Using models to guide field experiments: *a priori* predictions for the CO_2 response of a nutrient- and water-limited native Eucalypt woodland

Belinda E. Medlyn[1,2,*], Martin G. De Kauwe[2], Sönke Zaehle[3], Anthony P. Walker[4], Remko A. Duursma[1], Kristina Luus[3], Mikhail Mishurov[5], Bernard Pak[6], Benjamin Smith[5], Ying-Ping Wang[6], Xiaojuan Yang[4], Kristine Y. Crous[1], John E. Drake[1], Teresa E. Gimeno[1,7], Catriona A. Macdonald[1], Richard J. Norby[4], Sally A. Power[1], Mark G. Tjoelker[1], David S. Ellsworth[1]

1. Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith NSW 2751, Australia
2. Department of Biological Sciences, Macquarie University, North Ryde NSW 2109, Australia
3. Biogeochemical Integration Department, Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, D-07745 Jena, Germany
4. Oak Ridge National Laboratory, Environmental Sciences Division and Climate Change Science Institute, 1 Bethel Valley Road, Oak Ridge, TN, USA
5. Department of Physical Geography and Ecosystem Science, Lund University, Sölvegatan 12, 22362 Lund, Sweden
6. CSIRO Oceans and Atmosphere Flagship, Private Bag 1, Aspendale, Victoria 3195, Australia.
7. INRA UMR 1391 ISPA, 33140 Villenave d'Ornon Cedex, France

* Corresponding author.
Current address: Hawkesbury Institute for the Environment, Western Sydney University, Locked Bag 1797, Penrith NSW 2751, Australia
Phone: +612 4570 1372
Fax: +612 4570 1103
Email: b.medlyn@uws.edu.au

For submission to: Global Change Biology
Running head: Model predictions for the EucFACE experiment

Keywords: Carbon dioxide, ecosystem model, phosphorus, drought, *Eucalyptus tereticornis*

Type of Paper: Primary Research Article

No of Words: 7659

No of Tables: 2

No of Figures: 8

ABSTRACT

The response of terrestrial ecosystems to rising atmospheric CO_2 concentration (C_a), particularly under nutrient limited conditions, is a major uncertainty in Earth System models. The *Eucalyptus* Free-Air CO_2 Enrichment (EucFACE) experiment, recently established in a nutrient-and water-limited woodland, presents a unique opportunity to address this uncertainty, but can best do so if key model uncertainties have been identified in advance. We applied seven vegetation models, which have previously been comprehensively assessed against earlier forest FACE experiments, to simulate *a priori* possible outcomes from EucFACE. Our goals were to provide quantitative projections against which to evaluate data as they are collected, and to identify key measurements that should be made in the experiment to allow discrimination among alternative model assumptions in a post-experiment model intercomparison. Simulated responses of annual net primary productivity (NPP) to elevated C_a ranged from 0.5 to 25% across models. The simulated reduction of NPP during a low rainfall year also varied widely, from 24% to 70%. Key processes where assumptions caused disagreement among models included nutrient limitations to growth; feedbacks to nutrient uptake; autotrophic respiration; and the impact of low soil moisture availability on plant processes. Knowledge of the causes of variation among models is now guiding data collection in the experiment, with the expectation that the experimental data can optimally inform future model improvements.

INTRODUCTION

Ecosystem-scale manipulation experiments provide a unique opportunity to constrain the process-based vegetation models used to predict future productivity, carbon sequestration and land surface properties (Piao *et al.,* 2013; Dukes *et al.,* 2014; Medlyn *et al.,* 2015; Norby *et al.,* 2015). For example, one of the major uncertainties in projecting future atmospheric CO_2 concentration (C_a) is the uptake of C by terrestrial vegetation (Arora *et al.,* 2013; Friedlingstein *et al.,* 2014). Ecosystem-scale Free-Air CO_2 Enrichment (FACE) experiments, in which vegetation is exposed to elevated C_a (eC_a) continuously for a decade or more, are an essential source of data to both inform and test modelled responses. Previous FACE experiments have been especially valuable because they provided a comprehensive set of measurements of the major carbon, water, and nutrient fluxes, and were of long enough duration to capture the influences of year-to-year variation in weather, adjustment of transient responses, and shifts in vegetation structure (Norby & Zak, 2011). Such experiments can inform our ability to model vegetation responses to nutrient limitation and water stress as well as rising C_a.

The FACE Model-Data Synthesis project (Walker *et al.,* 2014, Medlyn *et al.,* 2015) tested 11 vegetation models against data from two forest FACE experiments and successfully identified a number of ways in which model assumptions could be improved based on these data, including recommendations for advancing the modelling of stomatal conductance (De Kauwe *et al.,* 2013), predicting nutrient limitations of growth (Zaehle *et al.,* 2014), and capturing flexibility of carbon allocation patterns (De Kauwe *et al.,* 2014). However, that model-data synthesis project took place after the experiments had been completed, too late to adapt the process of data collection to better inform the model simulations and reduce the identified uncertainties. Based on these and similar experiences, there is an increasing awareness that the knowledge advances from long-term ecosystem experiments can be maximised if models are employed from the outset (Luo 2001; Parton *et al.,* 2007; Norby *et al.,* 2015).

