



# The climate dependence of the terrestrial carbon cycle, including parameter and structural uncertainties

M. J. Smith, D. W. Purves, M. C. Vanderwel, V. Lyutsarev, and S. Emmott

Computational Science Laboratory, Microsoft Research Cambridge, 21 Station Road, Cambridge, CB1 2FB, UK

*Correspondence to:* M. J. Smith (matthew.smith@Microsoft.com)

Received: 17 July 2012 – Published in Biogeosciences Discuss.: 4 October 2012

Revised: 18 December 2012 – Accepted: 2 January 2013 – Published: 29 January 2013

**Abstract.** The feedback between climate and the terrestrial carbon cycle will be a key determinant of the dynamics of the Earth System (the thin layer that contains and supports life) over the coming decades and centuries. However, Earth System Model projections of the terrestrial carbon-balance vary widely over these timescales. This is largely due to differences in their terrestrial carbon cycle models. A major goal in biogeosciences is therefore to improve understanding of the terrestrial carbon cycle to enable better constrained projections. Utilising empirical data to constrain and assess component processes in terrestrial carbon cycle models will be essential to achieving this goal. We used a new model construction method to data-constrain all parameters of all component processes within a global terrestrial carbon model, employing as data constraints a collection of 12 empirical data sets characterising global patterns of carbon stocks and flows. Our goals were to assess the climate dependencies inferred for all component processes, assess whether these were consistent with current knowledge and understanding, assess the importance of different data sets and the model structure for inferring those dependencies, assess the predictive accuracy of the model and ultimately to identify a methodology by which alternative component models could be compared within the same framework in the future. Although formulated as differential equations describing carbon fluxes through plant and soil pools, the model was fitted assuming the carbon pools were in states of dynamic equilibrium (input rates equal output rates). Thus, the parameterised model is of the equilibrium terrestrial carbon cycle. All but 2 of the 12 component processes to the model were inferred to have strong climate dependencies, although it was not possible to data-constrain all parameters, indicating some potentially redundant details. Similar climate dependencies

were obtained for most processes, whether inferred individually from their corresponding data sets or using the full terrestrial carbon model and all available data sets, indicating a strong overall consistency in the information provided by different data sets under the assumed model formulation. A notable exception was plant mortality, in which qualitatively different climate dependencies were inferred depending on the model formulation and data sets used, highlighting this component as the major structural uncertainty in the model. All but two component processes predicted empirical data better than a null model in which no climate dependency was assumed. Equilibrium plant carbon was predicted especially well (explaining around 70 % of the variation in the withheld evaluation data). We discuss the advantages of our approach in relation to advancing our understanding of the carbon cycle and enabling Earth System Models to make better constrained projections.

## 1 Introduction

Whilst models of the Earth System (the thin layer that contains and supports life) have evolved in response to improvements in our understanding of different processes (Randall et al., 2007), wide differences in the predictions of different models still greatly limit decision making about how best to adapt to climate change (Cox and Stephenson, 2007; Kerr, 2011; Maslin and Austin, 2012). Improvements in characterising the consistency of models with empirical data and with each other, in terms of predictive accuracy and in terms of in-built assumptions, would improve clarity about why model predictions differ and hopefully enable critical improvements to be made that improve confidence in predictions.

Models of the terrestrial carbon cycle are one of the Earth System Model components most critically in need of improvement. The terrestrial carbon cycle has major effects on the dynamics of the Earth System over decadal or longer timescales (Denman et al., 2007), and, whilst this means that terrestrial vegetation currently accounts for approximately 60 % of the total annual flux in atmospheric carbon dioxide and absorbs around a quarter of anthropogenic carbon dioxide emissions (Denman et al., 2007), there is great uncertainty about how this balance will change in the future (Cramer et al., 2001; Friedlingstein et al., 2006; Denman et al., 2007; Sitch et al., 2008). This uncertainty is largely because models exhibit wide differences in their predictive accuracy (Keenan et al., 2012) and lead to widely diverging and inconsistent projections (Friedlingstein et al., 2006).

Resolving the problem of the differences in carbon model predictions is a major research challenge. Traditionally, carbon models have not been developed in a way that enables detailed intercomparisons to assess why their predictions differ. Their component processes and their parameterisations have been based on contemporary understanding but have not explicitly (quantitatively) incorporated confidence in how that understanding is based on empirical data. Further research has typically added more details to these models but has rarely gone back and characterised the consistency of the initial assumptions, or the overall model, with empirical data. Recent work has shown that explicitly constraining parameters of terrestrial carbon models with empirical data can lead to better understanding of uncertainty in their parameterisations and of the importance of that uncertainty for predictions (Knorr and Heimann, 2001; Scholze et al., 2007; Zhou and Luo, 2008; Rayner et al., 2011; Ricciuto et al., 2011). Recent systematic comparisons of alternative carbon models or their components have also shown how differences and inconsistencies between different models can be identified more precisely (Keenan et al., 2012; van Oijen et al., 2011; Randerson et al., 2009; Kloster et al., 2010). These analyses have been facilitated by the increased availability of more varied and detailed data sets on terrestrial carbon stocks and fluxes from around the globe.

Delivering better constrained projections of terrestrial carbon cycle dynamics could soon be achieved in light of these recent advances. However, delivering that goal is going to require improved methodologies for the construction, parameterisation and evaluation of terrestrial carbon cycle models, which enable the detailed analyses of the consistency of different model components and their parameterisations with empirical data. Here we develop such a methodology and implement it to fully decompose all component processes of a global terrestrial carbon cycle model in terms of their parameter uncertainty and the accuracy of their predictions with respect to different empirical data sets. Specifically, our goals were to (i) assess the degree of empirical support for simple functional representations of component processes of the carbon cycle, when assessed within a model of how the overall

system is connected; (ii) to assess whether the inferred relationships are consistent with current understanding; and (iii) to define a methodology by which we can build on from this model to identify the appropriate balance of details for making better constrained probabilistic projections of the carbon cycle into the future.

## 2 Data sources

### 2.1 Carbon stocks and fluxes

Our empirical data primarily came from field-data collation initiatives that targeted an individual terrestrial carbon stock or flow. These are summarised in Table 1. These data sets were selected on the basis that they (i) were informative about the stocks and fluxes of carbon in natural terrestrial vegetation, (ii) contained at least some data that were representative of vegetation in a state of dynamical equilibrium (selected via a filtering process; see Appendix A) (iii) could be used as information to constrain parameters in our model, (iv) had approximately global coverage, (v) could have single latitude and longitude coordinates assigned each a site-based estimate to enable cross-referencing to spatial climate data, and (vi) could be easily accessed and shared alongside our study to enable reproducibility, investigations of data processing steps, investigations into the importance of the selected data, and controlled comparisons of alternative models. Full details of how all of the empirical data sets were processed are given in Appendix A. Some of these data sets have been superseded by more recent data sets (e.g. our fire data set could now be replaced by the Global Fire Emissions Database (GFED) of Giglio et al., 2010). We hope to incorporate these improved data sets in the future.

### 2.2 Environmental data

All model components incorporated information from site-specific environmental variables to make predictions, either directly (e.g. the effect of temperature on net primary productivity) or indirectly by requiring input from another component model that itself required environmental data (e.g. the allocation to woody plant parts depends on net primary productivity). All environmental variables were calculated using data contained in either or both of the New et al. (2002) gridded monthly climate data set and the Global Soil Data Task Group (2000) “Global Data Set of Derived Soil Properties” data set. These have spatial resolutions of 10 arcmin and 0.5 decimal degrees, respectively. Full details of how the different environmental variables were calculated are given in Appendix B. Again, we hope these data sets will be upgraded in the future.

### 3 Model

#### 3.1 Full structure

We developed a terrestrial carbon model as a set of six carbon pools, connected by various flows (illustrated in Fig. 1). Mathematically, this corresponds to a series of six ordinary differential equations (one for each pool), each with three general components: an input rate (e.g. carbon fixation rate); an output rate (e.g. leaf mortality); and a carbon content (e.g. leaf carbon). The chosen level of complexity was biased by our ability to identify global data on at least two out of these three for each carbon pool, so that it would be possible to infer properties of the third. The full terrestrial carbon model is then expressed as follows:

$$\frac{dC_l}{dt} = \boxed{G} (1 - f_{\max} \boxed{f_s}) \frac{\boxed{\mu_l}}{\boxed{\mu_l} + \boxed{\mu_r}} - (\boxed{\mu_l} + \boxed{\mu_f} + \boxed{\mu_s}) C_l, \quad (1a)$$

$$\frac{dC_r}{dt} = \boxed{G} (1 - f_{\max} \boxed{f_s}) \frac{\boxed{\mu_r}}{\boxed{\mu_l} + \boxed{\mu_r}} - (\boxed{\mu_r} + \boxed{\mu_f} + \boxed{\mu_s}) C_r, \quad (1b)$$

$$\frac{dC_s}{dt} = \boxed{G} f_{\max} \boxed{f_s} - (\boxed{\mu_s} + S_f \boxed{\mu_f}) C_s, \quad (1c)$$

$$\frac{dC_m}{dt} = \boxed{f_m} \left( (\boxed{\mu_l} + \boxed{\mu_s}) C_l + (\boxed{\mu_r} + \boxed{\mu_f} + \boxed{\mu_s}) C_r \right) - k_m \boxed{A} C_m, \quad (1d)$$

$$\frac{dC_a}{dt} = (1 - \boxed{f_m}) \left( (\boxed{\mu_l} + \boxed{\mu_s}) C_l + (\boxed{\mu_r} + \boxed{\mu_f} + \boxed{\mu_s}) C_r \right) + \boxed{\mu_s} C_s - k_a \boxed{A} C_a, \quad (1e)$$

$$\frac{dC_b}{dt} = F_1 k_a \boxed{A} C_a - k_b \boxed{A} C_b, \quad (1f)$$

where  $C_l$ ,  $C_r$ ,  $C_s$ ,  $C_m$ ,  $C_a$ , and  $C_b$  are the amounts of organic carbon stored ( $\text{kg m}^{-2}$ ) in leaves, fine roots, structural plant parts, metabolic fraction of the soil, structural fraction of the soil and recalcitrant fraction of the soil (these are defined below);  $t$  is time in years (yr); symbols marked with a box are functions that are described below and fully defined in Appendix C;  $f_{\max}$  is the maximum fraction of net primary productivity allocated to structural plant parts;  $S_f$  scales the fire-induced mortality rate of structural plant parts relative to that of leaves and fine roots (inferred in this study);  $k_m$ ,  $k_a$  and  $k_b$  are the maximum loss rates of the metabolic, structural and recalcitrant soil fractions ( $\text{yr}^{-1}$ ; all inferred); and  $F_1$  is the fraction of the structural carbon pool that does not decompose directly to carbon dioxide but enters the recalcitrant pool (unitless; inferred).

The terms in boxes in Eq. (1) are functions that apply at a given location and given time, which depend on the physical

environment at that location and time. Our principal aim was to infer the environmental dependence for each of these functions. However, we also considered, for each component, a null model with no climate dependency (i.e. a constant applying to all locations and times: see below). We do not suggest that these are the “best” representations of the model component functions (boxes in Eq. 1). Rather, we identified reasonable functional forms from the literature, which could be used simply to infer climate-dependent relationships.  $\boxed{G}$  determines the plant carbon fixation rate ( $\text{kg m}^{-2} \text{yr}^{-1}$ ), based on the MIAMI model of Leith (1975).  $\boxed{f_s}$  scales the fraction of carbon allocated to structural plant parts over leaves and roots (unitless) and is new to the literature. The mortality rates of leaves ( $\text{yr}^{-1}$ ),  $\boxed{\mu_l}$ , is predicted from a new model based on the recent analysis by van Ommen Kloeke et al. (2011) of global patterns of leaf lifespan. The mortality rates of fine roots,  $\boxed{\mu_r}$ , whole plants  $\boxed{\mu_s}$  ( $\text{yr}^{-1}$ ), and the fraction of organic carbon in dead leaves and fine roots that enters the metabolic soil organic carbon pool  $\boxed{f_m}$  (unitless) are based on simple linear functions. Carbon loss rate due to fire  $\boxed{\mu_f}$  ( $\text{yr}^{-1}$ ) is based on the models of Thonicke et al. (2001), Kloster et al. (2010) and Arora and Boer (2005).  $\boxed{A}$  scales the decomposition rate of organic soil carbon based on classic climate-dependent relationships (Ise and Moorcroft, 2008; Schimel et al., 1996; Bolker et al., 1998; Adair et al., 2008).

When inferring the parameters, we further assume that all carbon pools (and thus all stocks) at the specific locations and times for the empirical data have reached equilibrium (the empirical data were filtered as described in Appendix A). Equation (1) then reduces to simple expressions for the equilibrium carbon contents of plant and soil carbon pools (omitted for brevity).

## 4 Parameter estimation and model assessment

### 4.1 Computational framework

We built a computational framework to enable the assembly, parameterisation and assessment of multi-component models of arbitrary complexity to enable our study to be conducted (illustrated in Fig. 2). The framework, model, and derivative data necessary to reproduce the results of this paper, as well as a user’s guide, are available from [research.microsoft.com/en-us/downloads/](http://research.microsoft.com/en-us/downloads/). The data resulting from conducting the analyses described in this study are available from [research.microsoft.com/en-us/downloads/](http://research.microsoft.com/en-us/downloads/).

