We use a carbon-cycle data assimilation system to estimate the terrestrial biospheric CO₂ flux until 2090. The terrestrial sink increases rapidly and the increase is stronger in the presence of climate change. Using a linearized model, we calculate the uncertainty in the flux owing to uncertainty in model parameters. The uncertainty is large and is dominated by the impact of soil moisture on heterotrophic respiration. We show that this uncertainty can be greatly reduced by constraining the model parameters with two decades of atmospheric measurements.

**Keywords:** carbon cycle; terrestrial uptake; data assimilation

**1. Introduction**

Uncertainty of the evolution of carbon uptake by the terrestrial biosphere is one of the largest uncertainties in climate forecasts [1]. Friedlingstein et al. [2], using several climate models with interactive carbon cycles, showed that much of this uncertainty comes from different interactions of the terrestrial biosphere and climate change. Although Friedlingstein et al. [2] displayed this as a large uncertainty in future CO₂ concentrations and temperatures, it is equally an uncertainty in the pathway required for stabilization of concentration.


*Author for correspondence (prayner@unimelb.edu.au).

One contribution of 17 to a Discussion Meeting Issue ‘Greenhouse gases in the Earth system: setting the agenda to 2030’.
Coupled ocean–atmosphere models with interactive carbon cycles suggest a role for climate change itself in reducing sink efficiency. The model of Cox et al. [9] showed considerable sensitivity of the terrestrial carbon uptake to climate change, with two-thirds of this owing to changes in soil carbon and one-third to the reduction in tropical forest, particularly the Amazon [10]. Jones et al. [11] showed how this uncertainty translated into an uncertainty of stabilization pathways for atmospheric concentration. The model of Dufresne et al. [12] showed less sensitivity, a difference only partly attributable to the lack of dynamic vegetation in that model. The comparison of Friedlingstein et al. [13] analysed the various contributions to this sensitivity and the larger comparison of Friedlingstein et al. [2] produced an ensemble estimate of this sensitivity, informing the statement in Denman et al. [1] that it constituted one of the largest uncertainties in climate prediction. The uncertainty range has only expanded as more recent studies [14,15] showed the different responses engendered when nitrogen feedbacks were included in the models. Finally, Gregory et al. [16] have recently pointed out that the sensitivity of the terrestrial carbon cycle to climate should also include the sensitivity of the carbon cycle to CO₂ concentrations even in the absence of climate change.

Given its importance there have been relatively few studies attempting to reduce this apparent uncertainty. Jones et al. [11] discussed the observational constraints available and Cadule et al. [17] introduced a benchmarking exercise for the models involved in such comparisons. There are now significant international projects for such benchmarking but the task of turning such benchmarks into probabilistic forecasts (e.g. [18]) is less advanced. The main aim of this paper is to prototype such an approach.

The outline of the paper is as follows. In §2, we describe the extensions necessary to a carbon-cycle data assimilation system (CCDAS) to use it for prediction. Section 3 describes the carbon-cycle predictions and their uncertainties along with the contribution of different model parameters to them. Finally, §4 points out caveats in the analysis and suggests some directions for future work.

2. Method

To assess the impact of current data on the uncertainty of future terrestrial uptake we need a carbon-cycle model capable of prediction and simultaneously of a data assimilation mode. Our starting point is the CCDAS described in Rayner et al. [19] and Scholze et al. [20], specifically in the version used in Koffi et al. [21]. In brief this used atmospheric concentration data from 68 stations over the period 1979–1999 to constrain the parameters of the terrestrial carbon-cycle model BETHY described by Knorr & Schulz [22]. BETHY is then forced with data from future climate simulations or with repetitions of the climate data used in the assimilation step. Various global outputs from these terrestrial simulations are then used as inputs to a global box model as described below. At each step, the model is linearized about its trajectory in order to calculate the sensitivities of the outputs with respect to the input parameters (the Jacobian of the model). The calculated sensitivities allow a linear error propagation from the input parameters to the outputs of the global model, in particular the future net
ecosystem productivity (NEP). This is possible since if \( v \) is a random vector with multivariate normal distribution described by the covariance matrix \( C(v) \) and \( M \) is any linear mapping then \( Mv \) is also multivariate normal with covariance
\[
Mv = MC(v)M^T,
\]
where the superscript \( T \) represents transpose. The assimilation step also calculates an approximate uncertainty for parameters. Comparison of the uncertainty in future NEP generated from prior and posterior parameter uncertainties is the primary result of our calculation.

