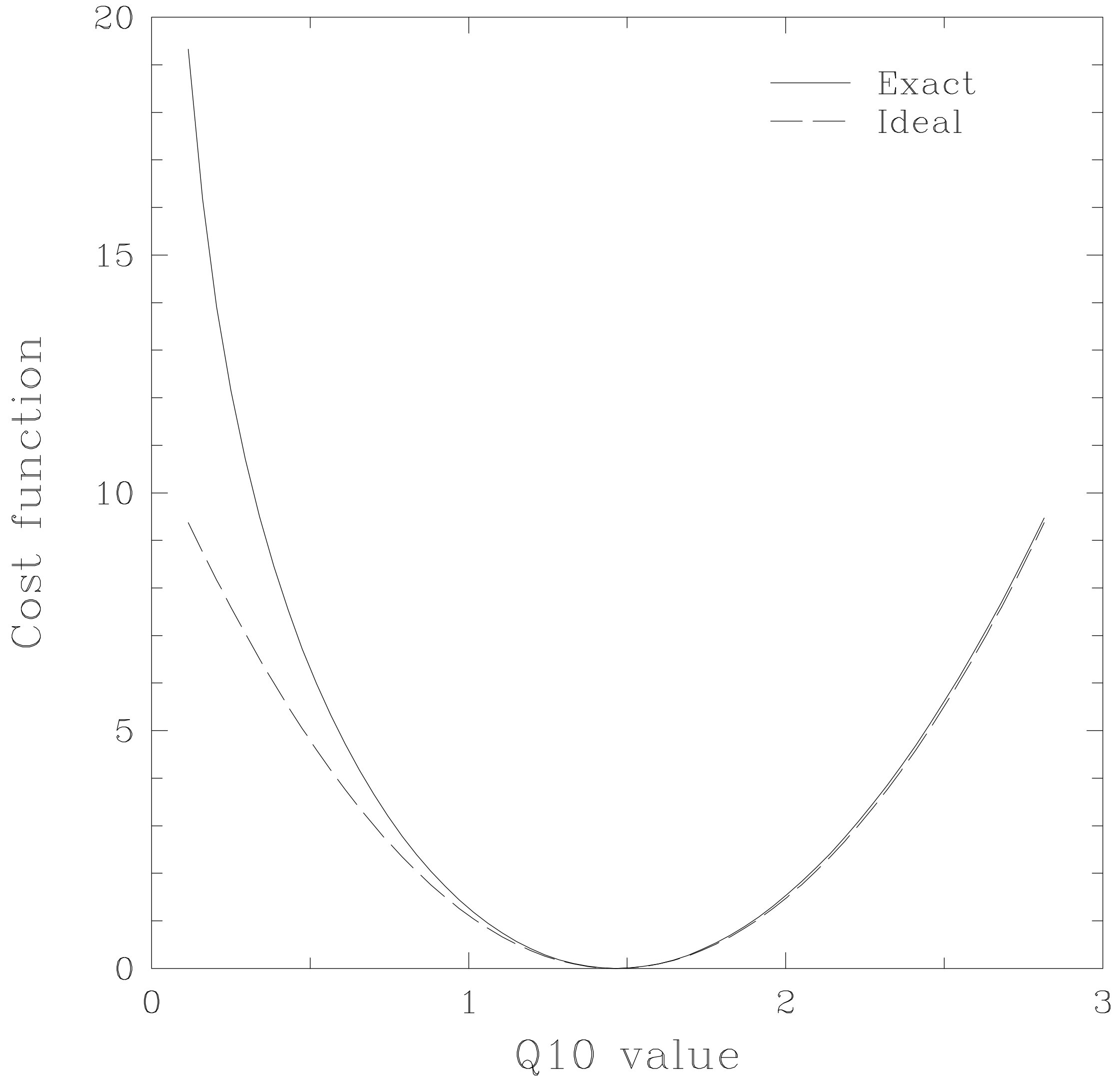

Figure 12. Exact and Ideal marginal cost function for Q_{10} for the Broadleaf Evergreen biome.