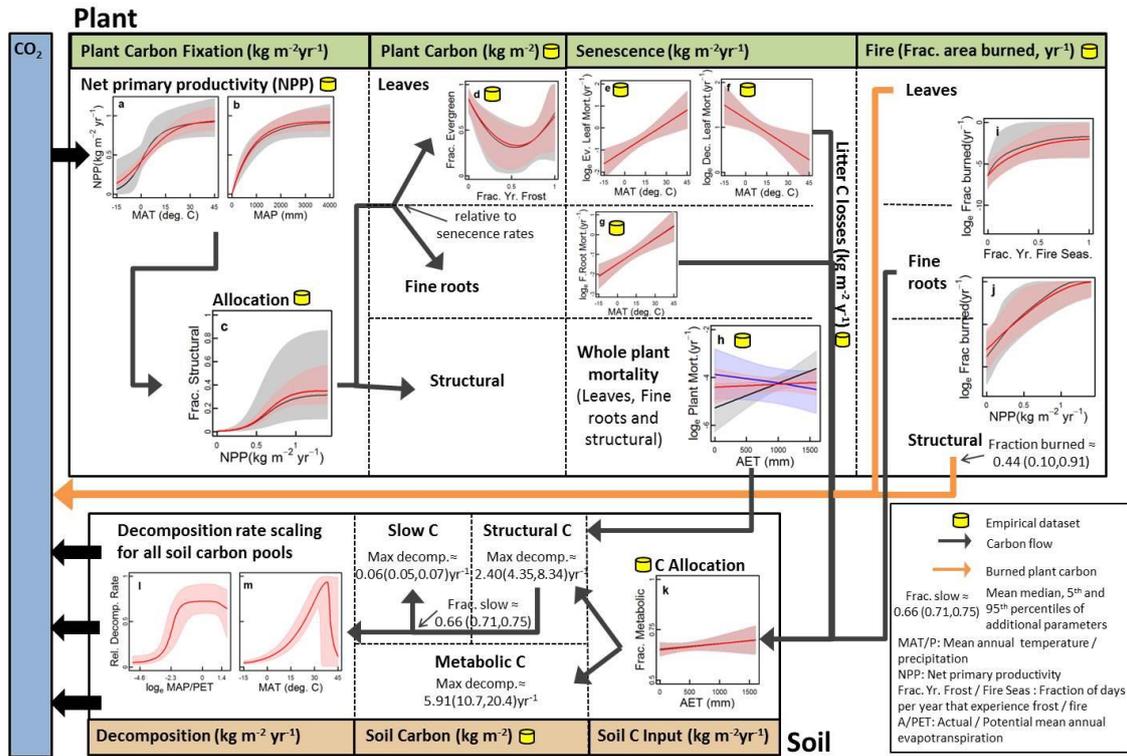
### 4.2 Data partitioning into training, evaluation and final test sets

The data sets were partitioned into training, evaluation and final test sets to avoid including parameters only because they help explain fluctuations specific to a particular data set, instead of the general phenomenon being inferred from the data

**Table 1.** Source data on carbon stocks and flows used in our study for model training and evaluation.

Data set name	Data description and typical units	Data provider	N. data	Likelihood function
Plant carbon: global biomass carbon map in the year 2000	The amount of carbon held in terrestrial vegetation, tonnes carbon ha <sup>-1</sup>	Carbon Dioxide Information; Analysis Center (cdiac.ornl.gov/epubs/ndp/global_carbon/carbon_documentation.html); Ruesch and Gibbs (2008)	1000*	Normal
Litter carbon production rate	Litter production rates, g dry matter m <sup>-2</sup> yr <sup>-1</sup>	Matthews (2003)	1000*	Normal
Soil carbon: global gridded surfaces of selected soil characteristics (IGBP-DIS)	Soil carbon density (kg m <sup>-2</sup> ) at a depth interval of 0–100 cm	Oak Ridge Nat. Lab. Distrib. Active Archive Center (ORNL DAAC) (daac.ornl.gov/SOILS/guides/igbp-surfaces.html); Global Soil Data Task Group (2000)	1000*	Lognormal
Plant carbon fixation rate: “Class B site” net primary productivity (NPP)	Net primary productivity (kg carbon m <sup>-2</sup> yr <sup>-1</sup> )	Oak Ridge Nat. Lab. Distrib. Active Archive Center (ORNL DAAC) (daac.ornl.gov/NPP/html_docs/EMDI_des.html); Olson et al. (2001)	933	Normal
Deciduous leaf mortality rate	Estimated lifespan (in months) of deciduous leaves	GLOPNET Authors Wright et al. (2004)	30	Lognormal
Evergreen leaf mortality rate	Estimated lifespan (in months) of evergreen leaves	GLOPNET Authors Wright et al. (2004)	46	Lognormal
Fraction of leaves that are evergreen	Categorical classification of leaves as “evergreen” or “deciduous”	GLOPNET Authors Wright et al. (2004)	155	Normal
Fine root mortality rate (lifespan)	Mean root turnover (yr <sup>-1</sup> )	Gill and Jackson (2000)	162	Lognormal
Plant mortality rate	Forest turnover rates (yr <sup>-1</sup> ) from different sites worldwide	Stephenson and van Mantgem (2005)	191	Lognormal
Global map of fraction of area burned per year, 1900–2000	Percentage of a grid cell burned per year for 100 yr (1900–2000)	Florent Mouillot (cefe.cnrs.fr/fe/staff/Florent_Mouillot.html); Mouillot and Field (2005)	1000*	Lognormal
Fraction of leaf and fine root carbon entering fast soil pool (that is “metabolic”)	Fraction of leaf and fine root carbon that is decomposed quickly by soil organisms (fraction)	Ise and Moorcroft (2006)	1000*	Normal
Global land cover in the year 2000. Used to infer fraction of plant carbon allocated to structural parts	Discrete classifications of land cover types represented as integer codes	European Commission (bioval.jrc.ec.europa.eu/products/glc2000/data_access.php); Bartholome and Belward (2005)	1000*	Logistic

Numbers represent those model parameters specifically needed to predict a particular data set. Some models may have taken the outputs of other models as their inputs, and therefore may implicitly include more model parameters. \* Approximate numbers of data points obtained through random stratified sampling of gridded global data.



**Fig. 1.** Summary of the inferred climate dependence of the terrestrial carbon cycle within our terrestrial carbon model. The inferred climate dependence of each component is shown in red (inferred using the full model and all data sets), grey (inferred using minimal subsets of the model) and, to illustrate the major structural uncertainty in plant mortality (*h*), in blue (inferred using the full model but omitting the plant mortality data). Lines and shading are average median, 5th and 95th percentiles from 10 training data subsets.

set (over-fitting). Over-fitting can still occur when adopting this approach if model refinement goes through many iterations and the same training and evaluation sets are used. We therefore first removed a fraction of each data set to be used as a final step to assess the performance of our models (the “final test data”). This allows us to test our models against data that played absolutely no role in the model refinement process. We constructed a land surface mask by randomly positioning 0.5 degree squares over the terrestrial land surface until approximately 25 % of the terrestrial land surface had been covered. Any data that fell under this mask were removed permanently as final test data. We performed 10-fold cross-validation within our model parameter inference experiments on the data remaining after the removal of the final test data.

### 4.3 Parameter inference

We used a Bayesian approach to infer the probability distributions for the model parameters given our empirical data sets. For every model we used flat (or “uninformative”) prior probability distributions for the parameter values. We assume that the probability of observing the data under all possible hypotheses is 1 (the marginal probability). We also made no

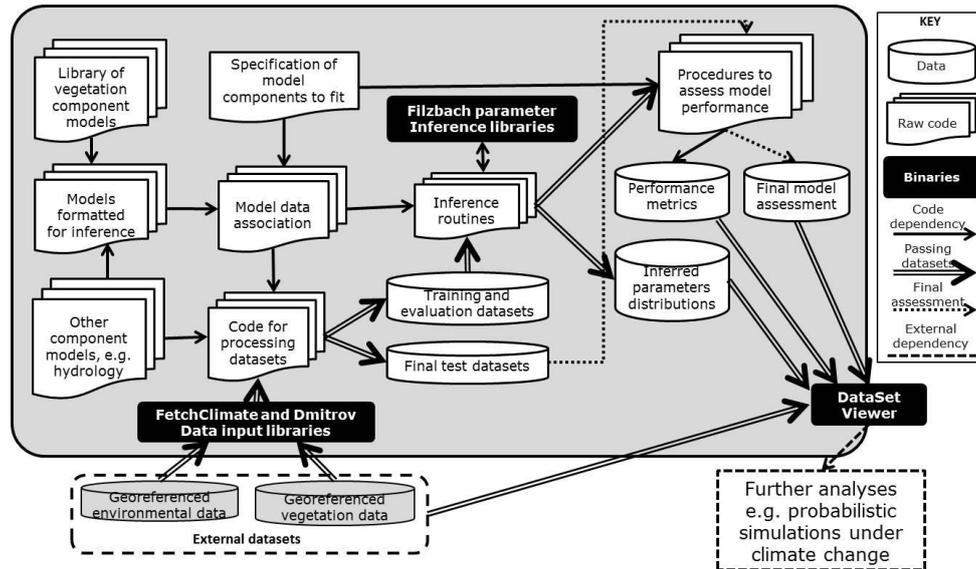
attempt in this study to incorporate or infer errors and biases associated with the observational data. This is potentially an important assumption, and accounting for these errors will be enabled in the future for empirical data sets that contain estimates of observational uncertainty

Under these assumptions, estimating the probability distributions of the parameters reduces to requiring the estimation of their likelihood, given the observed carbon data and environmental conditions. This required us to specify a likelihood function for each model component, which defines the probability of the data, given any combination of parameters. With this function defined, the posterior probability of each parameter could be estimated. Formally, we assume that

$$L(\mathbf{Pred}(\text{Model}(\mathbf{Pars}, \mathbf{Env})) | \mathbf{Obs}) \propto P(\mathbf{Obs} | \mathbf{Pred}(\text{Model}(\mathbf{Pars}, \mathbf{Env}))),$$

where bold text denotes a vector. In words, the likelihood *L* of the predictions **Pred** of the parameterised model given the observations, **Obs**, is proportional to the probability *P*, of the observations given the particular model predictions. The predictions arise from a particular model with parameters **Pars** and set of environmental conditions **Env**.

We used the Filzbach set of code libraries to find the posterior distributions for the parameters of a given model, given the observed carbon data (**Obs**) and environmental



**Fig. 2.** Our prototype automated system used for the construction, parameter estimation and assessment of multi-component models of arbitrary complexity. This automated system was used to develop and analyse our model. Inputs are multi-component models and empirical data. Data are partitioned into training, evaluation and final test sets. Input libraries provide a common interface to connect code with data. Inference routines use inference libraries to infer parameter probability distributions. Inference libraries implement approximate Bayesian inference using Markov chain Monte Carlo methods with the Metropolis–Hastings algorithm. A data viewer provides a standard interface for inspecting inputs and results. See Sect. 4.1. for details on how to download the framework.

conditions (**Env**; <http://research.microsoft.com/en-us/um/cambridge/groups/science/tools/filzbach/filzbach.htm>). Filzbach implements Markov chain Monte Carlo sampling of parameter space given a set of parameters to be varied (**Pars**) and likelihood function ( $L$ ; Gilks et al., 1996). It uses the Metropolis–Hastings algorithm to accept or reject sets of parameter values when compared to the likelihood associated with the parameter values of the previous iteration of the Markov chain (Gilks et al., 1996). In our study, the likelihood function used by Filzbach may depend on the likelihoods associated with several sub-component models, depending on the model parameter inference experiment being run (outlined below, and see Fig. 3 for details). The specific likelihood function chosen to assess each model component against its corresponding empirical data set is detailed in Table 1.

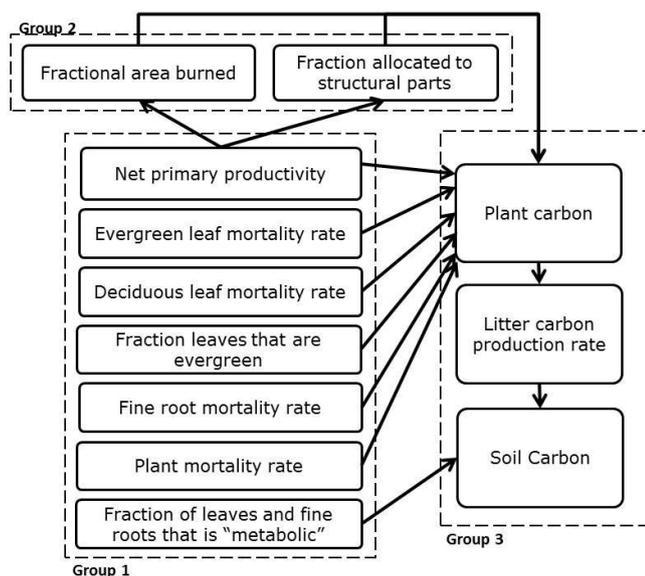
The data sets used contain different relative frequencies of data for different climate regions of the world. We down-weighted the log-likelihoods assigned to data points in direct proportion to the relative frequency of data in their respective Holdridge life zones (Holdridge, 1967). This avoids biasing the model parameter inference procedures towards parameter values that predict well those regions of the world that are most frequently represented in the data.

#### 4.4 Parameter inference experiments

We investigated the sensitivity of the inferred model functional forms and model performance metrics to using different combinations of model components and data sets by conducting three different parameter inference experimental protocols described below.

##### 4.4.1 Build-up experiments

We inferred parameters to each of the component models indicated in Fig. 3 alongside those of the models on which they depend to make predictions. We refer to these experiments as “build-up”, because we started by inferring the parameters to the group 1 models individually and then inferred those to the group 2 models (alongside those of the NPP model), before inferring parameters to each of the group 3 models (and those of the sub-components on which they depend), incrementally working towards inferring all the parameters in all components of the terrestrial carbon model simultaneously (the 12th experiment).



**Fig. 3.** The full terrestrial carbon model can be represented as a factor graph. All boxes represent model components with accompanying data. Arrows connect a model that acts as a subcomponent (tail of arrow) to another model (head of arrow). Models within group 1 do not require predictions from other models to predict their accompanying data sets. Group 2 models require predictions from the net primary productivity model. Group 3 models require input from several model components.

#### 4.4.2 Omit-data experiments

We sequentially omitted an entire data set associated with each model component in Fig. 3 prior to inferring the parameters of the full model. This enables investigation of how important the information contained in a given data set is for the inferred parameter probability distributions. Each evaluation step also included a fold of the omitted data sets. However, this approach does not allow us to estimate process error associated with a given model in the absence of its associated empirical data. We therefore used the posterior parameter estimates of process error obtained from inferring the parameters of the full model with all empirical data sets when performing model evaluation.

#### 4.4.3 Replace-null experiments

We sequentially replaced a model component in Fig. 3 with a null model consisting of an inferred constant and associated error. This allows us to investigate how important the climate dependency of a particular model is for the predictive performance of other components and their inferred parameter values.