Model simulations at the commencement of an experiment serve a number of purposes. They can highlight key areas of model uncertainty that could be addressed

experimentally (e.g. McMurtrie & Comins 1996), and can identify powerful sets of measurements to discriminate amongst competing hypotheses (e.g. Dietze *et al.,* 2014). They can also provide a benchmark against which to compare experimental results, making it possible to identify outcomes that are consistent or inconsistent with existing theory (Luo *et al.,* 2011). For example, Parton *et al.* (2007) used the DAYCENT model to predict ecosystem impacts of eC_a and warming in a semi-arid grassland in advance of the Prairie Heating and CO_2 Enrichment (PHACE) experiment. The model predictions of the response to nitrogen mineralization to increased soil temperature and reduced soil moisture helped to structure the measurements and reporting of results, which were shown to support the model predictions (Dijkstra *et al.,* 2010). Furthermore, application of models throughout the course of the experiment stimulates data collection, verification, and archival in a model-friendly format, which can be extremely difficult to achieve retrospectively (Norby *et al.,* 2015).

In this paper, we applied representative vegetation models to the recently-established EucFACE experiment. EucFACE is a FACE experiment in a mature, natural *Eucalyptus* woodland in western Sydney, growing in nutrient- and water-limited conditions. Soil at the experimental site is strongly depleted in organic matter and nutrients, particularly extractable phosphorus, and leaf nitrogen:phosphorus (N:P) ratios are high (see Supporting Table S1). At the start of the experiment, it was thought likely that P availability was limiting to tree growth at the site, a hypothesis that has since been supported experimentally (Crous *et al.* 2015). The site also has low water availability; mean annual potential evapotranspiration is 1300 mm while mean annual rainfall is 800 mm, with a 10th percentile of 530 mm. A nearby experiment with young *Eucalyptus* trees has found strong responses to both irrigation and fertilisation (Craig Barton and Burhan Amiji, unpublished data).

Previous experiments have demonstrated that water- and N-cycle interactions are key determinants of the plant growth response to eC_a (e.g. Dukes *et al.,* 2005, Reich *et al.,* 2014) and that divergent patterns of plant carbon (C) allocation can affect long-term growth under eC_a via feedbacks related to nutrient acquisition (e.g., contrasting results of McCarthy *et al.,* 2010; Norby *et al.,* 2010). Possible interactions between P availability and eC_a have received relatively little attention despite widespread P

limitation across much of the globe (McGroddy *et al.,* 2004; Wang *et al.,* 2010; Cleveland *et al.,* 2011). Several glasshouse-based experiments suggest a constrained growth response to eC_a under low soil P (Conroy *et al.,* 1990, Edwards *et al.,* 2005, Lewis *et al.,* 2010), but one experiment with *Eucalyptus grandis* seedlings found large proportional increases in growth with eC_a under very low soil P, enabled by a decreased tissue P content (Conroy *et al*., 1992). There are few field data exploring interactions of eC_a with P availability. Predicting likely outcomes for the EucFACE ecosystem thus presents a real challenge for vegetation models.

To quantify potential impacts of eC_a and nutrient and water limitation at the EucFACE site, we employed seven vegetation models. These models are based on generalised global or large-scale parameterisations of key processes and have not been calibrated against observations from specific sites. The experiment provides a unique opportunity to explore the validity of these generalised assumptions for the case of a forest growing under nutrient and water limitations as well as the effects of rising C_a. Six of the models were used in the previous FACE Model-Data Synthesis project and thus have been fully evaluated against data from the Oak Ridge and Duke Forest FACE experiments (De Kauwe *et al.* 2013; 2014; Medlyn *et al.* 2015; Walker *et al.* 2014; 2015; Zaehle *et al*. 2014), while the seventh model, CLM4-P, is an extension of CLM4 that incorporates phosphorus cycling. By evaluating multiple model differences at the outset of EucFACE, we aimed to uncover fundamental differences in modelling approaches that would enable future data collection to be optimised for ongoing model improvements or future intercomparisons. Specific goals were to (1) provide a range of quantitative predictions for the experiment against which data can be assessed as they become available; and (2) identify key sets of measurements that would allow us to discriminate among variable model predictions driven by alternative model assumptions.

MATERIALS AND METHODS

Overview

We examined the responses of stocks and fluxes of C, N, P and water predicted by seven process-based vegetation models to 12 years of both ambient and elevated C_a

exposure. Modellers were provided with meteorological forcing, C_a, N deposition data and general site information describing the physiology and structure of the stand. Models were spun-up using historical climate data for the site. We simulated a 12-yr experimental period assuming two alternative sets of weather forcing. The 'fixed' weather forcing simply repeated weather data from a relatively wet year for the site location, while the 'variable' weather forcing used a 12-year period of weather chosen from the recent historical record that included a reasonably strong variability in water availability.

The models

Of the seven process-based vegetation models used in this inter-comparison, two models include both N and P limitations to growth: Community Atmosphere Biosphere Land Exchange (CABLE) (Wang *et al.,* 2010, 2011), and Community Land Model 4 with phosphorus (CLM4.0-CNP, henceforth CLM4-P) (Yang *et al*., 2014). Four models represent nitrogen limitation only: Community Land Model 4.0 (CLM4.0, henceforth CLM4) (Oleson *et al.,* 2010), Generic Decomposition and Yield (GDAY) (Comins & McMurtrie, 1993), Lund-Potsdam-Jena General Ecosystem Simulator with carbon-nitrogen cycling (LPJ-GUESS) (Smith *et al.,* 2014) and Orchidee-C-N (O-CN) (Zaehle & Friend, 2010). Finally, the Sheffield Dynamic Global Vegetation Model (SDGVM) (Woodward & Lomas, 2004) is not stoichiometrically limited by N and considers only an empirical N limitation whereby leaf N is a function of soil C, and photosynthetic rates are a function of leaf N (Woodward *et al.,* 1995). Although we classify the models by their nutrient limitations, the models differ in many ways. The key assumptions of each model are summarised in Table 1.