Tables

Abbreviation	Full Name
BE	broadleaf evergreen
BDW	broadleaf deciduous and woodland
MCB	mixed coniferous, broadleaf deciduous and woodland
CF	coniferous forest and woodland
HLD	high lat deciduous and woodland
WC4	wooded C_4 grassland
C4	C_4 grassland
DES	shrubs and bare ground
TUN	tundra
AG	cultivation
WC3	C_3 wooded grassland
C3	C_3 grassland

Table 1. Descriptive Abbreviations used for Biomes

Code	Descriptive Name	RMS uncertainty (ppmv)	RMS error (ppmv)
ALT	Alert, N.W.T., Canada	0.73	1.25
AMS	Amsterdam Island	0.70	0.58
ASC	Ascension Island	0.70	0.54
AVI	St. Croix, Virgin Islands	0.71	1.10
AZR	Terceira Island, Azores	0.96	0.62
BAL	Baltic Sea	2.79	0.73
BME	St. David's Head, Bermuda	0.83	1.19
BMW	Southhampton, Bermuda	0.96	1.22
BRW	Barrow, Alaska	0.81	1.59
CBA	Cold Bay, Alaska	0.83	1.26
CGO	Cape Grim, Tasmania	0.70	0.48
CHR	Christmas Island, Kiribati	0.70	0.50
CMO	Cape Meares, Oregon	1.22	0.72
CRZ	Crozet, Indian Ocean	0.70	0.44
EIC	Easter Island	0.70	0.68
GMI	Guam, Mariana Islands	0.74	0.61
HBA	Halley Bay, Antarctica	0.70	0.27
ICE	Storhofdi, Heimaey, Vestmannaeyjar, Iceland	0.77	1.02
IZO	Tenerife, Canary Islands	0.70	0.65
KEY	Key Biscayne, Florida	0.81	0.95
KUM	Cape Kumukahi, Hawaii	0.70	0.84
MBC	Mould Bay, N.W.T., Canada	0.77	1.42
MHD	Mace Head, Ireland	0.84	1.22
MID	Sand Island, Midway	0.73	0.71
MLO	Mauna Loa, Hawaii	0.70	0.39
NWR	Niwot Ridge, Colorado	0.81	1.10
OPW	Olympic Peninsula, Washington	1.24	0.78
PSA	Palmer Station, Antarctica	0.70	0.18
RPB	Ragged Point, Barbados	0.70	1.24
SEY	Mahe Island, Seychelles	0.70	0.53
SHM	Shemya Island, Alaska	0.89	1.46
SMO	Tutuila, American Samoa	0.70	0.32
SPO	South Pole, Antarctica	0.70	0.20
STM	Atlantic Ocean (Polarfront)	0.75	1.50
SYO	Syowa Station, Antarctica	0.70	0.32
TAP	Tae-ahn Peninsula, Korea	2.25	0.31
UTA	Wendover, Utah	1.19	0.80
UUM	Ulaan Uul, Mongolia	1.12	1.05
WIS	Sede Boker, Israel	1.03	1.90
WLG	Mt. Waliguan, China	0.86	1.37
ZEP	Zeppelin Station, Svalbard	0.86	0.98

Table 2. List of station codes and names along with their RMS uncertainties and RMS errors for the control optimization.

Abbrev	NPP, GtC					
	prior	off	resp	npp-prior	npp	both
BE	16.0	8.0	3.3	15.2	6.8	14.6
BDW	2.1	2.1	2.4	1.8	1.9	2.1
MCB	5.0	6.2	3.7	4.6	5.6	3.7
CF	5.0	6.2	3.7	4.6	7.0	5.7
HLD	2.2	4.0	4.1	2.1	4.3	4.1
WC4	14.4	4.8	7.5	12.9	3.5	10.9
C4	2.0	-0.3	0.9	1.5	-0.5	0.3
DES	0.4	0.3	0.3	0.2	0.2	0.2
TUN	1.0	1.9	2.1	1.0	1.6	2.0
AG	6.7	4.8	8.6	5.7	3.9	6.1
WC3	3.2	2.2	0.3	2.9	3.0	2.2
C3	1.9	0.9	2.7	1.3	-1.0	1.6
global	60.0	41.1	39.5	53.8	36.4	53.5
mismatch	n/a	0.87	0.78	n/a	0.98	0.87

Table 3. Predicted NPP for various Model Configurations. Headings refer to whether water-stress α is included for each flux component. The NPP-prior case lists the unoptimized NPP when water-stress is included for calculating NPP.