(a) Simple model

As described in Rayner et al. [19] we need to make a simplification in the BETHY model for use in assimilation. For each gridpoint, we assume that the long-term mean NEP (defined here as the net flux from the undisturbed biosphere) is proportional to the long-term mean net primary productivity (NPP; the rate of plant growth) with the proportionality controlled by a regionally dependent parameter \( \beta \) [19, eqn (22)]. This parameter is a product of the assimilation under the assumption that the pool of carbon in the soil does not change. This assumption is reasonable for the two decades studied by Rayner et al. [19] but obviously cannot account for longer term changes in the real world. Since such changes are critical to the evolution of the sink in the long term, we must relax this constraint. We do this by building a diagnostic model based on outputs from the full model. This simple model (hereafter DIAGBETHY) takes as input the decomposition of litter, the NPP and the residence time of carbon in soil, all these are calculated by BETHY. DIAGBETHY is essentially a box model for soil carbon with the following governing equations:
\[
\frac{\partial C_S}{\partial t} = f_S D_L - \frac{C_S \omega^k Q_{10}^{T_a/10}}{\tau_B},
\]
where \( C_S \) is the mass of carbon in the soil, \( t \) is time, \( D_L \) is the rate of litter decomposition and \( f_S \) the fraction of \( D_L \) allocated to the soil pool. The term \( f_S D_L \) represents the input to the soil carbon pool from litter decomposition while the term \( C_S \omega^k Q_{10}^{T_a/10}/\tau_B \) represents the loss of soil carbon to the atmosphere by soil respiration. The parameter \( Q_{10} \) defines the exponential sensitivity of respiration to air temperature \( (T_a) \), while \( k \) describes the power-law relationship between respiration and fractional soil moisture \( \omega \). \( \tau_B \) is a normalization for the rate of soil decomposition. The bold-faced parameters \( f_S \) and \( D_L \) along with the term \( \omega^k Q_{10}^{T_a/10} \) are outputs from BETHY.

The NEP for DIAGBETHY is given by
\[
\text{NEP} = \text{NPP} - (1 - f_S) D_L - \frac{C_S \omega^k Q_{10}^{T_a/10}}{\tau_B}.
\]
Since equation (2.2) is a first-order differential equation its initial condition \( C_S(t = 0) \) is an input to the model, which is not available from BETHY. We must, therefore, treat it as an unconstrained parameter and its contribution to the uncertainty as an unavoidable nuisance. The term \( \tau_B \) determines the mean residence time and is set by the constraint that the NEP over the period of
the assimilation in DIAGBETHY must equal that in BETHY. That is, $\tau_B$ plays the role of the $\beta$ parameters from BETHY. $\tau_B$ is calculated as a preliminary calibration step in DIAGBETHY.

(b) Sensitivity analysis

The combination of the two models BETHY and DIAGBETHY constitutes a composition of two functions. Thus, the derivative of the combined model can be accomplished by application of the chain-rule. DIAGBETHY requires three outputs from BETHY: the NPP, $D_L$ and the term $\omega^* Q_{10}^{T_{an}/10}$ each at monthly resolution over the time of the simulation (here 1979–2090). We use the tool Transformation of Algorithms in Fortran [23] (http://www.fastopt.com) to calculate the derivative of each of these vector outputs with respect to the 57 parameters of BETHY which are exposed to the assimilation process.

DIAGBETHY is sufficiently simple that we can generate the equivalent derivative code by hand. Here, the output NEP is differentiated with respect to the inputs passed from BETHY plus three extra inputs. The parameter $f_s$, although itself a parameter to BETHY, also appears explicitly in DIAGBETHY. The initial size of the soil carbon pool $C_S$ must be treated as an unknown so we must take its potential contribution into account. Finally, the observed NEP over the BETHY assimilation period acts as a constraining pseudo-observation on DIAGBETHY and so any uncertainty in it could affect the output.