Simulation protocol

To generate hypothetical meteorological data for these simulations, we obtained a 20-year sequence from the recent past (1992-2011) from the closest 1.0-degree pixel to the experimental site from the global Princeton meteorological dataset (GPM; Sheffield *et al.,* 2006). Models were initialised by recycling this meteorological sequence, using pre-industrial C_a (277 μmol mol^{-1}) and N deposition (2.25 kg N ha^{-1} yr^{-1}), until model stocks of C and, where simulated, N and P, had equilibrated. Once equilibrated states had been obtained, models were run for a transient period to

account for changes in C, N and P balance induced by climate and C_a trends through the industrial period (1750-2012). Historical C_a concentration data were obtained from Vetter *et al.* (2008) and N deposition data from Dentener *et al.* (2006) for the closest location to the EucFACE site.

The site is dominated by *Eucalyptus tereticornis* trees. None of the models explicitly represents *Eucalyptus* and so they simulated the closest representative plant functional type, typically evergreen broadleaf trees. Estimated baseline data were provided for the site, including stand density, diameter at breast height (DBH) and standing biomass; physiological parameters; soil extractable water and texture; and soil nutrient contents. The information document provided to modellers is given as supporting information (Supplementary S1). This information was provided in June 2013, shortly after the start of the experiment. No information on experimental results from the site was available to the modellers at the time of the model runs, thus precluding calibration of the models against experimental data. Not all baseline information provided could be used in all models; in many cases models were run with their respective default or predicted parameter values.

Models that had the ability to represent stochastic fire events switched these mechanisms off for the course of the spin-up, transient and experimental period to facilitate comparison among models. Models were run for the forest canopy overstorey only, as most models were unable to simulate true mixtures of trees and grass, and the understorey forms a small component of total biomass at the site.

Experimental simulations were run for a hypothetical period 2012 – 2023, accounting for anticipated increase in ambient CO_2 concentrations, but no trends in climate drivers, during this period. For the ambient C_a simulations, we applied a conservative rate of increase in C_a, following the Representative Concentration Pathways (RCPs) emission pathway RCP3-PD (Meinshausen *et al.,* 2011). This scenario is identical to RCP4.5 in the simulation period and results in an increase in C_a in the ambient treatments from 393.8 to 418.6 ppm, a 24.8 ppm or 6% increase. For the elevated C_a simulations, we applied a ramp in C_a following the EucFACE experimental protocol (Drake *et al.* 2015): ambient + 30 ppm during September 2012; ambient +60 ppm during October 2012; ambient + 90 ppm during November 2012; ambient +120 ppm

during December 2012; and ambient + 150 ppm from January 2013 onwards (Figure 1a).

We ran simulations with both fixed and variable interannual meteorological data. For the fixed year we chose a relatively wet year (1998), while the varying 12-year sequence was selected to encompass periods of both wet and dry years (Figure 1b). The combination of fixed and variable simulations was designed to allow us to investigate interactions between eC_a and soil moisture availability. Hourly meteorological data were disaggregated from the 3-hourly GPM data using the CABLE weather generator (Haverd *et al.,* 2013). Disaggregated hourly temperature and precipitation data were compared to available local site meteorological measurements and were found to match annual means and intra-annual variability well (not shown).

Each model was used to perform four simulations:

1. AF: Ambient simulation with fixed meteorological data and RCP3-PD CO_2 concentrations.
2. AV: Ambient simulation with varying meteorological data and RCP3-PD CO_2 concentrations.
3. EF: Elevated simulation with fixed meteorological data and RCP3-PD CO_2 concentrations + ramp to 150 ppm.
4. EV: Elevated simulation with varying meteorological data and RCP3-PD CO_2 concentrations + ramp to 150 ppm.

Models were required to output up to 116 variables on a daily basis, including meteorological data and C, N, P and water fluxes and pools. Outputs were collated centrally and then checked to ensure mass balance of C, N, P and water in all models on an annual basis. Mass balance calculation details are shown in Supplement S5. We found that this level of output was required to verify that models were consistently using the same meteorological data and interpreting output variables in comparable ways. Daily and annual outputs were compared across models to quantify differences in predicted ambient ecosystem fluxes and their responses to elevated C_a.

RESULTS

Ambient C_a, constant climate

Modelled fluxes under ambient C_a conditions with constant climate are shown in Figure 2. Modelled GPP varied by as much as a factor of two among models, and depended on how nutrient limitation is represented in each model (Figure 2a). The lowest GPP was estimated by CLM4-P, which assumed strong limitations by both nitrogen (N) and phosphorus (P), and the highest GPP was estimated by SDGVM, which assumed no stoichiometric nutrient limitation. Among the N-limited models, a range of GPP values was predicted, with the lowest value (LPJ-GUESS) approaching that of CLM4-P, and the highest value (O-CN) approaching that of SDGVM.