#### 4.4.4 Assessing predictive performance

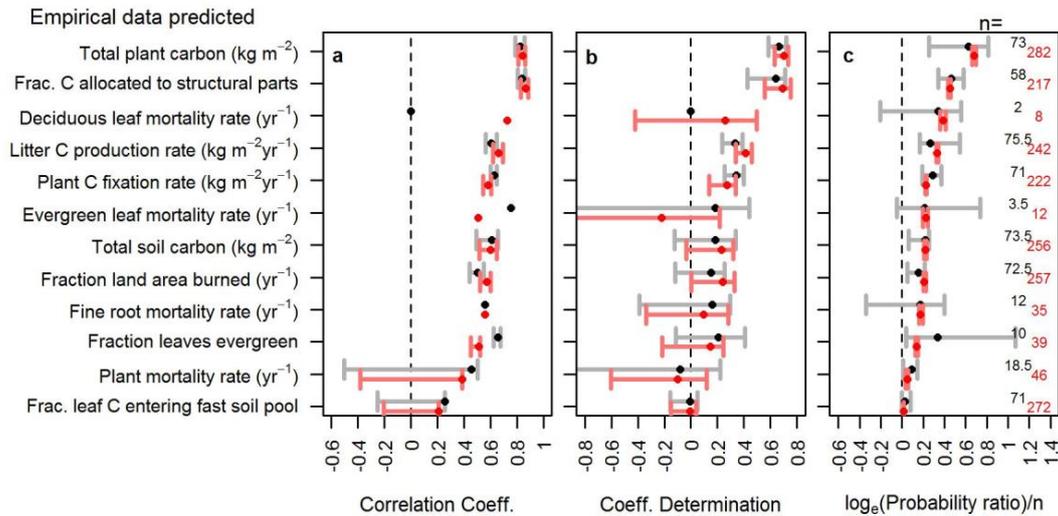
In general, we calculated performance metrics for each sample of parameter values from the Markov chain after the burn-in procedure. The burn-in period was always 20 000 iterations, and the Markov chain length used for parameter sampling was always 120 000 iterations and was subsampled every 100 iterations. This enabled us to also take mean, median and 95th percentiles of various performance metrics over the set of sampled parameter values.

## 5 Results

### 5.1 Full model

Overall we infer climatically varying functions for all component processes to our terrestrial carbon cycle model when fitted using all available data sets (red functional relationships in Fig. 1). The inferred climate dependencies of net primary productivity (NPP), increasing but saturating functions of temperature and precipitation (Fig. 1a, b), are consistent with what was established for the classic MIAMI model (Leith, 1975). The proportion of fixed carbon allocated to wood (versus leaves and fine roots) varies continuously as a sigmoid function of NPP and increases to around 0.35 for the most productive locations (Fig. 1c). The four processes then determining carbon loss rates have contrasting climate dependencies (Fig. 1d–h). Fire increases with dry season length (combustibility) and NPP (fuel), as expected (Fig. 1i, j; Kloster et al., 2010). Contrasting dependencies of evergreen and deciduous leaf mortality (Fig. 1e, f) are inferred, with the relative frequency of evergreen versus deciduous leaves being U-shaped against annual frost frequency (Fig. 1d). This highlights the relatively complex climate dependence of leaf lifespan globally (van Ommen Kloeke et al., 2011). Fine root mortality rate is inferred to increase with mean annual temperature as expected (Fig. 1g; Gill and Jackson, 2000). A relatively flat relationship for the climate dependency of plant mortality is inferred (Fig. 1h); however, this actually results from contradictory information from different data sets under our assumed model formulation (details below). We infer no strong climate dependency in the fraction of dead leaves and roots initially allocated between the different soil pools (Fig. 1k). For the soil component, we infer temperature and moisture dependencies of the classical three-pool soil model that are consistent with previous findings (Fig. 1l, m; Ise and Moorcroft, 2008).

The lack of a relationship for plant mortality was unexpected, because a previous study (Stephenson and van Mantgem, 2005), using the same empirical data on plant mortality rates, identified a positive relationship between mortality rates and productivity – a close correlate of evapotranspiration rates (Leith, 1975). Further analysis reveals this inconsistency to be due to differences in the information implied



**Fig. 4.** Performance assessments of the terrestrial carbon model: (a) Pearson's correlation, (b) the coefficient of determination and (c) the probability of the model relative to a null model (Gelfand and Day, 1994). Separate data subsets were used for parameter inference and model evaluation to avoid over-fitting. A final test data subset was reserved (never used during model development) to provide an independent estimate of the likely predictive ability when the model is applied to locations that have not been observed. Assessments against evaluation and final test data are in grey and red, respectively, with  $n$  being the respective mean or absolute number of data points per assessment. Dots and error bars are average medians, 5th and 95th percentiles in (a) and (b) and medians, maxima and minima in (c) using parameter distributions inferred from 10 training data subsets. Insufficient evaluation data existed to calculate (a) and (b) for the deciduous leaf mortality model.

by different empirical data sets under our assumed model formulation. We infer qualitatively different climate dependence from the plant mortality data alone (a positive relationship, Fig. 1h (grey)), as found by Stephenson and van Mantgem, 2005), from all model components and empirical data sets together (a flat relationship, Fig. 1h (red)), or using all model components but omitting individual data sets (omitting plant mortality data gives a negative relationship, Fig. 1h (blue), omitting the plant carbon data gives a positive relationship, similar to Fig. 1h (grey)). These results indicate a clear discrepancy between the information on the climate dependency of plant mortality implied by the mortality data itself and that implied by the other plant carbon data sets under the assumed model formulation. On this basis we identify global plant mortality rates as a major “structural uncertainty” in our terrestrial carbon model.

Other than the non-climatically dependent functions, the climate dependencies inferred for the full terrestrial carbon model tend to make predictions that both are significantly positively correlated with the evaluation data sets (Fig. 4a), and tend to explain a positive fraction of the variation within each data set (Fig. 4b). The plant carbon data set is predicted particularly well (final test data median Pearson's  $r = 0.84$  (5% and 95% confidence intervals = 0.81, 0.86), coefficient of determination = 0.70 (0.63, 0.77)), as are data on litter production rate, plant carbon fixation rate and the fraction of carbon allocated to structural plant parts, for all of which the model always explains a positive fraction of the varia-

tion in the data at a 95% confidence level (Fig. 4b). Comparing the performance of the full model to one in which the relevant model component is replaced with a null model supports choosing the climate-dependent model for all processes (Fig. 4c) except for the two component processes for which we inferred no climate dependencies.

When applied at global scales, the terrestrial carbon model predicts global patterns of equilibrium plant and soil carbon that match the known patterns well (Fig. 5a, b; calculated using the New et al., 2002; and Batjes, 2000 gridded data sets for environmental variables). We estimate that the absolute uncertainty is positively related to the median (Fig. 5c, d) for both carbon pools. For plant carbon, relative uncertainty (absolute uncertainty/median prediction) tends to be higher in areas in which the model predicts vegetation that is intermediate between being maximally woody and completely non-woody (Fig. 5e), owing to the contrasting mortality rates of these different carbon pools (Fig. 1e–h). Relative uncertainty is also higher over Greenland, and we expect this is because of the inflation of uncertainty under extrapolation: the extreme environments in that region are out of the ranges represented in our data (Fig. 5e). We believe a similar phenomenon explains the relatively higher uncertainty in predictions of soil carbon in extremely dry or cold environments (Fig. 5f).

The inferred climate dependencies in the terrestrial carbon model, other than the plant mortality function, generally support those that have been found in previous studies,

indicating their consistency when considered as part of the overall system. Although the qualitative nature of many of these dependencies is consistent with previous knowledge, our estimates of parameter values for these functions (and the associated uncertainty in those parameter estimates) are new. Our analysis also highlighted a number of new insights into the performance of these climate-dependent functions as detailed below.

Although the inferred climate dependencies of NPP (Fig. 1a, b) are consistent with the MIAMI model (Leith, 1975), one parameter is not well constrained by the data: the parameter controlling NPP at zero degrees Celsius ( $t_1$  in Eq. C1b) does not converge. Instead, the sigmoid response of NPP to temperature is constrained by the parameter controlling the gradient of the temperature dependency,  $t_2$ . This implies that the temperature-dependent function may actually be over-complex for our purposes. Further investigations (omitted for brevity) involving removing this parameter and refitting the model reveal that this and other under-constrained parameters discussed below have little effect on model predictive accuracy. Plots of predictions versus observations also reveal a noisy positive relationship, with the model underestimating NPP for sites with NPP greater than around  $1.0 \text{ kg m}^{-2} \text{ yr}^{-1}$  – a property that has been noted previously for the MIAMI model (Fig. 6a; Friedlingstein et al., 1992; Dai and Fung, 1993).

The shape of the function controlling the proportion of fixed carbon allocated to structural components implies that structural tissue only makes up around 10% of vegetation carbon at NPP values of around  $0.5 \text{ kg m}^{-2} \text{ yr}^{-1}$  (Fig. 1c). We expect this probably underestimates structural carbon in vegetation types dominated by low-productive woody vegetation such as some boreal forests (Kicklighter et al., 1999), although we have not verified this.

The wide confidence intervals in the function predicting the fraction of vegetation leaves that are evergreen imply relatively high uncertainty, and inspection of the relationship between predictions and observations makes clear why this is the case, with several observations lying far from the 1 : 1 line (Fig. 6b). Despite this, the correlation between observations and predictions is relatively high (Fig. 4a), probably due to the dominance of sites in the data set that are either entirely evergreen (44%) or entirely deciduous (14%). We anticipate that the low quantity of data in the data sets on leaf characteristics (the most sparse data in our collection; Table 1) strongly influences the variation in model predictive performance for those climate dependencies.

Our inference of the climate dependencies of fire frequency globally again highlights some redundant model parameters: the scaling constant  $c_f$  and the two half saturation constants,  $l_{\text{halfsat}}$  and  $\text{NPP}_{\text{halfsat}}$ , are poorly constrained. This implies that the model could be reformulated with fewer parameters and still predict the data with the same accuracy. Although the correlation between predictions and observations is relatively strong for this data set (Fig. 4a, b), visual

inspection of observational data versus predictions implies a lower predictive performance at low fire return intervals (fraction of burned per year) and a tendency to underestimate the fire return interval overall (Fig. 6g).

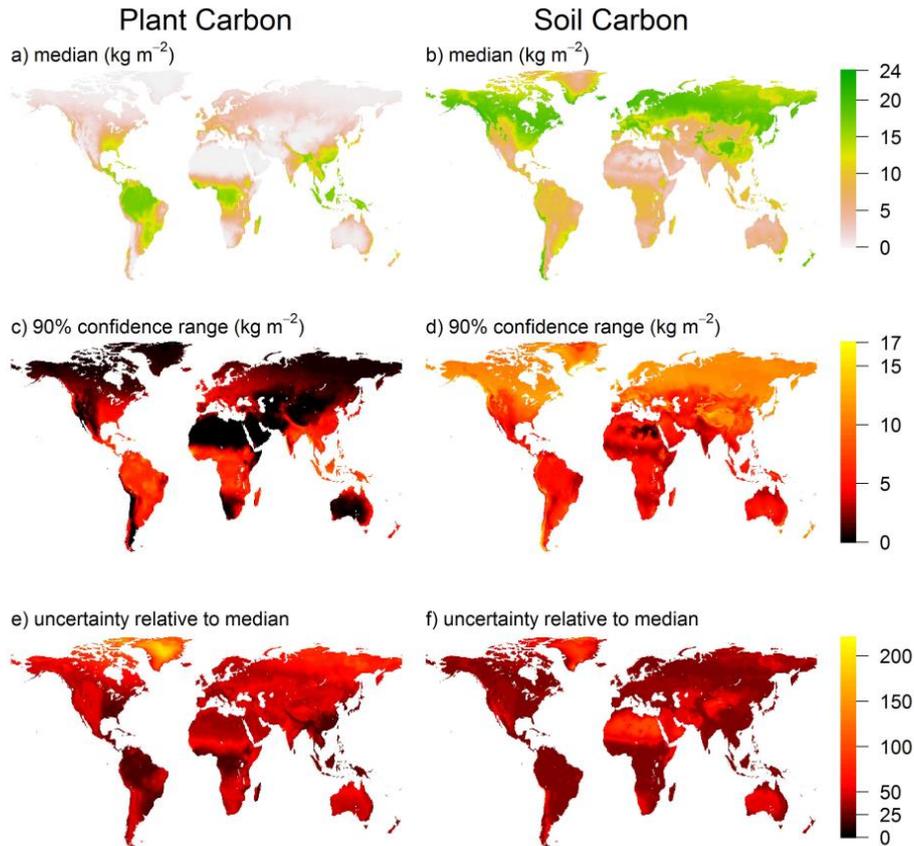
Although plant carbon is predicted well, the plots of predictions versus observations indicate some notable outliers at low carbon contents, where carbon is predicted to be much higher than observed for some sites (Fig. 6h). These sites appear to be associated with tropical vegetation that has been classified as grasslands and shrublands in the global land cover map and have been assigned low carbon content (data omitted for brevity), in contrast to the predictions of the model.

For the soil model we could, again, not fully constrain all model parameters. The parameter downscaling soil decomposition rates as a function of extremely high temperatures ( $t_{\text{sc}}$  in Eq. C8a) and the parameters controlling the optimum moisture content for decomposition and downscaling parameter of decomposition rates in extremely wet conditions ( $m_{\text{sc}}$  and  $m_{\text{thresh}}$ , in Eq. C8c) are probably all poorly constrained, a result of lacking sufficient data representing such extreme environments.

## 5.2 Build-up experiments

The build-up experiments show that the performance of some model components changes as they become part of larger model structures (Fig. 7). The major result from these experiments is the qualitative change in the inferred climate dependency of plant mortality upon being connected to the full model, as mentioned above (Fig. 1k). The consequent change in model predictions is clearly seen in the plots of predictions versus observed data in which a noisy positive relationship between predictions and observations is apparent when the plant mortality model is fitted to the plant mortality data alone, whereas a relatively flat relationship is observed for the model fitted as part of the full model structure (Fig. 6k, l).

The net primary productivity model (NPP) improves in predictive performance when it is connected to the full model, showing higher correlation coefficients and an improved fit to the training data (lower deviance information criterion values; Fig. 7). Lower uncertainties in NPP functional forms are also visible for the full model compared to when the model is fitted to the NPP data alone (grey versus red in Fig. 1a, b). This implies that information from other model components helps to further constrain the climate dependence of NPP. The component predicting the fraction of plant material that is woody exhibits a poorer fit to the training data as it is connected to other model components but still slightly increases in predictive performance (Fig. 7). Connecting this component to plant carbon also enables the parameter controlling the maximum carbon allocation fractions to wood to be inferred (grey versus red in Fig. 1c). Connecting the model predicting litter production to the full model structure improves its fit to the training data, whereas the



**Fig. 5.** Predicted global distributions of equilibrium plant and soil carbon. These match the known patterns well based on visual inspection. (a), (b) Mean median prediction of plant and soil carbon from parameter distributions inferred from 10 training data subsets. (c), (d) Uncertainty range spanned by the 5th and 95th % confidence intervals. (e), (f) Uncertainty relative to the median. Spatial resolution is 10 arcmin (18.5 km at the Equator).

opposite is the case for the model predicting fraction of land area burned (Fig. 7). However, none of these effects significantly alter the correlation between predictions and observations in the evaluation data, and they only cause minor reductions in the confidence intervals of the functional forms of the fire model (grey versus red in Fig. 1i, j).