(c) Initial values and uncertainties

Since this is a study of the propagation of uncertainty, the prior probability distributions chosen for parameters are as important as their initial values. For most of the parameters of BETHY only positive values are physically meaningful and we thus choose lognormal distributions. The detailed specification is listed in Koffi et al. [21] but we note here that the choice represents an optimistic view of the prior uncertainty on the biosphere. For example, it generates a prior uncertainty in the gross primary productivity (GPP) of only 3 Pg C yr$^{-1}$ over the assimilation period. The formulation of BETHY means that all parameters are dimensionless with prior uncertainties unity and the difference in the uncertainty of physical parameters is transmitted by different sensitivities within the model. Most of this is irrelevant to DIAGBETHY except the parameter $f_s$, which appears explicitly. We list its apparent uncertainty along with the other inputs to DIAGBETHY in table 1.
(d) Experiments

Following the useful distinction of Gregory et al. [16], we wish to investigate the uncertainty in the uptake with and without climate change, and before and after confronting the model with the atmospheric data. The experiments are complicated by the fact that BETHY is solely a model of the carbon balance. As noted in Rayer et al. [19], it is abstracted from a larger model FULLBETHY, which includes modelling of hydrology, phenology and energy exchange. Output of soil moisture and leaf area index from FULLBETHY are used as inputs to BETHY. This poses two problems. Firstly, our sensitivity analysis is necessarily incomplete since it does not include the parameters in FULLBETHY that do not appear in BETHY. Secondly, the simulations of FULLBETHY are performed with prior parameters. Thus, there is likely to be a slight inconsistency between, say, the carbon available for the formation of leaves in FULLBETHY and in BETHY once the model parameters have been optimized. These inconsistencies are probably not important for the uncertainties which are the focus of this paper.

We need a chain of models in the calculation. First, we extract climate output from the Institut Pierre-Simon Laplace climate model, here the version used in Solomon et al. [24]. We must choose a particular concentration scenario and, for consistency with Friedlingstein et al. [2] we choose the SRES-A2 scenario. For the climate-change scenario, we need consistency with the observed forcing during the assimilation period but should also avoid a sharp discontinuity between the observed and predicted periods. We achieve this by adjusting the output of the climate model by the mean difference between the model and observed forcing over the assimilation period so that for any quantity $x$

$$x(t) = \begin{cases} 
  x_O(t) & (t < 2000) \\
  x_M(t) - \left[ x_M(t^*) - x_O(t^*) \right] & (t \geq 2000)
\end{cases} \tag{2.4},$$

where $t^*$ represents the month of year in a seasonal climatology. The required fields to run FULLBETHY are the maximum and minimum air temperature, precipitation, downward shortwave radiation, surface relative humidity and surface wind speed. For the no-climate-change scenario, we repeat the observed forcing six times to cover the period 1979–2100. For all scenarios, we increase atmospheric CO2 concentration in line with the SRES-A2 scenario.

With these fields we run FULLBETHY that calculates fields of leaf area index and soil moisture which, along with the fields from the climate model, are used as inputs to BETHY whose outputs are, in turn, used as inputs to DIAGBETHY.

3. Results

(a) Fit to data

The overall performance of the CCDAS was discussed in detail in Rayner et al. [19] and Scholze et al. [20] and the performance of this version was described in Koffi et al. [21]. Central to the long-term response to climate change are the slow processes surrounding heterotrophic respiration. Within the 20 year assimilation window these processes will manifest themselves in the interannual variability of fluxes. Thus, a useful indicator of model performance is the interannual
variability in the CO2 growth rate simulated by the prior and optimized models. Following Rayner et al. [19], we use the linear combination of three-quarters the concentration at Mauna Loa plus one-quarter the concentration at the South Pole and the procedures of Thoning et al. [25] to calculate interannual variability in the growth rate.

Figure 1 shows the variation in growth rate for the observations, the prior model and the optimized model. We see that the prior model produces a bad fit with many of the variations occurring out-of-phase with the observations, while the situation is greatly improved for the optimized model. In particular, we see a large error in 1982. Although this is the first year for which we have data at both stations, it is not a transient related to model spin-up but rather a response to the anomalous climate associated with the 1982 El Niño event. Of course, this is not an independent check of model quality since we are only testing the fit to data that were assimilated. The significant change in the fit does suggest, however, that important processes active on interannual time scales have been greatly changed by the optimization.

(b) Simulated net ecosystem productivity

Although the focus of this paper is the uncertainty in future NEP rather than the NEP itself, the two cannot be entirely separated. Figure 2 shows the decadal mean NEP for the period 2000–2090 with the mean for 2000–2010 subtracted. We denote this quantity δNEP. The figure shows the four cases, with and without climate change and with optimized and prior parameters.

All four cases show large increases in δNEP, particularly at the start of the period. The optimized cases (solid lines) produce greater δNEP than the unoptimized cases. The impact of climate change is, however, different. For the unoptimized cases (broken lines), climate change increases δNEP throughout the period while for the optimized cases (solid lines) climate change reduces δNEP at the end of the period.