There were sizeable differences in predicted ~~autotrophic respiration rates (R_{auto})~~carbon use efficiency (CUE) among models~~-~~, ~~due to~~which was caused by different model assumptions for autotrophic respiration, R_{auto}. ~~Differences among the models in assumptions relating to R_{auto} can be most clearly seen by comparing model estimates of carbon use efficiency (CUE),~~CUE ~~which~~ is the fraction of GPP not lost to respiration, calculated as~~or~~ $1 - R_{auto}$ / GPP. CUE was set to 50% in GDAY, and a similar ratio was predicted by both SDGVM and CABLE (Figure 2c). In O-CN, it is assumed that nutrient limitation leads to excess C being respired, here resulting in a relatively low CUE. High leaf and root respiration rates in CLM4 and CLM4-P, and high sapwood respiration rates in LPJ-GUESS, led to particularly low (~25%) CUE in these models. This low CUE meant that ambient NPP predicted by these models was low (Figure 2b), as little as 300 g C m^{-2} yr^{-1} in CLM4-P and LPJ-GUESS, compared to 1200 g C m^{-2} yr^{-1} predicted by the non-nutrient limited, high-CUE model SDGVM.

Predicted LAI was high, compared to the site estimate provided to modellers of 1.5 m^2 m^{-2}, in three of the seven models (Figure 2d). It was highest (> 4.5 m^2 m^{-2}) in SDGVM, a consequence of this model's optimal LAI allocation. In SDGVM, LAI is increased to the point where the carbon balance of the lowest layer of foliage in the canopy is zero. This optimisation scheme resulted in a high fraction of NPP allocated to foliage, well above that of the other models (see Supporting Figure S2). The variability among other models in predicted LAI depended on simulated NPP as well as the fraction of NPP allocated to foliage (which varied from 8 – 33%, Figure S2)

and the values of leaf mass per area (LMA) and leaf turnover time. Values of LMA in models ranged from 104 to 229 g dry matter m^{-2} (the site estimate provided was 229 g m^{-2}) while leaf turnover times ranged from one to three years (the site estimate provided was 18 months).

There was as much as a twofold difference among models in predicted canopy transpiration, ranging from ca. 350 mm yr^{-1} in GDAY up to ca. 760 mm yr^{-1} in O-CN (Figure 2e). This range represents 40 – 85% of annual precipitation in the constant climate scenario. Water use efficiency (WUE, calculated as GPP / transpiration) was similar in most models, meaning that differences in transpiration were correlated with differences in GPP. However, the CLM4-P model had particularly low WUE because canopy transpiration is calculated from potential GPP (i.e. GPP calculated without N and P limitations) rather than nutrient-limited GPP, resulting in relatively high transpiration rates.

Nutrient pools and fluxes also varied strongly among models (Figure 2f,g). While N deposition was specified as an input (averaging 3.3 g N m^{-2} yr^{-1} over the experimental period), the models differed in the amount of biological N fixation predicted, ranging from close to zero (O-CN, LPJ-GUESS) up to 5 g N m^{-2} yr^{-1} in CABLE, which uses the N fixation model of Houlton *et al.* (2008). Nitrogen losses from the system also varied among models, from close to zero up to 2.5 g N m^{-2} yr^{-1} gaseous losses (CLM4) and 4.8 g N m^{-2} yr^{-1} leaching (CABLE). Total soil nitrogen varied from 140 g N m^{-2} (LPJ-GUESS) to 810 g N m^{-2} (CLM4). Net N mineralisation and plant N uptake both ranged from ca 3 (in LPJ-GUESS) up to 11 g N m^{-2} yr^{-1} (in CABLE). Predicted canopy N content was lowest in LPJ-GUESS (~3 g N m^{-2}) and highest in CLM4 (~10 g N m^{-2}), bracketing the site estimate of 6 g N m^{-2}. The two models simulating P cycling, CABLE and CLM4-P, differed in their predictions of P uptake by a factor of 4 (Figure 2h).

Elevated C_a responses, constant climate

Differences among the models' predicted responses to elevated C_a are related to the alternative hypotheses for nutrient cycling embedded in the models (Figure 3). The non-stoichiometrically limited model, SDGVM, showed a sustained response of GPP to eC_a over the 12-yr simulation (Figure 3a). There was a lagged response of

respiration and thus CUE as biomass increased (Figure 3c), meaning that there was an initial strong stimulation of NPP (Figure 3b), which then relaxed to a slightly lower stimulation than that of GPP. Although sustained, the SDGVM response was smaller than predicted by some of the N-limited models, principally because SDGVM predicted a high LAI under ambient conditions (Figure 2d). Leaf photosynthesis was thus largely light-limited and was less responsive to C_a than in a more open canopy where light is less limiting.

The two P-limited models both simulated a minimal response of NPP to eC_a (Figure 3b) because the plants were unable to increase P uptake (Figure 3h) to support any increase in C uptake. The mechanism is similar in both models, although it is implemented in a slightly different way. In CLM4-P, potential GPP is down-regulated to the GPP that is able to be supported by P uptake, whereas in CABLE, GPP is unchanged and respiration rate is up-regulated (Figure 3a,c). Thus, CLM4-P showed a minimal eC_a effect on GPP, CUE, and NPP, whereas CABLE showed a substantial eC_a effect on GPP and CUE, yielding no overall increase in NPP.

Of the N-limited models, O-CN predicted the largest sustained GPP response to eC_a (Figure 3a). The initial response of G'DAY was similar to that of O-CN but this response declined over time. In G'DAY, the initial stimulation of GPP and NPP caused leaf N concentration ([N]) to fall (Figure 3f), driving a decline in litter quality that led to N immobilization in the soil. Reduced N availability for uptake (Figure 3g) then caused leaf [N] to fall further, causing a strong feedback on the eC_a response of GPP. In the O-CN model, limited N available for uptake drove a large increase in plant respiration, meaning that the NPP response to eC_a was small despite the large increase in GPP.