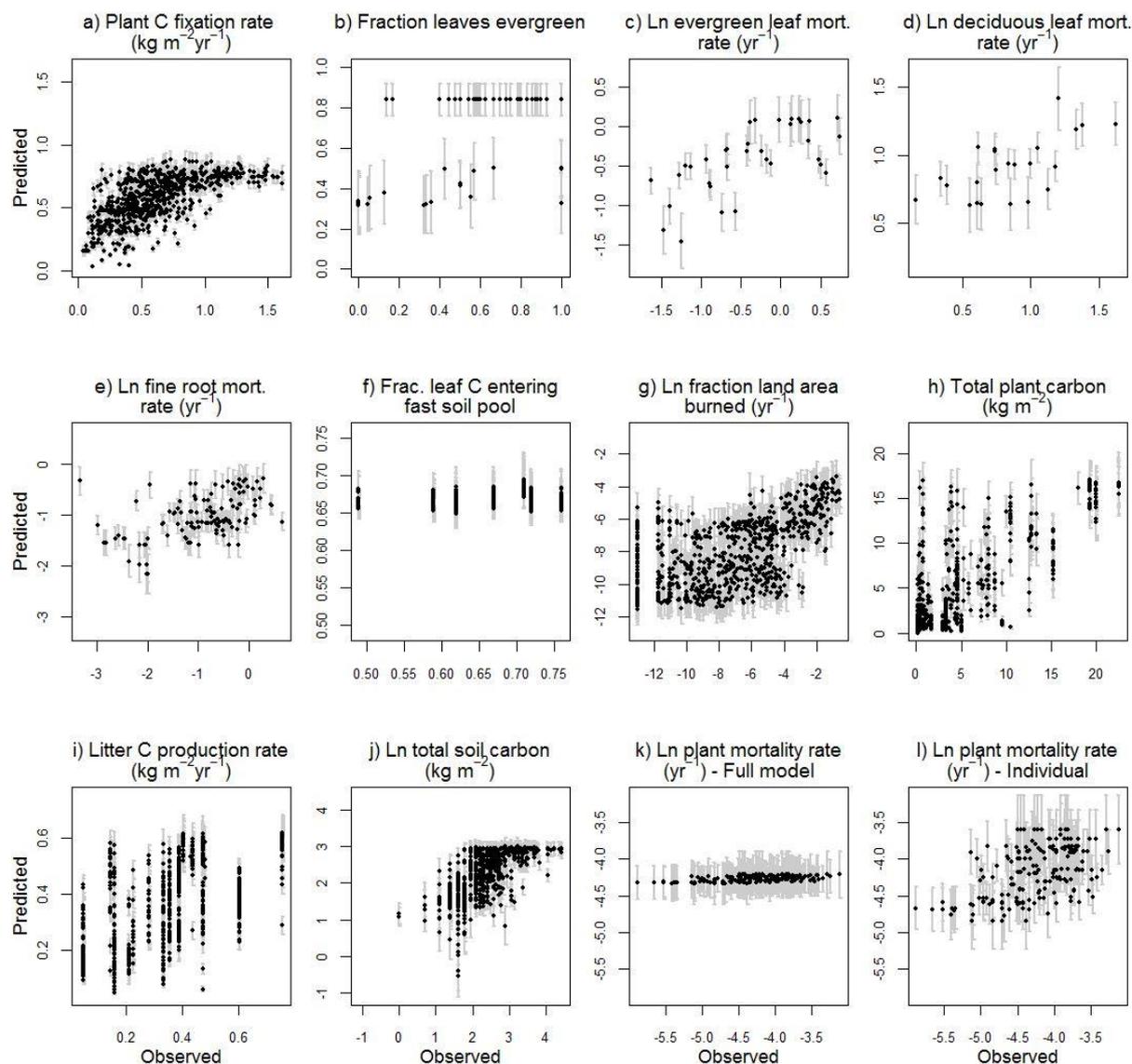
### 5.3 Omit-data experiments

The most notable effects of omitting data sets when inferring parameters to the full models were on the parameters of the plant mortality climate dependency, as reported above. Otherwise, omitting individual data sets when fitting the full model does not dramatically influence the performance of the model at predicting the data sets that had not been removed (we omit details for brevity). Constraining the parameters of some components is entirely dependent on the presence of their corresponding data set. This is the case for the fraction of leaves that are evergreen, the mortality rates of evergreen and deciduous leaves, fine root mortality rates and soil decomposition rates and indicates that the prediction of other connected components, such as plant carbon for example, is

not dependent on the predictive accuracy of those components. In contrast, omitting the data for the NPP model still results in the inference of very similar parameter values, indicating a strong dependency on those parameter values for predicting other empirical data sets.

### 5.4 Replace-null experiments

In general, and as expected, replacing a component with a model predicting no climate dependency strongly influences the predictive performance of that component in the cases where evidence exists for a climate dependency (Fig. 8). For some components this results in negligible effects on the predictive performance of the rest of the model. However, for the NPP model, replacement strongly influences the predictive performance of other components. This is not surprising, given that it is the initial carbon input term for the plant carbon pools.



**Fig. 6.** Model predictions versus observed empirical data (omitting the 25 % final test data). Predictions were made using all 10 posterior parameter probability distributions from 10-fold model fitting. Points show the average median prediction, and error bars show the average upper and lower 95 % confidence intervals. (a–k) show predictions from the full model trained to all empirical data sets, and (l) shows predictions from the plant mortality model trained to the plant mortality data alone.

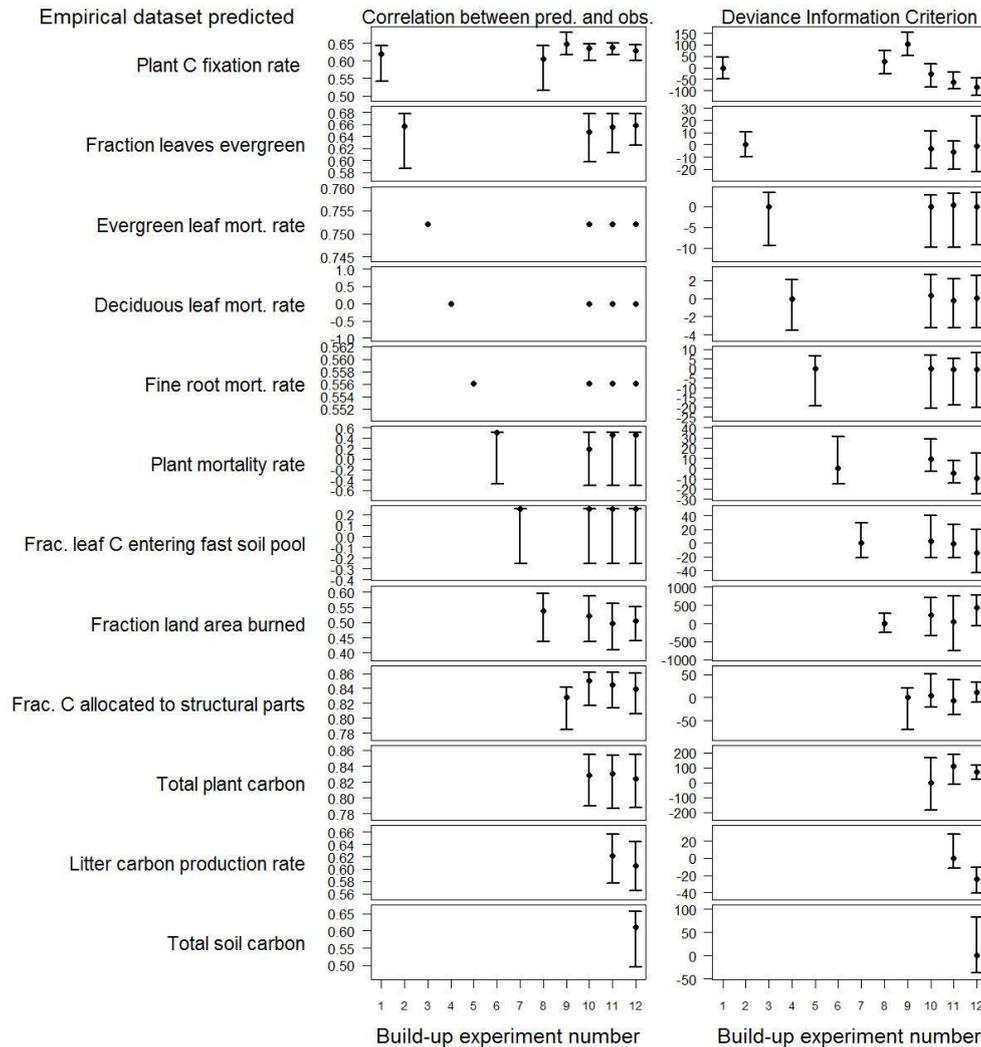
## 6 Discussion

### 6.1 The degree of empirical support for the component processes

One of our key aims was to establish a baseline terrestrial carbon model to support future work in which we had assessed the empirical support for every component process, including clearly characterising and incorporating parameter uncertainties. Our new methodology has clearly achieved that, and it also illustrates how the adoption of such a methodology in the future development of terrestrial carbon models could ac-

celerate the identification of where their key inconsistencies lie, both with each other and with empirical data.

We adopted a relatively simple model here, and we hypothesise that it is probably too simple and uncertain to provide informative projections of future terrestrial carbon dynamics. For example, while we adopted the MIAMI model as a convenient and well-recognised, simple model of net primary productivity, it has several well-known limitations (e.g. Bonan, 1993) including being insensitive to changes in atmospheric  $\text{CO}_2$  concentrations (discussed further below). Rather, what we aimed to do was to develop and implement methodology to enable quantitative comparisons of



**Fig. 7.** Correlations between model predictions and hold-out evaluation data (left) and deviance information criterion (DIC) values (right) obtained from different model fitting experiments (Gelman et al., 2004). DIC values have been normalised relative to the lowest build-up experiment number predicting the data set indicated. The presence of confidence intervals for a given experiment number indicates whether a model component was used in a model fitting experiment. Points represent the average median estimate using the parameter probability distributions obtained from 10-fold model fitting, and the whiskers are average 95 % confidence intervals.

alternative representations of component processes such as NPP in the future (e.g. those documented in Adams, 2004). Setting out to data-constrain all component processes also forced us to propose new models for some components, such as that for leaf mortality rates based on van Ommen Kloeke et al. (2011). Conducting model development in this way enables the empirical evidence for such new component formulations to be rapidly assessed and modified if necessary before becoming a longer-term feature of the model. The omit-data experiments also illustrated that the parameters of a new component can be estimated even without any direct data on the process in question, if enough data are included on other connected model components. Subsequently, the direct data,

if they become available, could further reduce uncertainty in the functional form and parameters for the new component.

Skill at predicting the data on the terrestrial carbon cycle at equilibrium, as we have shown here, is no guarantee that the model can accurately capture the temporal dynamics of the carbon cycle that other data assimilation studies have focussed on (Scholze et al., 2007; Ricciuto et al., 2011). For example, our model does not account for carbon dioxide fertilisation effects that are likely to have had, and could continue to have, a major influence on vegetation carbon fixation potential into the future (Friedlingstein et al., 2003). We therefore see a pressing need to build on our rigorous approach and begin the process of comparing alternative formulations of component models, model structures, and

different data sets. Such studies would obviously consider alternative representations of canopy photosynthesis (Purves and Pacala, 2008), including not only carbon dioxide fertilisation effects (Friedlingstein et al., 2003), but also the dynamics of other resources such as water and nitrogen (Goll et al., 2012), the importance of different plant functional types, vegetation traits, successional processes and permafrost carbon thawing, to name a few (see Arneeth et al., 2012; for a detailed discussion). Although we developed our model within a generalized framework so that it can become a foundation upon which future intercomparisons can be conducted (see Sect. 5.1), it will require further improvements to truly enable detailed intercomparisons of state-of-the-art terrestrial carbon models. Enabling the inference of the parameters of dynamical models using time-varying data, and the quantification, assessment and propagation of errors in the observational data are two key necessary refinements to enable such studies.

Future studies should also investigate the effects of incorporating more and improved data sets for data assimilation and model testing than we did here. We deliberately selected data sets that were relatively easy to obtain, process and share. It is reassuring that these were sufficient to infer most of the known climate dependencies. However, further constraining and testing the more detailed terrestrial carbon models will require new data sets that document temporal dynamics and, more generally, whichever data have the greatest potential to minimise uncertainty in model predictions. Here we made no use of the valuable FLUXNET (<http://fluxnet.ornl.gov/>) data sets on water and carbon fluxes nor the FPAR data ([http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD\\_NUMBER=15](http://modis.gsfc.nasa.gov/data/dataproduct/dataproducts.php?MOD_NUMBER=15)). We expect that incorporating such data could assist greatly in constraining and comparing any dynamical models developed in the future. There have already been developments in this direction, especially in constraining terrestrial carbon models with carbon flux data (Knorr and Heimann, 2001; Kaminski et al., 2012; Scholze et al., 2007; Ricciuto et al., 2011). Future developments should build from these studies to enable more detailed process-based models to be data-constrained within our framework, although it will be important to take a balanced approach to refining components and not, for example, focus on refining the representation of photosynthesis if it occurs at the expense of refining other components, such as mortality, which has received much less attention to date.

While we demonstrated here how multiple component processes can be data-constrained using multiple data sets, there is obviously a question about how to assess models when they are projected under future scenarios, for which no observational data can exist. This problem can never be solved fully, and in principle vegetation could exhibit changes, or suffer perturbations, that have never been observed, either naturally or under artificial perturbations. Our approach offers the promise of being able to identify model components for which improved understanding would be

most important, by decomposing changes in the output variables of interest (e.g. soil carbon) to changes in different processes (e.g. the effects of extreme temperatures on soil decomposition rates), and then finding out which data sets are most important for improving those components. Under our framework the results of key experiments, such as manipulative experiments, or of other passive observations could then be included as additional data constraints.