The rapid increase in δNEP without climate change suggests large sensitivity to CO2. We verified this by running the model over the assimilation period with CO2 set at 355 ppm then at 356 ppm. The mean GPP was 0.3 PgC yr$^{-1}$ larger with
1 ppm extra CO$_2$. This is a large sensitivity. It suggests, for example, an increase of about 10 PgC yr$^{-1}$ over the assimilation period and would, for reasonable values of residence times, suggest an unrealistically large terrestrial uptake over the twentieth century. We therefore expect the $\delta$NEP values from BETHY to be an upper estimate.

(c) Uncertainty in net ecosystem productivity

Using linear error propagation through the tangent linear version of DIAGBETHY, we can calculate the uncertainty in $\delta$NEP arising from uncertainty in BETHY and DIAGBETHY parameters. Figure 3 shows this uncertainty as a standard deviation. The parameter uncertainties used are the prior uncertainties or the posterior covariance as generated by Koffi et al. [21].

As another measure we can calculate the uncertainty of the integrated $\delta$NEP between 2000 and 2090. The results are shown in table 2 along with the values for summed $\delta$NEP. Both figure 3 and table 2 show that uncertainty in $\delta$NEP is large. For the case most pertinent to a climate model (that with prior parameters and climate change), we see that the uncertainty in $\delta$NEP is larger than $\delta$NEP itself.

Considering the differences between the ‘climate change’ and ‘no climate change’ scenarios for prior parameters, we see that the uncertainty of the interaction of climate change with the biosphere dominates other contributions. Another way to express this is to calculate the uncertainty in the gain $G = \delta$NEP$_C/\delta$NEP$_N$, where the subscripts C and N refer to ‘climate change’ and ‘no climate change’ respectively. To do this, we first calculate the Jacobian $\nabla_p G$ (where $P$ is a vector of the parameters of DIAGBETHY) using the product rule:

$$\nabla_p G = \frac{\nabla_p \delta$NEP$_C}{\delta$NEP$_N} - \frac{\nabla_p \delta$NEP$_N \delta$NEP$_C}{\delta$NEP$_N^2},$$

\text{(3.1)}
Figure 3. Uncertainty in $\delta$NEP (expressed as 1σ standard deviation). Broken lines show the case for prior parameters and solid for optimized. Thick lines show the cases with climate change and thin lines with the observed climate for 1979–1999 repeated.

Table 2. Uncertainty in integrated $\delta$NEP from 2005 to 2085 for the four experiments.

<table>
<thead>
<tr>
<th>case</th>
<th>sum (PgC)</th>
<th>uncertainty (1σ PgC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>prior no-clim</td>
<td>278</td>
<td>126</td>
</tr>
<tr>
<td>prior clim-change</td>
<td>656</td>
<td>1141</td>
</tr>
<tr>
<td>optimized no-clim</td>
<td>717</td>
<td>78</td>
</tr>
<tr>
<td>optimized clim-change</td>
<td>799</td>
<td>107</td>
</tr>
</tbody>
</table>

then use conventional error propagation. We note that all the terms in equation (3.1) are available from the tangent linear model. Applying equation (3.1) to the prior model yields an uncertainty for $G$ of 3.5. This is unrealistically large but is an indicator of the immense uncertainty of the climate feedback even within the same model. We can go a little further to ask what parameter contributes the most to this uncertainty. This is possible because with the diagonal $C(P)$ for the prior parameters, equation (2.1) simplifies to

$$C(G) = \sum_i \nabla^2_{P_i} G C(P)_{i,i}. \quad (3.2)$$

This calculation shows that 80 per cent of the variance is contributed by the parameter $\kappa$ from equation (2.3) while $f_s$ contributes 14 per cent. This suggests that while Jones et al. [26] found the sensitivity of respiration to temperature (embodied in the parameter $Q_{10}$) dominated the climate response of the biosphere, here it is the sensitivity to soil moisture, or rather this sensitivity weighted by its uncertainty. Soil moisture and its response to climate change are highly variable suggesting this dominance may be even larger at smaller scales. Our results underline the importance of detecting soil moisture changes and their interaction with the atmosphere.

Phil. Trans. R. Soc. A (2011)
The analysis of equation (3.2) also enables us to allay an obvious concern with DIAGBETHY. The initial pool size is almost unconstrained. It makes, however, almost no contribution to the uncertainty of either $\delta$NEP or $G$. The main reason for this is the use of $\delta$NEP in which the NEP for a baseline period is subtracted. This removes most of the sensitivity to the initial pool size.