In contrast to O-CN and GDAY, the CLM4 model assumes fixed C:N stoichiometry in plant tissues (Table 1). Rather than a gradual feedback via increasing tissue C:N, this model assumes that the eC_a response of NPP and GPP are immediately limited if there is insufficient N available to build additional new tissue, so any increase in NPP is determined by whether or not there is additional N available for uptake. CLM4 has a high rate of denitrification, which is suppressed due to higher plant N demand under elevated C_a, leading to a ~10% increase in plant N uptake (Figure 3g) and

consequently a ~10% increase in NPP under elevated C_a (Figure 3b). By contrast P does not have a gaseous loss pathway and thus CLM4-P only increases P uptake via increased competitiveness for P with microbes, leading to a small (3-4%) eC_a response of P uptake, NPP, and GPP (Figures 3a, b, h).

The approach taken in LPJ-GUESS is quite different from other models because it uses an optimisation scheme for canopy photosynthesis that predicts canopy N based on the balance between photosynthesis and respiration (Haxeltine & Prentice 1996). This scheme results in an acclimation of photosynthesis to elevated C_a via a reduction in canopy N content. The eCa response of GPP was, therefore, strongly limited by available mineral N. However, the reduction in leaf N:C ratio under eC_a (Figure 3f) led to a reduction in respiration rate and thus a strong increase in CUE (Figure 3c), yielding a large response of NPP to eC_a (Figure 3b). It can also be observed that there is a year-to-year oscillation in the LPJ-GUESS outputs. This oscillation is due to the dependence of mineral soil N on leaf litter input, which occurs on an annual time-step. Annual time-steps can be problematic in global models because the appropriate date for processes may differ in northern and southern hemispheres. Here, the oscillation is exacerbated because, in the southern hemisphere, leaf litterfall accumulated from the previous calendar year becomes available for mineralisation in mid-winter, which is six months later, on 1st July.

Overall, the models can be divided into those with a strong positive NPP response to eC_a (> 20%; GDAY, LPJ-GUESS), an intermediate response (10 – 15%; SDGVM, CLM4) or minimal response (0-5%; O-CN, CABLE, CLM4-P) (Figure 3b). Predicted responses of LAI were similar in size to the respective NPP responses except in SDGVM, which predicted a small response of LAI due to a minimal increase in the optimal LAI under eCO_2 assumed by this model (Figure 3d). The two models that predicted the largest increases in LAI were GDAY and LPJ-GUESS, with predicted increases of ca. 16-20%.

The eC_a effect on transpiration varied among the models from -8% to +8% (Figure 3e). The differences among the models can be explained in terms of the eC_a effect on GPP (see above) and that on water-use efficiency (WUE). Inter-model differences in WUE are well-understood following the extensive analysis by De Kauwe *et al.*

(2013). The models considered here make very similar assumptions about the effect of eC_a on stomatal conductance – the ratio of photosynthesis to stomatal conductance is assumed to be nearly proportional to C_a – but they differ in how strongly transpiration is coupled to stomatal conductance (Table 1). The O-CN and GDAY models predicted large eC_a effects on WUE because transpiration is closely coupled to stomatal conductance. However, these models both had strong eC_a responses of GPP, meaning that the eC_a effect on transpiration was relatively small. In the CABLE model, the eC_a response of WUE was somewhat smaller than that of O-CN and GDAY because transpiration is less strongly coupled to stomatal conductance. However, CABLE's GPP response was also small, due to P limitation, so the limited increase in WUE translated into the largest reduction in transpiration among the models. Transpiration and stomatal conductance are not closely coupled in LPJ-GUESS, which predicted an increase in transpiration with eC_a despite a reduction in stomatal conductance.

Ambient C_a, variable climate

Models agreed that GPP and NPP would be reduced in years with low rainfall (2016, 2017, 2019) and increased in the wettest year (2021) but the size of the impact varied considerably among models (Figure 4a,b). In the driest year (2016), for example, GPP was reduced between 18% (CABLE, LPJ-GUESS) and 38% (CLM4, SDGVM) while NPP was reduced between 24% (CABLE) and 70% (CLM4, LPJ-GUESS). Thus, the spread among models in the predicted effect of interannual variability on GPP and NPP is of a similar magnitude to the spread in the predicted effect of increased C_a.

The principal reason for this variation among models is that different models represent the impact of drought on gas exchange very differently (Table 1). To contrast this effect among models, we requested each model output a reduction factor (β), which is the ratio of predicted gas exchange at a given soil water content (SWC) to that at field capacity. Figure 5 shows that the relationship between β and SWC differs greatly among models. Some models calculate β as a function of total SWC (GDAY, SDGVM) but differ in their parameterisation of this function, such that gas exchange is reduced at higher SWC values in GDAY than in SDGVM. The β function in GDAY is taken from Landsberg & Waring (1997) while the function used in

SDGVM is based on Gollan *et al.* (1992). The low drought sensitivity in SDGVM resulted in a very large reduction in soil water content (Figure 4f), because the plants continue transpiring until low SWC is reached.

The CABLE model calculates β as a function of root-weighted SWC, such that the water content of the upper soil layers has more impact on β than the lower layers. As a result, the same β may be obtained for a very wet soil that has undergone a short drought, or a very dry soil that has had recent rainfall. The CLM4 and CLM4-P models similarly use a root-weighting function, but it is applied to soil water potential (SWP) rather than SWC. Since SWP is a strongly non-linear function of SWC, this function implies that small changes in SWC can have a large impact on β. Figure 5 shows that the β factor declined very steeply with SWC in these models, resulting in an on-off behaviour of gas exchange as soil dried down. Further causing differences among models, the calculated β factor may be applied to stomatal conductance (CABLE), the maximum rate of Rubisco activity (CLM4, CLM4-P), or both of these (GDAY, O-CN, SDGVM). In models such as CABLE where β is only applied to stomatal conductance, water use efficiency increases in drought, such that transpiration can decline much more strongly than photosynthesis (e.g. compare Figure 4a,e for CABLE).