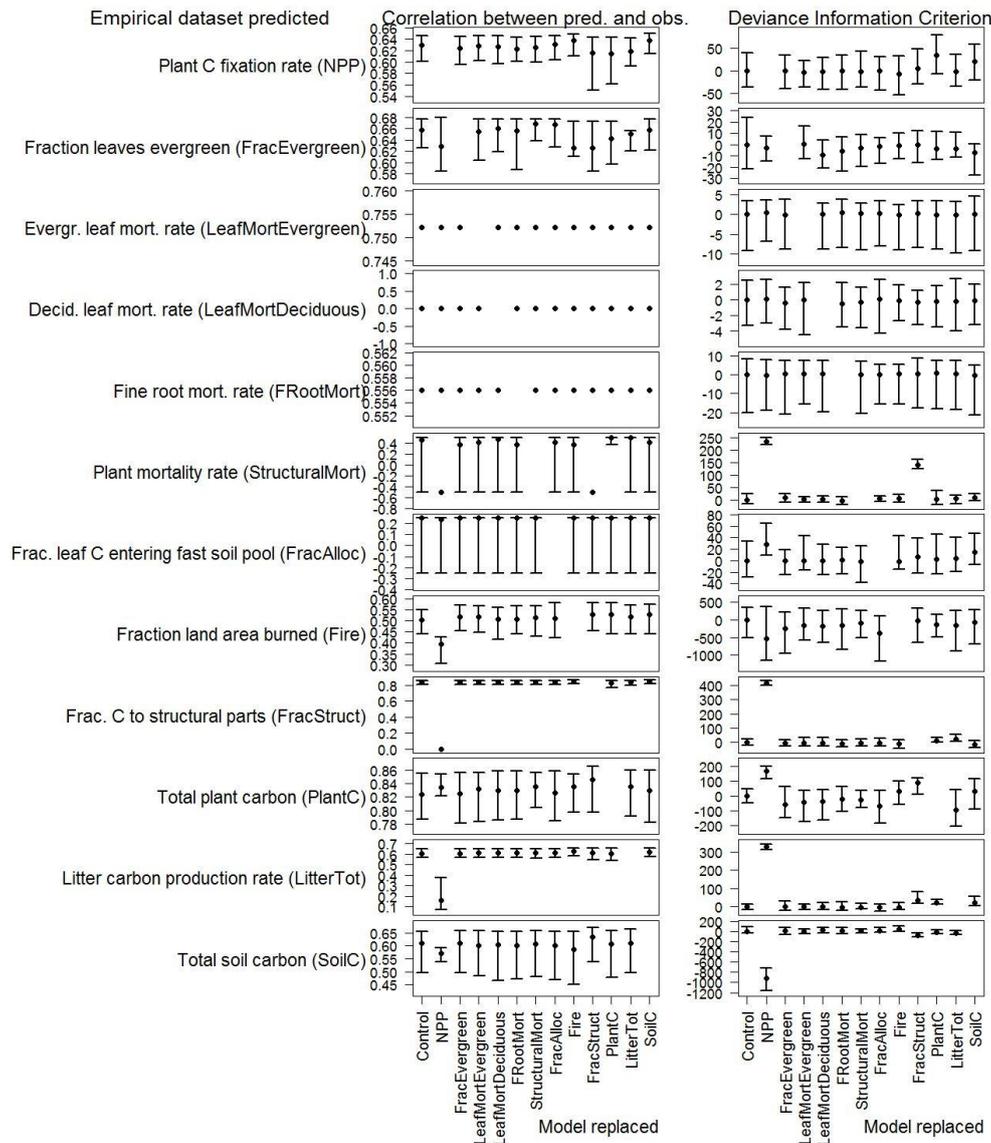
## 6.2 Inferred relationships in relation to current understanding

The climate dependencies inferred here mostly confirm those that have been identified previously. However, the identification of qualitatively different climate dependencies for plant mortality depending on the model formulation and empirical data used, as well as some other more subtle adjustments to other climate dependencies, highlights the value of the systemic approach: it enables us to assess how consistent our model of how the overall system functions is with empirical evidence and identify where discrepancies lie.

At present, we do not understand the reason for the structural uncertainty in the climate dependency of plant mortality rates. As a group, the plant carbon data sets imply that mortality rates decrease with actual evapotranspiration under our assumed overall model formulation, whereas the plant mortality data imply the opposite. Perhaps the high plant mortality rates observed in highly productive sites are inconsistent with the high carbon values recorded for those sites under the inferred climate dependencies. This could have caused the systematic underestimation of net primary productivity in highly productive sites (Fig 6a), leading to lower than expected predictions for plant carbon, which in turn could cause the inferred plant mortality rates for those sites to be higher than observed. There are however other possibilities, such as issues to do with the quality of the plant mortality data or the other plant carbon data sets, the possibility of plant mortality patterns obeying more complex relationships than represented by our mortality component, or missing processes in the model (such as the separation of structural carbon losses between whole plant mortality and the loss of other structural parts such as branches). These possibilities add to recent calls to increase understanding of the role of plant mortality in the global carbon cycle (Stephenson and van Matgem, 2005; van Matgem et al., 2009; Allen et al., 2010).

## 6.3 Building a model for future predictions

We do not know how accurate predictions of temporal dynamics of the carbon cycle from our terrestrial carbon model might be, because we have done no assessments of our model in predicting carbon dynamics. Nonetheless, we can use it to illustrate the concept of using a data-constrained model to inspect the relative importance of parameter and structural uncertainty in projections. For example, it could be that, despite

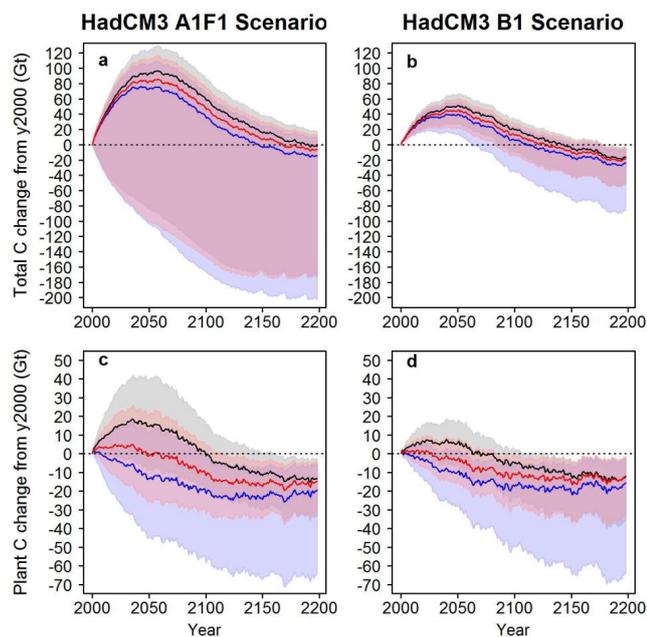


**Fig. 8.** Correlations between model predictions and hold-out evaluation data (left) and DIC values (right) obtained from different replace-null model fitting experiments. Confidence intervals and points have been removed where the empirical dataset predicted corresponds directly to the model replaced. This is because model performance for that component is typically very poor, making it difficult to visualise the performance of the other components. Points represent the average median estimate using the parameter probability distributions obtained from 10-fold model fitting, and the whiskers are average 95 % confidence intervals. “Control” indicates the results obtained for the full model fitted to all data sets.

the uncertainties identified, the model makes relatively well-constrained projections for some features of the carbon cycle, or it could be that the uncertainties tend to combine to give extremely uncertain dynamics. To illustrate this we set the plant and soil carbon pools across the terrestrial land surface to equilibrium in the year 2000 (as in Fig. 5 but at 0.5 decimal degree global resolution), and simulated climate change under pessimistic (A1F1) and optimistic (B1) anthropogenic emissions scenarios using our model to explore the plausible importance of these uncertainties (The IPCC Data Distri-

bution Centre; Lowe, 2005; [http://www.mad.zmaw.de/IPCC\\_DDC/html/SRES\\_AR4/index.html](http://www.mad.zmaw.de/IPCC_DDC/html/SRES_AR4/index.html); see Appendix D for details). We decided to set the land surface to equilibrium carbon levels to ensure that any changes in the carbon balance were solely due to climate changes, and the climate scenarios were chosen to create two extremes for comparison.

The resulting projected changes in the total terrestrial carbon balance are similar under both climate forcing scenarios, both predicting a net carbon sink up to about 2150 before becoming a carbon source (Fig. 9). However, the time



**Fig. 9.** Projected changes in terrestrial carbon under two climate change scenarios. These highlight the potential importance of parameter and structural uncertainty. Lines and shading represent the average median, 5th and 95th percentile projected changes (representing parameter uncertainty) from the terrestrial carbon model using parameter probability distributions inferred from 10 training data subsets. Red, grey and blue correspond to different climate dependences for plant mortality, representing the major structural uncertainty in the model (see Fig. 1). Details of how we simulated the model are given in Appendix C. The additional code necessary to run these simulations is available at [research.microsoft.com/en-us/downloads/](http://research.microsoft.com/en-us/downloads/), and the resulting simulation data are available from [download.microsoft.com/download/1/F/D/](http://download.microsoft.com/download/1/F/D/).

courses of uncertainty in the projections are quite different, with uncertainty higher under the A1F1 scenario and showing that the terrestrial carbon cycle could be either a carbon source or a sink over the simulated time period (note that these are not intended as reliable projections). Substituting the alternative models of plant mortality inferred under the different model fitting experiments indicates that this structural uncertainty has only minor quantitative effects on the projected change in the terrestrial carbon balance. Remarkably, however, this structural uncertainty leads to qualitatively different predictions for the projected changes in total vegetation carbon up to 2050 (Fig. 9), with positive, relatively flat, or negative changes in global vegetation carbon, depending on the mortality model parameterisation chosen. Despite these differences, all simulations predict that vegetation becomes a net source by 2020. Although these are exploratory simulations, they do emphasise the potential importance and value of considering parameter and structural uncertainties in terrestrial carbon models when attributing confidence to projections of the terrestrial carbon cycle in Earth

System Models. Such uncertainty could be decomposed further into the contributions from different component processes and even individual parameters and data. We could also begin such simulations with vegetation out of equilibrium. Such diagnoses are likely to help identify the different sources of uncertainty in predictions and projections, enabling the most important sources of uncertainty to be targeted for reduction.

Overall, our results complement the progress that has been made in data-constraining terrestrial carbon models (Rayner et al, 2011; Knorr and Heimann, 2001; Scholze et al., 2007; Ricciuto et al., 2011; Zafele et al., 2011) and the development of frameworks to enable such studies to be performed in a repeatable fashion (Scholze et al., 2007; <http://pecanproject.org/>). We hope that, by combining these insights, the biogeosciences community can rapidly move towards identifying the best models for specific purposes, that is, those models which have been shown to be most consistent with the available empirical evidence, and which incorporate uncertainties in model structure and parameter values into their predictions.

## Appendix A

### Processing of empirical data on carbon stocks and fluxes

We did not actively seek the “best data” pertaining to individual stocks and fluxes, because we prioritised enabling reproducibility and transparency over obtaining the most accurate results. A number of the data sets were only available as gridded land surface data, constructed through analyses of site-based data but for which the original site data were not available (Table 1). For these we generated pseudo-site based data by stratified random sampling. We anticipate that some of our data sets only coarsely represent true carbon stocks or flows at global scales. Moreover, although not all of the data in all data sets represent vegetation in a state of dynamical equilibrium, we screened the data sets that incorporate disturbance and anthropogenic effects using the Global Land Cover 2000 vegetation classification map (Bartholome and Belward, 2005) to only select data representative of vegetation at equilibrium.

### A1 Vegetation carbon

Global site-based data sets of carbon held in natural terrestrial vegetation have been compiled previously and used to produce global gridded maps. Unfortunately, the original source data have not been made generally available. Recently, Ruesch and Gibbs (2008) produced a global biomass carbon map for the year 2000 for the entire global land surface at 1 km resolution (approximately). This was derived from unique estimates of vegetation carbon values for 124 different carbon zones that were defined by considering the

continent, ecoregion, vegetation type and degree of human disturbance. We used this map to generate pseudo-site based data. We performed stratified random sampling of the Global Land Cover 2000 vegetation classification map (Bartholome and Belward, 2005; which has the same 1 km resolution as the vegetation carbon map) and then for each randomly selected point sampled the corresponding vegetation carbon value, conditional on whether the vegetation classification indicated specific vegetation types. Specifically, this included classification types 1–9 (generally “tree cover”), excluded type 10 (“tree Cover, burnt”), included types 11–15 (generally shrub or herbaceous cover), and excluded the remaining types 16–22 (not natural or not vegetation).

## A2 Litter carbon production

Different studies use terms like “litter production” to mean different aspects of vegetation biomass loss (Matthews, 2003). Here, we define “litter carbon production” as the net mass of carbon lost through natural leaf and fine root mortality (not including fire-induced mortality) per unit area of land surface ( $\text{m}^2$ ) per year (yr). This does not include woody litter production through the loss of branches, stems or coarse roots.

Leaf litter production is commonly estimated in the field using litter traps (Matthews, 2003). Studies employing this method rarely estimate root litter production. This led us to consider modelling leaf and fine root litter production separately. However, studies estimating leaf litter production sometimes use it to approximate net primary productivity by assuming that the gains and losses of carbon in the vegetation are balanced (Matthews, 2003). This implies that source data on leaf litter production might not be independent of some of our net primary production data (although we did not check to confirm this). These challenges led us to consider alternative methods for estimating litter carbon production.

Matthews (2003) used multiple methods to estimate litter production. One of these infers litter production from data on soil respiration rates and root respiration rates, by assuming that carbon stocks and flows are at equilibrium. Neither of these data sets was used in our model, so we used data generated from this method. A further complication however was that Matthews (2003) never made the source data available. Instead the author presented averages for 30 different vegetation types according to the UNESCO vegetation classification system (Matthews, 1999). We therefore generated pseudo-site based data. To do this we performed stratified random sampling of the UNESCO vegetation classification map (which comes in 1 degree resolution), and for each randomly selected point we associated the vegetation type with the litter production value given in column 3D of Table 5 in Matthews (2003). This gives litter dry matter production figures, so we multiplied by 0.5 to approximately convert from dry matter to carbon.

## A3 Soil carbon

Global site-based estimates of carbon held in soils under natural terrestrial vegetation have been compiled for decades. We initially considered using the NDP018 data set analysed by Post et al. (1982, 1985) in their studies of global patterns of plant carbon and nitrogen. However, preliminary investigations revealed significant differences in the Holdridge climate classifications (Holdridge, 1967) associated with the sites in the NDP018 data and the classifications we obtained by using georeferenced climate data from the New et al. (2002) data set. This led us to suspect that the GPS coordinates associated with the NDP018 are not sufficiently accurate to enable a sufficiently accurate estimate of the climatic conditions associated with the site. Instead we chose the “Global Gridded Surfaces of Selected Soil Characteristics” (IGBP-DIS) data set produced through the “Global Soil Data Task” project, which was designed specifically to assemble a “reliable and accessible data set on pedosphere properties on a global scale” (Batjes, 2000). We generated pseudo-site based data by performing stratified random sampling of the map of soil-carbon density at depth interval of 0–100 cm (which comes in 5 arcmin resolution).