The case for optimized parameters is very different. Along with the rise in $\delta$NEP evident from figure 2, table 2 shows a dramatic reduction in the uncertainty of $\delta$NEP by a factor of 10. Even more remarkably the uncertainty in the climate feedback is reduced from 3.5 to 0.04. This is driven by an uncertainty reduction of 99 per cent for the parameter $\kappa$. Thus, it appears there is considerable scope for reducing the uncertainty of the carbon-cycle climate feedback by formally confronting carbon-cycle models with contemporary data.

4. Discussion

The very large uncertainty of $\delta$NEP for our control case (prior parameters and climate change) is likely overestimated for two reasons. Firstly, the large increase in NPP (more than 50% between the decades 2000–2009 and 2080–2089) means that any uncertainty in respiration will act on large carbon stocks and so be inflated. In this respect, $G$ is perhaps a more robust model property. Also, this is an off-line experiment where other negative feedbacks on CO2 concentration (such as ocean uptake or CO2 fertilization itself) are unable to act. We can roughly estimate this effect by using the current airborne fraction of 40 per cent suggesting our sensitivities should be roughly halved.

Even taking these into account, the uncertainty in integrated $\delta$NEP is important. Halving it to account for the missing feedbacks then converting to parts per million, we see an uncertainty around 270 ppm, large when compared with the range in Friedlingstein et al. [2]. It also suggests that our result of a negative feedback of climate change on the carbon cycle (i.e that climate change increases NEP and hence reduces radiative forcing) should not be given too much weight since it is highly uncertain.

One of the most striking results is the dominant role of soil moisture in determining the carbon-cycle/climate feedback, particularly compared with Jones et al. [26]. There are some structural reasons for this in BETHY and its derived models. Parameter $\kappa$ appears in the equations for the respiration of both the litter pool and soil pool while there is a different $Q_{10}$ value for each pool. This might reduce the dominance somewhat but even the summed impacts of the two $Q_{10}$ parameters on the uncertainty are negligible compared with that of $\kappa$. Analysis of the Jacobians which are multiplied to generate the term $\partial G/\partial \kappa$ shows that the main impact comes through its impact on the respiration of the litter pool. This is important since the dynamics of the litter pool are generally fast enough to be observed within the 21 year assimilation window we use.

The relation of the important time scales for centennial changes in NEP to the two decades of observations we use to constrain them could limit our approach; clearly if important processes are too slow to be observed in 20 years then our observations will not help much. This is not the case with the model as it stands since the observations do provide a strong constraint. However, much of the dynamics missing from BETHY is slow, e.g. succession. The large-scale mortality
seen in Cox et al. [9] and responsible for about one-third of the carbon loss would undoubtedly contribute its own uncertainty and it is unlikely the observational period would constrain it. Furthermore, climate sensitivity was an important contributor to the feedback shown by Friedlingstein et al. [13] so its uncertainty is likely to be comparably important.

The simplicity of our model leaves our analysis somewhere between a technical demonstration of a methodology and a physically significant result. The model is missing various important processes that could respond to climate change, such as fires, ecosystem composition and weathering. The results are interesting enough, and the methodology simple enough, to recommend a similar approach be tried on a spatially explicit model without the simplifications present in DIAGBETHY. The spatial richness available in such an approach will yield a great deal of information on model dynamics and suggest what and where to measure to reduce uncertainty. There also seems no reason to limit the approach to biogeochemical models. One could easily imagine a tangent linear version of a climate model giving a good deal of information on the uncertainties of the climate sensitivity.

5. Conclusions

We constructed a diagnostic box model of the terrestrial carbon cycle consistent with and driven by the BETHY model. We forced this chain of models with outputs from climate models to estimate future terrestrial carbon fluxes. Using a tangent linear version of the models, we calculated uncertainties in this future uptake. The conclusions can be summarized as follows.

— NEP increases rapidly into the future and climate change enhances this uptake. This is most probably owing to unrealistic sensitivity of photosynthesis to CO₂ concentration.
— Uncertainty in future uptake is large, especially the interaction with climate change. The dominant component in this uncertainty is the sensitivity of heterotrophic respiration to soil moisture.
— Assimilating 21 years of atmospheric observations to constrain model parameters can greatly reduce this uncertainty.

P.R. is the recipient of an Australian Research Council Professorial Fellowship (DP1096309).

References

Carbon-cycle predictions