The LPJ-GUESS model takes a fundamentally different approach to calculating the drought impact: it calculates whether soil water supply (based on root-weighted soil water content and maximum transpiration rate) is sufficient to meet demand (based on photosynthetic rate) and reduces photosynthesis on days where supply is less than demand (Table 1). As a result, low soil water availability has most impact on days of high demand (those with high incident PAR) and may have no impact at all on days of low demand (such as days of cooler temperature). The β factor is more strongly related to incident PAR than to SWC.

Another reason for the difference among models is the non-drought LAI: models with high non-drought LAI and transpiration rate (e.g., CLM4, SDGVM) tended to show larger reductions in soil water content over time during drought. For example, the impact of drought was significantly larger in CLM4 than CLM4-P (Figure 4a)

although they use the same drought response function. Due to the P limitation on growth in CLM4-P, non-drought LAI was predicted to be considerably lower in this model (1.5 cf. 3.5 m^2 m^{-2} in CLM4, Figure 2d); consequently, transpiration rate was lower and soil water content remained higher than in CLM4.

The relative impact of drought years on NPP (as opposed to GPP) was greatest in those models that have a low baseline CUE, particularly CLM4 and LPJ-GUESS. In these models, a large fraction of GPP was used in respiration (up to 75%) (Figure 2c). In consequence, a relatively small reduction in GPP (ca. 20% in LPJ-GUESS) translated into a large reduction of about 80% in NPP in the first severe drought year, when respiration rates were still high (Figure 4b). In subsequent years, however, respiration rates were lower, due to lower biomass production, such that the impact of drought on NPP was lessened. This 'catch-up' effect on respiration is an example of a lagged effect of drought. Lagged effects can also be seen in the response of LAI in several models: in SDGVM, for example, the drought impact on GPP in 2016 and 2017 resulted in lower LAI in 2018, with consequences for GPP in that year and further feedbacks to LAI in subsequent years (Figure 4a, d).

Elevated C_a responses, variable climate.

The interaction between eC_a and rainfall is complex. Low water availability may lead to higher eC_a responses because stomatal closure can yield water savings, providing an additional stimulation to photosynthesis. However, feedbacks via increased LAI in eC_a may also negate water savings. To examine differences among models, we evaluated whether they predicted increased soil water content, increased LAI, both, or neither (Figure 6).

In the wettest year (Figure 6a), there are clear differences among models in whether the impact of eC_a is principally on LAI or SWC, which can be understood in terms of their underlying assumptions. The models SDGVM, CLM4 and GDAY tended to show an increase in LAI, because these models all simulated an increase in NPP that led to increased LAI, the resultant larger surface area for transpiration outweighing the effects of stomatal closure. In contrast, the O-CN model showed an increase in soil moisture, because it simulated no change in NPP or LAI. The two P-limited models CLM4-P and CABLE tended to show neither effect. In both these models

there was no change in NPP, due to P limitation. However, they also both simulated a relatively small eC_a effect on WUE (Table 1; De Kauwe *et al.,* 2013), so transpiration was not reduced either, leading to little effect on soil moisture content. The model LPJ-GUESS shows a reduction in soil moisture content, rather than an increase. In this model, transpiration is largely decoupled from stomatal conductance under non-water stress conditions and hence the eC_a-induced reduction in stomatal conductance does not lead to water savings (Table 1; De Kauwe *et al.,* 2013). Thus, the predicted increase in LAI in this model was accompanied by a reduction in average soil water content.

The differences among models are less obvious in the summer of the driest year (Figure 6b). None of the models show large soil water savings, which is because soil water differences are transient; in a dry year, soil moisture is still depleted, albeit at a slower rate. Similarly, no models show large increases in LAI, largely because of shifts in allocation from leaves to roots during periods of water stress.

This wide range of potential feedbacks leads to considerable variation in model predictions of the interaction between eC_a and rainfall. In Figure 7 we examine whether or not the models predicted a higher eC_a response in years of low rainfall. Five of the seven models showed a trend of decreasing GPP response with increasing rainfall. However, this only translated to a trend of decreasing NPP response with increasing rainfall in two of the seven models, due to lagged responses of respiration and LAI to changes in GPP, as well as feedbacks via soil nutrient availability. For example, O-CN showed a negative relationship between GPP response and annual rainfall, but showed the strongest NPP response in the wettest year. This reversal is related to soil water impacts on N mineralisation, which allowed an NPP response in wet years but not in dry years.

Although most models did not predict an interaction between eC_a and rainfall in their effects on NPP, they did predict that eC_a could ameliorate the effects of drought, at least to some degree. Figure 8 shows ambient and elevated C_a GPP and NPP during the three driest years (2016, 2017, 2019), compared to their values in the fixed climate simulations. As noted above, the reductions in ambient C_a range across models from -20 to -70%. For GPP, every model predicted that production under eC_a is higher than

under aC_a, while for NPP only the O-CN model predicted that productivity in drought years would not be increased with eC_a. The impact of drought on NPP was particularly strongly ameliorated by eC_a in three models: GDAY, LPJ-GUESS and SDGVM.