## A4 Net primary productivity

Net primary productivity is estimated as the net mass (kg) of carbon fixed by living vegetation per unit area of land surface ( $\text{m}^2$ ) per year (yr). We selected NPP data compiled for the Ecosystem Model Data Intercomparison project (EMDI, Olson et al., 2001). We selected the Class B “intermediate quality” data set, because it provided a relatively high quantity of data (933 unique sites) and represented all of the major vegetation zones of the world.

## A5 Mortality rates of deciduous and evergreen leaves, and the fraction of vegetation that is evergreen

We know of no global data sets containing site-based estimates of leaf turnover rate at the whole vegetation stand level. Data from satellite observations are likely to fill this gap in the future. The lead authors of Wright et al. (2004) provided the GLOPNET database, which, according to Wright et al. (2004), “spans 2548 species from 219 families at 175 sites” (approximately 1 % of the extant vascular plant species). This database contains georeferenced data for a variety of species-specific leaf traits, for multiple species at a given site, including an estimate of leaf lifespan and whether the leaf is classified as deciduous or evergreen. In a recent analysis of global patterns of leaf mortality, using the GLOPNET database (Wright et al., 2004) supplemented with additional data, van Ommen Kloeke et al. (2011) found clear trends with environmental variables when the deciduous and evergreen leaves were treated separately. As a result we calculated the geometric average mortality rate at each

site, separately for deciduous and evergreen leaves. We also calculated the fraction of the records at each site that were of evergreen plants and used this as a coarse approximation of the fraction of vegetation at each site that is evergreen.

#### A6 Fine root mortality rate

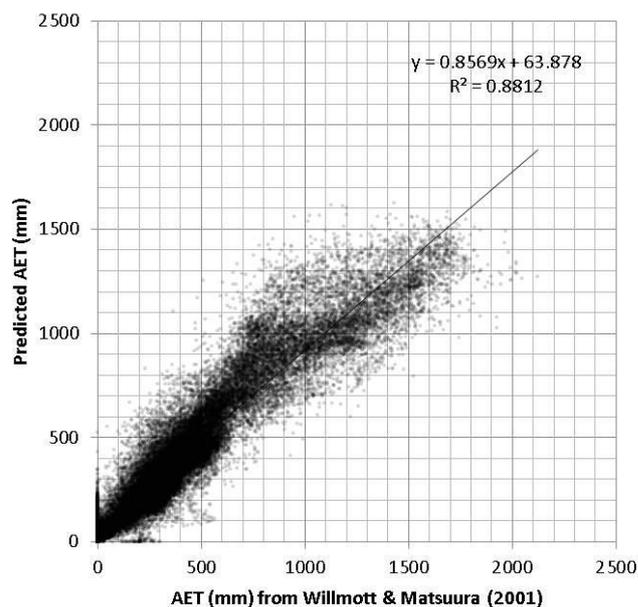
Data on root turnover rates are available for most vegetation types, even though estimating root mortality rates is notoriously difficult and prone to error. As is conventional, we distinguished between “coarse roots” and “fine roots”, but only explicitly modelled the mortality rate of fine roots (structural roots are implicitly represented in plant structural carbon). We define “fine roots” as those roots whose primary function is to acquire water and nutrients, with little role in structural support or resource storage (Eissenstat and Yanai, 1997). We obtained data from the Appendix of Gill and Jackson (2000) who studied global patterns in root turnover. We included all data from their table with the exception of any entries that had “root type” classified as “coarse”. We also corrected three obvious errors in their data set: “Adiopodoum, Ivory Coast” should have the longitude 4°30' W; “Portugal” should have longitude 9°24' W; and the two entries for “Macquarie Island, Subantarctic” were corrected to have longitude 158°57' E.

#### A7 Plant mortality rate

Any plant matter that is not “leaves” or “fine roots” in our model is classified as “structural”. We equated the turnover rate of structural plant parts in the absence of fire to the rate of plant mortality. We could find no data on plant mortality rates at global scales. Instead we used the data compiled by Stephenson and van Mantgem (2005) on forest turnover rates at global scales. This omits data on non-forested vegetation.

#### A8 Fractional area burned

A number of projects have sought to obtain accurate estimates of patterns of fire frequency at global scales. Mouillot and Field (2005) generated a global fire map of “fractional area burned per year” at 1° resolution for the terrestrial land surface by synthesizing available data and extrapolating when data were absent. We used this data set to generate pseudo-site based data on the fractional area burned per year by performing stratified random sampling of their map. We then took averages over the same time period as in the New et al. (2002) data set (1961–1990), including only samples from sites that were classified as natural vegetation (codes < 16) in the Global Land Cover 2000 map (Bartholome and Belward, 2005). In our predictive model we equate the fraction of land area burned per year with the rate of plant carbon losses due to fire, although we allow for this to be downscaled for structural plant parts.



**Fig. A1.** Comparison between our model-derived estimates of annual actual evapotranspiration (AET) and those predicted by the model of Matsuura and Willmott (2009). Predicted AET (mm) was calculated using the methodology described in Sect. 3.2.6 for land grid squares at 0.5 degree resolution. Points are partially transparent to help emphasise differences in data density.

#### A9 Fraction of carbon in leaves and fine roots in “metabolic” carbon pool

It is common in dynamic soil modelling to distinguish between the pool of dead plant carbon that decomposes relatively slowly (cellulose and lignin) and that which decomposes relatively rapidly (nucleic acids and cytoplasmic constituents; Ise and Moorcroft, 2006; Schimel et al., 1996; Bolker et al., 1998). However, we have not found any actual compilations of site-based estimates of the “metabolic fraction” to use to infer global patterns, nor of the two metrics typically used to calculate it: the ratio of carbon to nitrogen in plant tissues or the lignin fraction of plant tissue mass. What does exist are “representative” figures for different vegetation types: for example, the litter fall of boreal evergreen forest trees can be calculated to have a metabolic fraction of 0.49, whereas grassland litterfall has a metabolic fraction of 0.76, a typically high value. We therefore generated pseudo-site based data. To do this we performed stratified random sampling of the IBIS vegetation classification map (Ramankutty and Foley, 1999; <http://www.sage.wisc.edu/atlas/data.php?incdataset=PotentialVegetation>, which comes in 0.1 degree resolution), and for each randomly selected point we associated the vegetation type with the metabolic fraction value given in Table 1 of Ise and Moorcroft (2006).

### A10 Fraction of plants that is “structural”: everything but leaves and fine roots

The dominant vegetation types across the globe exhibit highly contrasting patterns in their ratios of leaves and fine roots to woody plant parts such as stems and woody roots. In this study we treated all of the biomass that is not leaves and fine roots as “structural”. Previous global vegetation modelling assumed different allocation patterns between leaves, roots and structural components for different plant functional types, such as boreal forests (high fraction “structural”) versus grasses (low fraction “structural”). In our study we inferred the most likely allocation patterns from data. However, we know of no global data explicitly documenting leaf : fine root : structural allocation ratios at the vegetation stand scale (although such data do exist for individual species). Instead, we assumed that the fraction of net primary productivity allocated to structural plant parts was zero in grasslands and some maximum in evergreen tropical rainforests. To acquire the data we performed stratified random sampling of the Global Land Cover 2000 map (Bartholome and Belward, 2005) and only recorded samples if they happened to be classified as these vegetation types. We recorded an integer “1” alongside the sample GPS coordinates if the vegetation was recorded as evergreen tropical rain forest (codes 1) or a “0” if the vegetation was recorded as grassland (codes 13 and 14).

## Appendix B

### Calculation of environmental variables

#### B1 Mean annual temperature

Mean annual temperature, used in several of the model components, was calculated as the arithmetic mean of monthly temperatures in the New et al. (2002) gridded climate data set. If elevation figures were present in georeferenced data, then the difference between this and the associated elevation figure in the New et al. (2002) data set was used to calculate a corrected temperature, assuming a lapse rate of 6.49 degrees Celsius per km elevation.

#### B2 Mean annual precipitation

Mean annual precipitation, used in several of the model components, was calculated as the sum of the mean monthly precipitation values in the New et al. (2002) gridded climate data set.

#### B3 Mean annual biotemperature

Mean annual biotemperature, used to calculate Holdridge life zones (Holdridge, 1967), was calculated as the arithmetic mean of monthly temperatures in the New et al. (2002) grid-

ded climate data set after setting monthly temperatures less than 0 °C or greater than 30 °C equal to zero.

#### B4 Fraction of the year experiencing frost

The fraction of the year experiencing frost (FYF), used to calculate the fraction of leaves that are deciduous or evergreen, was calculated using the method employed by van Ommen Kloeke et al. (2011) (P. van Bodegom, personal communication, 2011). The algorithm uses the number of frost days per month from the New et al. (2002) gridded climate data set. If the number of frost days was greater than 15, then the whole month was classified as being a “frost month”. If the number was less than 15, then it was classified as being a “non-frost month”. However, if a non-frost month followed a frost month, or if a non-frost month was to be followed by a frost month, then the fraction of the month in the non-frost month was calculated as 1/15 the number of frost days. The fraction of the year experiencing frost was then the sum of the number of frost months divided by 12.

#### B5 Mean annual potential evapotranspiration

Mean annual potential evapotranspiration rate was calculated using the Penman–Monteith algorithm as specified in Allen et al. (1998) for calculating monthly evapotranspiration rates. The algorithm is rather lengthy and we omit details here for brevity, although all details of the algorithm were checked using the test data provided in Allen et al. (1998). All environmental variables for the algorithm are available in the New et al. (2002) gridded climate data set. The sum of the monthly evapotranspiration rates gave PET. This calculation requires making an assumption about the vapour pressure deficit and stomatal resistance, both of which can be influenced by vegetation type. We simplified the calculation by calculating the PET according to grassland (the “reference evapotranspiration”; Allen et al., 1998). A natural area for future work will be to explore the importance of alternative formulations for calculating PET.

#### B6 Mean annual actual evapotranspiration

Soil moisture content was simulated on a daily time step to obtain estimates of the actual evapotranspiration rate and the length of the “fire season”. It was calculated using a modified version of the algorithm reported by Prentice et al. (1993). Climate variables for the algorithm came from the New et al. (2002) gridded climate data set, and soil maximum water capacity (field capacity) came from the Global Soil Data Task Group (2000) “Global Data Set of Derived Soil Properties” data set. Daily changes in soil water content were calculated using the balance equation specified in Prentice et al. (1993):

$$\omega_i = \min([\omega_{i-1} + P_i - E_i] \omega_{\max}) \quad (B1)$$

where  $\omega_i$  is soil water content (mm),  $\omega_{\max}$  is soil field capacity,  $i$  is time in days,  $P_i$  is daily precipitation and  $E_i$  is actual daily evapotranspiration. Actual daily losses due to evaporation are calculated as

$$E_i = ET_i \left( \frac{\omega_i - 1}{\omega_{\max}} \right) \quad (\text{B2})$$

where  $ET_i$  is daily potential evapotranspiration. Our method is a modification of the algorithm used by Prentice et al. (1993). Here the supply of water is taken to be proportional to maximum evaporative demand (potential evapotranspiration) scaled by the relative soil wetness. Prentice et al. (1993) calculated  $E_i$  as the minimum of daily supply and demand, where supply was calculated according to Eq. (B1) but using a maximum evapotranspiration rate constant instead of  $ET_i$ , and using a slightly different algorithm to calculate potential evapotranspiration rate (we used Allen et al., 1998). A natural extension in the future will be to use the methodology for model building that we report here to also assess the modelling of evapotranspiration rates.

For each site we initialised the soil water content at field capacity and simulated Eqs. (B1) and (B2) for 10 yr, which was long enough for the annual dynamics to converge to an equilibrium annual cycle. Like Prentice et al. (1993) we used a daily time step, because we found that adopting a coarser time step led to extreme numerical artefacts in the time series of soil water balance. Values for  $P_i$  and  $ET_i$  were obtained by linear interpolation of the monthly precipitation and potential evapotranspiration values, respectively. Monthly and annual actual evapotranspiration was then calculated by summation of the  $E_i$  values.

We checked that our actual evapotranspiration rate calculation yielded sensible predictions by comparing our estimates with model-derived estimates of global actual evapotranspiration rates (Willmott and Matsuura, 2001) averaged between 1961 and 1990, the same period as our New et al. (2002) climate data. We obtained a good agreement with their calculations (Fig. A1;  $r^2 = 0.88$ ,  $n = 44\,225$ ). However, a natural area for future work will be to explore the importance of alternative methods for calculating AET and soil water balance.

## B7 Length of the fire season

Daily soil water content predictions (detailed above) were used to estimate the length of the fire season – the fraction of days of the year over which fire is likely to occur. We based our algorithm on that specified in Thonicke et al. (2001), which calculates the length of the fire season as a function of the daily soil moisture status throughout the year. However, unlike Thonicke et al. (2001) we did not impose any constraints on the amount of biomass present for the daily probability of fire to be greater than zero (instead, this is part of the fire model), and we added the constraint that daily temperature (interpolated from monthly temperature) had to be

greater than zero for the daily probability of fire to be non-zero (as in Kloster et al., 2010). The algorithm was therefore

$$\text{LFS} = \sum_{i=1}^{360} \text{fp}(\omega_i, T_i), \text{ where} \quad (\text{B3a})$$

$$\text{fp} = \begin{cases} \exp\left(-\pi \left(\frac{\omega_i}{\omega_e}\right)\right), & T > 0, \omega_i < \omega_e \\ 0, & T \leq 0, \omega_i \geq \omega_e \end{cases} \quad (\text{B3b})$$

where  $\omega_i$  is the daily soil water content on day  $i$ ,  $\omega_e = 0.3$  is the soil moisture content at which fires become impossible (moisture of fire extinction) and  $T_i$  is the daily temperature, which was linearly interpolated from monthly values. Note that we assume a 360-day year. The parameter  $\omega_e = 0.3$ , used by Thonicke et al. (2001), is clearly one that could be inferred from data in future studies.