Consequences for ecosystem C storage

The C budgets for each model over the 12 simulation years are given in Table 2. Values are given for the fixed climate simulations; similar values are obtained in the variable climate simulations. Following initial spin-up, the models predicted the plant C pool to be in the range of 5 kg m^{-2} (CLM4-P) to 25 kg m^{-2} (O-CN), straddling the best estimate for the site of 12.7 kg m^{-2} (Supplementary S3), while the soil C pool was in the range of 1.5 kg m^{-2} (LPJ-GUESS) to 8.5 kg m^{-2} (CABLE). Under ambient conditions, the changes in plant, soil and litter carbon over the 12 years of the experiment are relatively small fractions of total NPP, which is expected as the simulations should have been roughly in equilibrium following spin-up. Litterfall exceeded NPP in the O-CN model; as a result of the strict self-thinning assumptions in O-CN (Walker *et al.,* 2015), this model predicted net loss of plant C as the tree stand grew and densified over the 12 years.

The fraction of the extra NPP due to eC_a remaining in the system after 12 years (Table 2, lower half) varied among models depending on their assumptions about allocation and turnover (cf. De Kauwe at al. 2014). In CABLE, CLM4 and CLM4-P, allocation patterns and turnover are assumed fixed during the growing season; all three models predicted approximately 40% of the extra NPP would remain in the plants. The four models with variable allocation patterns all predicted an increase in the fraction of C allocated to wood with elevated C_a (Figure S2). In the three functional-balance type models (GDAY, LPJ-GUESS, O-CN), this shift occurred because the alleviation of water stress by eC_a outweighed the increase in nutrient stress, while in SDGVM a smaller foliage allocation is required to reach optimal LAI under eC_a (Figure S2). In LPJ-GUESS and GDAY a shift of allocation towards wood resulted in 60-70% of the extra NPP remaining in the plants. The two models SDGVM and O-CN have contrasting self-thinning assumptions (cf. Walker *et al.,* 2015). In the O-CN model, there was a net loss of carbon from the plant C pool under eC_a despite an overall 7.5% increase in NPP, due to accelerated stand decline. In

contrast, in the SDGVM model, all of the 16% increase in NPP remained in the plant, due to a decrease in turnover predicted by this model.

When additional NPP due to eC_a remains in the plant, the stimulation of litterfall and, consequently, heterotrophic respiration (R_{het}), under eC_a will be smaller than the stimulation of NPP. Thus, the only model in which R_{het} increased more than NPP was O-CN, which predicted an increase in R_{het} of 9%. The change in R_{het} was no more than 6% in the other models. Predicted changes in soil C over time were similarly small; the maximum predicted change in soil C stock due to eC_a was 122 g m^{-2} over the twelve years of simulation (CLM4).

DISCUSSION

Large-scale ecosystem experiments like EucFACE, provide a major opportunity to improve ecosystem models by investigating responses to a perturbation in a situation where drivers and most responding processes can be directly measured and quantified. To maximise this opportunity for EucFACE, we took the important step of carrying out a model intercomparison towards the start of the experiment, rather than at the end. We had two principal goals for this model intercomparison. Firstly, we aimed to provide a range of baseline model outputs against which experimental results could be compared, such as predicted responses of GPP, NPP, LAI, transpiration and soil respiration to eC_a (Figures 2, 3; Table 2). Such baseline model predictions can be used to help optimize sampling regimes by indicating the likely effect size for key variables. They also enable competing model assumptions to be evaluated as data emerge from the experiment (e.g. Duursma *et al.,* 2015; Gimeno *et al.,* 2015).

Secondly, we aimed to identify key measurements that would allow discrimination among competing model hypotheses, thus ensuring that the experiment would be able to address outstanding model uncertainties. Some of the important differences among models shown in this comparison have been previously documented in model intercomparisons against the Duke and ORNL Forest FACE experiments, including differences in the response of transpiration to stomatal conductance (De Kauwe *et al.,* 2013), the flexibility of carbon-nitrogen stoichiometry (Zaehle *et al.,* 2014), and influences of elevated C_a on allocation patterns (De Kauwe *et al.,* 2014) and woody

biomass turnover (Walker *et al.,* 2015). These model uncertainties remain key questions to address in the EucFACE experiment. In addition, our comparison has highlighted three further critical areas of uncertainty: P limitation, autotrophic respiration, and drought stress.

Phosphorus limitation

Of the seven models considered here, one incorporated no stoichiometric nutrient limitation, four incorporated N limitation, and two incorporated both N and P limitation. The two models considering P limitation predicted the lowest eC_a response (Figure 3). The strong P-limitation arises in part because the P-cycle is assumed in the models to be relatively 'closed' (Kirschbaum *et al.,* 1998; Wang *et al*., 2010), meaning that there are few gains and losses from the system and therefore little capacity for the vegetation to increase P uptake. However, this assumption may be incorrect, as there is ample evidence from P-limited ecosystems that incoming C can enable plants to mineralise or acquire more P (Lambers *et al.,* 2008; 2012; Nazeri *et al.,* 2013). For example, biochemical mineralization of organic P through phosphatase activity can be enhanced under eC_a through increased production of phosphatase enzymes, in response to increasing P limitation and increased amounts of fine root biomass, mycorrhizal fungi and soil microbes. A simple approach of modelling biochemical mineralization may not be able to accurately capture this feedback pathway. Also, it has been suggested that mycorrhizae play an important role in P uptake (Bolan, 1991; Cairney *et al*., 2011). Under eC_a, mycorrhizal associations can be stimulated (Treseder, 2004; Nie *et al.,* 2013). Therefore the models may be underestimating P uptake under eC_a by not including mycorrhizal associations. Mycorrhizae also represent a potential C sink, unaccounted for in most models (Fransson, 2012). Another possible pathway for increasing P availability under eC_a is through increased desorption of secondary P minerals, which occurs as a result of increased SOM and root exudation of carboxylates that compete for sorption sites with phosphate ions (Lambers *et al.,* 2006). A modelling sensitivity study by Yang *et al.* (2014) showed that these processes can significantly affect soil P availability and determine the extent of P limitation in P-limited ecosystems under eC_a.