## Appendix C

### Model component functions

#### C1 Net primary productivity

We used the so-called “MIAMI” model, developed by Leith (1975), to predict annual net primary productivity (NPP) as a function of two of the main environmental variables known to limit plant productivity: temperature and available water. It has the following form:

$$\boxed{G} = \min(G_T, G_P) \text{ kg m}^{-2} \text{ yr}^{-1}, \quad (\text{C1a})$$

where  $\boxed{G}$  is NPP ( $\text{kg carbon m}^{-2} \text{ yr}^{-1}$ ). The functions  $G_T$  and  $G_P$  are calculated according to

$$G_T = \frac{G_{\max}}{1 + \exp(t_1 - t_2 \text{MAT})}, \quad (\text{C1b})$$

$$G_P = G_{\max} (1 - \exp(-p_1 \text{MAP})), \quad (\text{C1c})$$

where  $G_{\max}$ ,  $t_1$ ,  $t_2$ , and  $p_1$  are unknown constants (inferred parameters),  $\max$  is the maximum NPP,  $t_1$  and  $t_2$  scale the temperature dependency of NPP,  $p_1$  scales the precipitation dependency of NPP, MAT is mean annual temperature and MAP is mean annual precipitation.

#### C2 Leaf mortality rates and fraction of vegetation that is evergreen

We formulated a new leaf mortality model based on the recent analysis by van Ommen Kloeke et al. (2011) of global patterns of leaf lifespan. Van Ommen Kloeke et al. (2011) revealed contrasting global patterns of leaf lifespans (the inverse of leaf mortality rates) for evergreen and deciduous

plants. The mean leaf mortality rate  $\overline{\mu}_1$  ( $\text{yr}^{-1}$ ) is calculated as a weighted average according to

$$\overline{\mu}_1 = \exp(f_e \ln(\mu_e) + (1 - f_e) \ln(\mu_d)) \text{yr}^{-1}, \quad (\text{C2a})$$

where

$$f_e = a_{fe} \text{FYF}^2 + b_{fe} \text{FYF} + c_{fe}, \quad (\text{C2b})$$

$$\mu_e = \exp(m_e \text{MAT} - c_e), \quad (\text{C2c})$$

$$\mu_d = \exp(-(m_d \text{MAT} + c_d)). \quad (\text{C2d})$$

$a_{fe}$ ,  $b_{fe}$ ,  $c_{fe}$ ,  $c_e$ ,  $m_e$ ,  $c_d$ , and  $m_d$  are unknown constants (inferred parameters);  $0 < f_e < 1$  is the fraction of the vegetation that has evergreen leaves with parameters  $a_{fe}$ ,  $b_{fe}$  and  $c_{fe}$  scaling the quadratic Eq. (C2b);  $\mu_e$  is the mortality rate of evergreen leaves with parameters  $c_e$  and  $m_e$  scaling that exponential function;  $\mu_d$  is the mortality rate of deciduous leaves with parameters  $c_d$  and  $m_d$  scaling that function; and FYF is the fraction of the year that experiences frost (calculated using the method of van Ommen Kloeke et al., 2011).

### C3 Fine root mortality rate

Gill and Jackson (2000) analysed a database on root turnover rates from all major terrestrial vegetation types and found clear log-linear positive relationships between turnover rates and site mean annual temperatures for fine roots in forests, and for roots in shrublands and grasslands. Reflecting their findings we predicted fine root turnover rates according to

$$\overline{\mu}_r = \exp(m_{frm} \text{MAT} + c_{frm}) \text{yr}^{-1}, \quad (\text{C3})$$

where  $\overline{\mu}_r$  is root-turnover rate ( $\text{yr}^{-1}$ ), and  $c_{frm}$  and  $m_{frm}$  are unknown (assumed constant) parameters scaling the response of fine root mortality rate to MAT.

### C4 Plant mortality rate

Stephenson and van Mantgem (2005) analysed patterns of tree mortality rates across temperate and tropical forests worldwide and revealed a tendency for mortality rates to be higher in higher productivity areas. We therefore modelled plant mortality rates as

$$\overline{\mu}_s = \exp(p_2 \text{AET} + p_1) \text{yr}^{-1} \quad (\text{C4})$$

where  $\overline{\mu}_s$  is plant mortality rate ( $\text{yr}^{-1}$ ), and  $p_1$  and  $p_2$  are unknown constants (inferred parameters) that scale the plant mortality rate as a function of annual actual evapotranspiration, AET.

### C5 Mortality rate due to fire

We developed a fire model based on the models of Thonicke et al. (2001), Kloster et al. (2010) and Arora and Boer (2005). We predicted the per capita vegetation mortality rate due to fire as

$$\overline{\mu}_f = c_f f_1(\text{LFS}) f_2(\overline{G}) \text{yr}^{-1}, \quad \text{where} \quad (\text{C5a})$$

$$f_1(\text{LFS}) = \frac{1}{1 + \exp(-\text{lfs}_{\text{scalar}}(\text{LFS} - \text{lfs}_{\text{halfsat}}))} \quad \text{and} \quad (\text{C5b})$$

$$f_2(\overline{G}) = \frac{1}{1 + \exp(-\text{NPP}_{\text{scalar}}(\overline{G} - \text{NPP}_{\text{halfsat}}))}. \quad (\text{C5c})$$

Here,  $\overline{\mu}_f$  is the mortality rate due to fire ( $\text{yr}^{-1}$ ), and  $c_f$ ,  $\text{lfs}_{\text{scalar}}$ ,  $\text{lfs}_{\text{halfsat}}$ ,  $\text{NPP}_{\text{scalar}}$ , and  $\text{NPP}_{\text{halfsat}}$  are unknown constants (inferred parameters). The constant  $c_f$  scales the overall mortality rate due to fire,  $\text{lfs}_{\text{scalar}}$  and  $\text{lfs}_{\text{halfsat}}$  scale the logistic response of this mortality rate to the length of the fire season LFS, and  $\text{NPP}_{\text{scalar}}$  and  $\text{NPP}_{\text{halfsat}}$  scale the logistic response of fire return interval to  $\overline{G}$ . To infer the parameters to this model, we assume that the mortality rate due to fire is equivalent to the fractional area burned per year.

### C6 Metabolic fraction

The fraction of leaf and fine root carbon allocated to components that decompose relatively rapidly (nucleic acids and cytoplasmic constituents) notably varies between different plant functional types, with gymnosperms, for example, tending to have a relatively low “metabolic fraction”. Metabolic fraction also tends to be positively associated with environmental variables such as actual evapotranspiration rate, even when controlling for changes in plant functional types (Aerts, 1997). Rather than introduce plant functional types, we chose to model metabolic fraction as a simple linear function:

$$\overline{f}_m = c_{fm} + m_{fm} \text{AET}, \quad (\text{C6})$$

where  $\overline{f}_m$  is metabolic fraction and  $c_{fm}$  and  $m_{fm}$  are unknown constants (inferred parameters) that scale the response of metabolic fraction to AET.

### C7 Fraction of carbon allocated to structural components

We developed a simple model that predicts the fraction of carbon to woody plant parts as a logistic function of the net primary productivity of the vegetation. This model has the following form:

$$\overline{f}_s = \min \left( \frac{0.01 \exp \left( f_{\text{scalar}} \overline{G} \right)}{1 + 0.01 \exp \left( f_{\text{scalar}} \overline{G} - 1 \right)}, 0.99 \right), \quad (\text{C7})$$

where  $f_s$  scales the fractional carbon allocation to woody plant parts and  $f_{\text{scalar}}$  is an unknown constant (inferred parameter), scaling the response of  $f_s$  to  $G$ . In this form,  $0.01 < f_s < 0.99$ , with the maxima and minima set to allow maximum likelihood estimation of the parameters to the continuous function  $f_s$  given the binary data. However, this gets rescaled by the maximum fractional carbon allocation parameter  $f_{\text{max}}$  in the full vegetation model (Eq. 1).

**C8 Relative soil decomposition rate**

A commonly used approach to modelling the decomposition of carbon in different soil carbon pools is to assume that they decompose with different maximum rates, but with the same dependence on environmental conditions (Ise and Moorcroft, 2008; Schimel et al., 1996; Bolker et al., 1998; Adair et al., 2008). This is normally implemented by scaling the maximum decomposition rates for each pool with a common environmentally dependent decomposition index. We developed a decomposition index that is highly analogous to those that have been adopted in other studies, and takes the form

$$A = A_{\text{MAT}} A_{\text{P}}, \quad \text{where}$$

$$A_{\text{MAT}} = \begin{cases} \left( \frac{0.99 \left( \frac{1}{1 + t_{\text{d}_a} \exp(t_{\text{d}_b} \text{MAT})} \right) + 0.01}{\exp(t_{\text{S}_c} (\text{MAT} - t_{\text{thres}}))} \right) / A_{\text{P, max}}, & \text{MAT} < t_{\text{thres}} \\ \exp(t_{\text{S}_c} (\text{MAT} - t_{\text{thres}})), & \text{MAT} \geq t_{\text{thres}} \end{cases}, \quad \text{where} \quad (\text{C8a})$$

$$A_{\text{MAT, max}} = 0.99 \left( \frac{1}{1 + t_{\text{d}_a} \exp(t_{\text{d}_b} t_{\text{thres}})} \right) + 0.01, \quad \text{and} \quad (\text{C8b})$$

$$A_{\text{P}} = \begin{cases} \left( \frac{0.99 \left( \frac{1}{1 + m_{\text{d}_a} \exp\left(\frac{m_{\text{d}_b} \text{MAP}}{\text{PET}}\right)} \right) + 0.01}{\exp(m_{\text{S}_c} \left(\frac{\text{MAP}}{\text{PET}} - m_{\text{thres}}\right))} \right) / A_{\text{P, max}}, & \frac{\text{MAP}}{\text{PET}} < m_{\text{thres}} \\ \exp(m_{\text{S}_c} \left(\frac{\text{MAP}}{\text{PET}} - m_{\text{thres}}\right)), & \frac{\text{MAP}}{\text{PET}} \geq m_{\text{thres}} \end{cases}, \quad \text{where} \quad (\text{C8c})$$

$$A_{\text{P, max}} = 0.99 \left( \frac{1}{1 + m_{\text{d}_a} \exp(m_{\text{d}_b} m_{\text{thres}})} \right) + 0.01. \quad (\text{C8d})$$

Here,  $A$  is the relative soil decomposition rate, and  $t_{\text{d}_a}$ ,  $t_{\text{d}_b}$ ,  $t_{\text{S}_c}$ ,  $t_{\text{thres}}$ ,  $m_{\text{d}_a}$ ,  $m_{\text{d}_b}$ ,  $m_{\text{S}_c}$ , and  $m_{\text{thres}}$  are unknown constants (inferred parameters) and PET is mean annual potential evapotranspiration. We assume a minimum relative decomposition rate returned by either “wetness” or temperature functions of 0.01.

**Appendix D**

**Methodology for projections under climate change**

The full terrestrial carbon model is simulated by solving Eq. (1) using a standard forward Euler method with time step  $\Delta t = 1/12$  yr. Preliminary investigations were made to ensure that this time step was sufficiently small to avoid

extreme numerical artefacts in the simulated time series. Changes in soil water content were solved at a higher temporal resolution ( $\Delta t = 1/360$ , or 30 times per update in Eq. 1) in between updates of the carbon stocks to maintain consistency with the method used in model parameter inference and to avoid extreme numerical artefacts in the dynamics of soil water balance.

The model (Eq. 1) was simulated across the terrestrial land surface at 0.5 degree resolution. Initial values were calculated by solving Eq. (C1) to obtain equilibrium carbon stocks. Environmental variables were obtained using the same method as for model parameter inference – using the New et al. (2002) gridded climate data and the Global Soil Data Task Group (2000) “Global Data Set of Derived Soil Properties” data set for soil field capacity.

We obtained two data sets arising from simulations of the HadCM3 model run under different anthropogenic emissions scenarios from the IPCC Data Distribution Centre ([http://www.mad.zmaw.de/IPCC\\_DDC/html/SRES\\_AR4/index.html](http://www.mad.zmaw.de/IPCC_DDC/html/SRES_AR4/index.html); Lowe, 2005). These detail predicted monthly values of environmental variables for the surface of the Earth gridded at a  $2.5 \times 3.75$  degree resolution from the year 2000 through to 2199. We did not use these environmental data directly to drive the model but instead applied the difference between a given year and that at 2000 to the values in the New et al. (2002) data set.

The future climate data sets did not contain all of the environmental variables we had used for model parameter inference. In particular, they did not include several of the variables needed for calculating potential evapotranspiration (PET) using the Penman–Monteith algorithm. We therefore resorted to using the simpler Malmstrom algorithm for calculating changes in PET for the global simulations, which only uses data on monthly temperatures. We also used a different algorithm for calculating the number of days of the year in which frost is present prior to calculating yearly changes in the fraction of the year that is frost (FYF). This was calculated by associating predicted monthly temperatures with the middle day of each month, then linearly interpolating between these to calculate the fraction of the year over which temperatures are less than zero.

Climate change simulations were run for all 1200 sampled joint posterior parameter estimates resulting from parameter inference of the full dynamic global vegetation model (DGVM), for the parameter sets returned from each of the 10 different subsets of training data (thus, 12 000 simulations). To run simulations with different plant mortality models we simply replaced the list of sampled mortality model parameter values with those obtained from either inferring the parameters for the mortality model alone to the data on plant mortality rates, or from inferring the parameters for the full DGVM in the absence of the data on plant mortality rates.

To obtain estimates of changes in the stocks of terrestrial plant and soil carbon we multiplied the estimated values

for each grid point (in units of  $\text{kg C m}^{-2}$ ) by the area of the grid cell.

**Acknowledgements.** We thank the Computational Science Laboratory and the Machine Learning and Perception group at Microsoft Research Cambridge for technical assistance and comments on the manuscript. We thank Colin Prentice, Bjorn Brooks, Eva Falge and an anonymous reviewer for comments that greatly improved this manuscript. We thank 3 anonymous reviewers to a previous version of the manuscript for detailed constructive comments. We thank Florent Mouillot, Ian Wright, Peter Reich, H. Gibbs, Elaine Matthews, Rob Jackson, Nathan Stephenson, Philip van Mantgem, Takeshi Ise, the Joint Research Centre of the European Commission, Springer, John Wiley & Sons, the American Geophysical Union, Oak Ridge National Laboratory Distributed Active Archive Center (ORNL-DAAC) and the Climate Research Unit and the University of East Anglia, for their kind permission in allowing us to make derivatives of their data available with our publication, and the same authors as well as Mac Post, Robert Hijmans, the Center for Sustainability and the Global Environment (SAGE) and The IPCC Data Distribution Centre for making their data available for us to use in our research. We thank G. Mace, A. Friend, P. van Bodegom, Neil Crout, Pete Smith, Jessica Bellarby, Tim Lenton, Ben Adams, Kirsten Thonicke, Stephen Beckett and the GREENCYLCES2 members for providing advice for our study. We thank Pablo Tapia, Carole Boelitz, Chuck Needham, Rachel Free and Rob Knies for technical assistance.

Edited by: E. Falge

## References

- Adair, E. C., Parton, W. J., del Grosso, S. J., Silver, W. L., Harmon, M. E., Hall, S. A., Burke, I. C., and Hart, S. C.: Simple three-pool model accurately describes patterns of long term litter decomposition in diverse climates, *Global Change Biol.*, 14, 2636–2660, 2008.
- Adams, B., White, A., and Lenton, T. M.: An analysis of some diverse approaches to modelling terrestrial net primary productivity, *Ecol. Mod.*, 177, 351–391, 2004.
- Aerts, R.: Climate, leaf litter chemistry and leaf litter decomposition in terrestrial ecosystems: a triangular relationship, *Oikos*, 79, 439–449, 1997.
- Allen R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration – Guidelines for computing crop water requirements, *FAO Irrigation and drainage paper*, 56, <http://www.fao.org/docrep/X0490E/X0490E00.htm>, 1998.
- Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J. H., Allard, G., Running, S. W., Semerci, A., and Cobb N.: A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *For. Ecol. Manage.*, 259, 660–684, 2010.
- Arneth, A., Mercado, L., Kattge, J., and Booth, B. B. B.: Future challenges of representing land-processes in studies on land-atmosphere interactions, *Biogeosciences*, 9, 3587–3599, doi:10.5194/bg-9-3587-2012, 2012.
- Arora, V. K. and Boer, G. J.: Fire as an interactive component of dynamic vegetation models, *J. Geophys. Res.*, 110, G02008, doi:10.1029/2005JG000042, 2005.
- Bartholome, E. M. and Belward, A. S.: ; a new approach to global land cover mapping from Earth Observation Data, *Int. J. Remote Sens.*, 26, GLC2000, doi:10.1080/01431160412331291297, 1959–1977, 2005.
- Batjes, N. H.: Global Data Set of Derived Soil Properties, 0.5-Degree Grid (ISRIC-WISE), Data set, from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, available at: <http://www.daac.ornl.gov>, 2000.
- Bolker, B. M., Pacala, S. W., and Parton Jr., W. J.: Linear analysis of soil decomposition: insights from the century model, *Ecol. App.*, 8, 425–439, 1998.
- Bonan, G.: Physiological derivation of the observed relationship between net primary productivity and mean annual air temperature, *Tellus B*, 45, 379–408, doi:10.1034/j.1600-0889.1993.t014-00001.x, 1993.
- Cox, P. and Stephenson, D.: A changing climate for prediction, *Science*, 317, 207–208, 2007.
- Cramer, W., Kicklighter, D. W., Bondeau, A., Moore III, B., Churkina, G., Nemry, B., Ruimy, A., and Schloss, A. L.: Global responses of terrestrial ecosystem structure and function to CO<sub>2</sub> and climate change: results from six dynamic global vegetation models, *Global Change Biol.*, 7, 357–373, 2001.
- Dai, A., and Fung, I. Y.: Can climate variability contribute to the “missing” CO<sub>2</sub> sink?, *Global Biogeochem. Cy.*, 7, 599–609, 1993.
- Denman, K. L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P. L., Wofsy, S. C., and Zhang, X.: Couplings Between Changes in the Climate System and Biogeochemistry, in: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Eissenstat, D. M. and Yanai, R. D.: The ecology of root lifespan, *Ads. Ecol. Res.*, 17, 1–60, 1997.
- Friedlingstein, P., Delire, C., Muller, J. F., and Gerard, J. C.: The climate induced variation of the continental biosphere: a model simulation of the Last Glacial Maximum, *Geophys. Res. Lett.*, 19, 897–900, 1992.
- Friedlingstein, P., Dufresne, J.-L., Cox, P. M., and Rayner P.: How positive is the feedback between climate change and the carbon cycle?, *Tellus B*, 55, 692–700, 2003.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison, *J. Climate*, 19, 3337–3353, 2006.

- Gelfand, A. E. and Day, D. K.: Bayesian model choice: asymptotics and exact calculations, *J. Roy. Statist. Soc. B*, 56, 501–514, 1994.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: *Bayesian Data Analysis*, Chapman and Hall, Boca Raton, USA, 2004.
- Giglio, L., Randerson, J. T., van der Werf, G. R., Kasibhatla, P. S., Collatz, G. J., Morton, D. C., and DeFries, R. S.: Assessing variability and long-term trends in burned area by merging multiple satellite fire products, *Biogeosciences*, 7, 1171–1186, doi:10.5194/bg-7-1171-2010, 2010.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J.: *Markov Chain Monte Carlo in Practice*, Chapman and Hall, Boca Raton, USA, 1996.
- Gill, R. and Jackson, R. B.: Global Patterns of root turnover for terrestrial ecosystems, *New Phytol.*, 81, 275–280, 2000.
- Global Soil Data Task Group: Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS, [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme – Data and Information System)], from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, available at: <http://www.daac.ornl.gov>, 2000.
- Goll, D. S., Brovkin, V., Parida, B. R., Reick, C. H., Kattge, J., Reich, P. B., van Bodegom, P. M., and Niinemets, Ü.: Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling, *Biogeosciences Discuss.*, 9, 3173–3232, doi:10.5194/bgd-9-3173-2012, 2012.
- Holdridge, L. R.: *Life Zone Ecology*, Tropical Science Centre, San Jose, Costa Rica, 1967.
- Ise, T. and Moorcroft, P. R.: The global-scale temperature and moisture dependencies of soil organic carbon decomposition: an analysis using a mechanistic decomposition model, *Biogeochem.*, 80, 217–231, 2006.
- Kaminski, T., Knorr, W., Rayner, P. J., and Heimann, M.: Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle, *Glob. Biogeochem. Cy.*, 16, 1–8, 2002.
- Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P.-P.: Consistent assimilation of MERIS FAPAR and atmospheric CO<sub>2</sub> into a terrestrial vegetation model and interactive mission benefit analysis, *Biogeosciences*, 9, 3173–3184, doi:10.5194/bg-9-3173-2012, 2012.
- Keenan, T. F., Baker, I., Barr, A., Ciais, P., Davis, K., Dietze, M., Dragoni, D., Gough, C. M., Grant, R., Hollinger, D., Hufkens, K., Poulter, B., McCaughley, H., Raczka, B., Ryu, Y., Schaefer, K., Tian, H., Verbeek, H., Zhao, M., and Richardson, A. D.: Terrestrial biosphere model performance for inter-annual variability of land-atmosphere CO<sub>2</sub> exchange, *Glob. Change Biol.*, 18, 1971–1987, doi:10.1111/j.1365-2486.2012.02684.x, 2012.
- Kerr, R. A.: Time to Adapt to a Warming World, But Where's the Science?, *Science*, 334, 1052–1053, 2011.
- Kicklighter, D. W., Schloss, A. L., Kaduk, J., McGuire, A. D., and Bondeau, A.: Comparing global models of terrestrial net primary productivity (NPP): global pattern and differentiation by major biomes, *Global Change Biol.*, 5, Supplement 1, 16–24, 1999.
- Kloster, S., Mahowald, N. M., Randerson, J. T., Thornton, P. E., Hoffman, F. M., Levis, S., Lawrence, P. J., Feddes, J. J., Oleson, K. W., and Lawrence, D. M.: Fire dynamics during the 20th century simulated by the Community Land Model, *Biogeosciences*, 7, 1877–1902, doi:10.5194/bg-7-1877-2010, 2010.
- Knorr, W. and Heimann, M.: Uncertainties in global terrestrial biosphere modelling, Part I: A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme, *Global Biogeochem. Cy.*, 15, 207–225, 2001.
- Leith, H.: *Modelling the primary production of the world*, edited by: Leith, H. and Whittaker, R. H., Springer-Verlag, New York, USA, 238–263, 1975.
- Lowe: IPCC DDC AR4 UKMO-HadCM3 SRESB1 run1. World Data Center for Climate. CERA-DB, available at: [http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=UKMO\\_HadCM3\\_SRESB1\\_1](http://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=UKMO_HadCM3_SRESB1_1), 2005.
- Malmstrom, V. H.: A new approach to the classification of climate, *J. Geog.*, 68, 351–357, 1969.
- Maslin, M. and Austin, P.: Climate models at their limit?, *Nature*, 486, 183–184, 2012.
- Matsuura, K. and Willmott, C. J.: *Terrestrial Water Budget Data Archive: Monthly Time Series (1900–2008)*, available at: [http://climate.geog.udel.edu/~climate/html\\_pages/WB2009/README.wb2009\\_ts.html](http://climate.geog.udel.edu/~climate/html_pages/WB2009/README.wb2009_ts.html), 2009
- Matthews, E.: Global litter production, pools and turnover times: Estimates from measurement data and regression models, *J. Geophys. Res.*, 102, 18771–18800, 2003.
- Matthews, E.: *Global Vegetation Types, 1971–1982* (Matthews), from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, available at: <http://daac.ornl.gov>, 1999.
- Mouillot, F. and Field, C. B.: Fire history and the global carbon budget, *Global Change Biol.*, 11, 398–420, 2005.
- New, M., Lister, D., Hulme, M., and Makin, I.: A high-resolution data set of surface climate over global land areas, *Clim. Res.*, 21, 1–25, 2002.
- Olson, R. J., Scurlock, J. M. O., Prince, S. D., Zheng, D. L., and Johnson, K. R.: NPP Multi-Biome: NPP and Driver Data for Ecosystem Model-Data Intercomparison, from the Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA, available at: <http://www.daac.ornl.gov>, 2001.
- Post, W. M., Emanuel, W. E., Zinke, P. J., and Stangenberger, A. G.: Soil carbon pools and world life zones, *Nature*, 298, 156–159, 1982.
- Post, W. M., Pastor, J., Zinke, P. J., and Stangenberger, A. G.: Global patterns of soil nitrogen storage, *Nature*, 317, 613–616, 1985.
- Prentice, I. C., Sykes, M. T., and Cramer, W.: A simulation model for the transient effects of climate change on forest landscapes, *Ecol. Mod.*, 65, 51–70, 1993.
- Purves, D. W. and Pacala, S.: Predictive models of forest dynamics, *Science*, 320, 1452–1453, 2008.
- Ramankutty, N. and Foley, J. A.: Estimating historical changes in land cover: North American croplands from 1850 to 1992, *Global Ecol. Biogeog.*, 8, 381–396, 1999.
- Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R. J., Sumi, A., and Taylor, K. E.: *Climate Models and Their Evaluation*, in: *Climate Change 2007: The Physical Science Basis*, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M.

- Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- Randerson, J. T., Hoffman, M. F., Thornton, P. E., Mahowald, N. M., Lindsay, K., Lee, Y.-H., Nevison, C. D., Foney, S. C., Bonan, G., Stockli, R., Covey, C., Running, S. W., and Fung, I. Y.: Systematic assessment of terrestrial biogeochemistry in coupled climate – carbon models, *Global Change Biol.*, 15, 2462–2484, 2009.
- Rayner, P. J., Koffi, E., Scholze, M., Kaminski, T., and Dufresne, J.-L.: Constraining predictions of the carbon cycle using data, *Phil. Trans. R. Soc. A*, 369, 1955–1955, 2011.
- Ricciuto, D. M., King, A. W., Dragoni, D., and Post, W. M.: Parameter and prediction uncertainty in an optimized terrestrial carbon cycle model: Effects of constraining variables and data record length, *J. Geophys. Res.*, 136, G01033, doi:10.1029/2010JG001400, 2011.
- Ruesch, A. and Gibbs, H. K.: New IPCC Tier-1 Global Biomass Carbon Map For the Year 2000, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, available at: <http://cdiac.ornl.gov>, 2008.
- Schimel, D. S., Braswell, B. H., McKeown, R., Ojima, D. S., Parton, W. J., and Pulliam, W.: Climate and nitrogen controls on the geography and timescales of terrestrial biogeochemical cycling, *Global Biogeochem. Cy.*, 10, 677–692, 1996.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *J. Geophys. Res.*, 112, D17305, doi:10.1029/2007JD008642, 2007.
- Sitch, S., Huntingford, C., Gedney, N., Levy, P. E., Lomas, M., Piao, S. L., Betts, R., Ciais, P., Cox, P., Friedlingstein, P., Jones, C. D., Prentice, I. C., and Woodward, F. I.: Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five dynamic global vegetation models (DGVMs), *Global Change Biol.*, 14, 1–25, 2008.
- Stephenson, N. L. and van Mantgem, P. J.: Forest turnover rates follow global and regional patterns of productivity, *Ecol. Lett.*, 8, 524–531, 2005.
- Thonicke, K., Venevsky, S., Sitch, S., and Cramer, W.: The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model, *Global Ecol. Biogeog.*, 10, 661–677, 2001.
- van Mantgem, Stephenson, N. L., Byrne, J. C., Daniels, L. D., Franklin, J. F., Fule, P. Z., Harmon, M. E., Larson, A. J., Smith, J. M., Taylor, A. H., and Veblen, T. T.: Widespread increase of tree mortality rates in the Western United States, *Science*, 323, 521–524, 2009.
- van Oijen, M., Cameron, D. R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., and Yreluripati, J. B.: A Bayesian framework for model calibration, comparison and analysis: Application to four models for the biogeochemistry of a Norway spruce forest., *Ag. Forest Met.*, 151, 1609–1621, 2011.
- van Ommen Kloeke, A. E. E., Douma, J. C., Ordonez, J. C., Reich, P. B., and van Bodegom, P. M.: Global quantification of contrasting leaf life span strategies for deciduous and evergreen species in response to environmental conditions, *Global Ecol. Biogeog.*, 21, 224–235, 2011.
- Wright, I., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M. L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., and Villar, R.: The worldwide leaf economics spectrum, *Nature*, 428, 821–827, 2004.
- Zahele, S., Friedlingstein, P., and Friend, A. D.: Terrestrial nitrogen feedbacks may accelerate future climate change, *Geophys. Res. Lett.*, 37, L01401, doi:10.1029/2009GL041345, 2010.
- Zahele, S., Ciais, P., Friend, A. D., and Prieur, V.: Carbon benefits of anthropogenic reactive nitrogen offset by nitrous oxide emissions, *Nature Geosci.*, 4, 601–605, 2011.
- Zhou, T. and Luo, Y.: Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States, *Global Biogeochem. Cy.*, 22, GB3032, doi:10.1029/2007GB002939, 2008.