In addition, the P-limited models (CABLE, CLM4-P) assumed or simulated limited flexibility in whole-plant C:P ratios. Flexibility in these ratios would allow increased

plant C uptake even where there is no increase in P uptake. Unfortunately, little is known about the flexibility of these ratios under eC_a. Furthermore, there is still considerable debate concerning how P limitations to photosynthesis should be represented (Reich *et al.,* 2009, Ellsworth *et al.,* 2015). Clearly, measurements related to P uptake and P use efficiency at EucFACE would be valuable for the further development and improvement of the models by helping to quantify these important plant-nutrient feedbacks.

Autotrophic respiration

The intercomparison also highlighted very different ~~assumptions for the~~ effects of eC_a on ~~autotrophic respiration~~carbon use efficiency under nutrient limitation, due to different assumptions about autotrophic respiration. The GDAY model simply assumes a constant CUE, meaning that plant respiration varies in direct proportion to modelled GPP. The LPJ-GUESS model predicts that leaf N concentration should decline under eC_a, leading to a reduction in respiration rates per unit biomass (Ryan *et al.,* 1991; Reich *et al.,* 2008) and an increase in CUE. In contrast, CABLE and O-CN assume that excess labile, non-structural C accumulates when the plant is unable to build new tissue due to nutrient limitation, and this C accumulation, in effect, drives an increase in tissue respiration rates and a decrease in CUE. All three competing hypotheses are testable. Targeted measurements of tissue respiration, in conjunction with tissue N content and tissue carbohydrate content, are needed at EucFACE in order to distinguish among these hypotheses.

Previous experiments suggest that respiration rates in plant tissues may correlate positively with both N and increasing carbohydrate contents in plants grown under eC_a (Tjoelker *et al*., 1999a), with no change in CUE (Tjoelker et al., 1999b). As yet, there is little evidence for increased respiration with excess C accumulation as a direct response to reduced nutrient availability and uptake. Specific rates of respiration in plant tissues, particularly roots, typically decline with nutrient limitation, owing to lower respiratory costs from reduced ion uptake, transport and assimilation and reduced growth demands for adenylates and C skeletons (Lambers *et al.*, 1983; Amthor, 2000). Engagement of the alternative oxidase pathway could provide a means to consume C without feedbacks from adenylate cycling, altering the efficiency of ATP synthesis (Millar *et al*., 2011), though total respiratory C fluxes would not

necessarily be expected to change. At the whole-plant scale, increases in proportional allocation of C to root biomass under nutrient limitations would, in effect, increase respiratory C losses relative to GPP, reducing CUE (Poorter *et al*., 1995).

These observations tend not to support the hypothesis of excess respiration following carbohydrate build-up. However, it can be argued that this hypothesis is not a plant physiological hypothesis as such, but rather is a mechanism used in models to deplete excess carbpm under nutrient limitation, acting as a proxy for carbon never assimilated, or losses from the plant through other mechanisms such as root exudation. The use of such proxies in modelling reflects the fact that at a fundamental level, we do not fully understand, and thus do not know how correctly to model, the effect of nutrient limitation on the carbon balance of plants. Some of the models simulate a reduction in photosynthetic C uptake under nutrient limitation, either via an implausibly large reduction in leaf nutrient content (GDAY) or an unspecified mechanism of strong photosynthetic down-regulation (CLM4, CLM4-P). Other models (CABLE, O-CN) effect nutrient limitation of growth by simulating an increase in respiration. Both approaches are imperfect solutions to the problem that the reduction of photosynthesis under low nutrient availability is generally smaller than the observed reduction in growth (Reich, 2012). At the nutrient-limited EucFACE site, it will be particularly helpful to construct an ecosystem-scale mass balance of carbon to enable the models to close the carbon budget through a correct balance of component fluxes.

Impacts of low rainfall

Another striking outcome of the comparison was that the models disagreed as much, if not more, about the effect of low-rainfall years, as about the effect of eC_a on productivity (Figure 4). Intensive ecosystem-scale experiments such as FACE have previously provided a wide range of insights into ecosystem function well beyond responses to eC_a (Norby & Zak 2011). In the same way, low-rainfall periods at the EucFACE experiment provide an excellent opportunity to reduce model uncertainty related to the impacts of drought.

Our analysis of low-rainfall years showed that the models embed several plausible alternative hypotheses for effect of drought on plant function (Table 1). However, it

was striking that there was a limited empirical basis for most of the drought-stress functions used in the models. The function used in O-CN, for example, can be traced back to Abramopoulos *et al.* (1988), who do not refer to experimental data. No source is given for the functions used in CLM4/CLM4-P (Oleson *et al.,* 2010) or LPJ-GUESS (Sitch *et al.,* 2003). The soil moisture functions employed in both CABLE and SDGVM can be traced back to work on sunflower by Gollan *et al.* (1986, 1992). The shape of the soil moisture function in GDAY is based on work on corn by Denmead & Shaw (1961), but Landsberg & Waring (1997) state that their parameter values are chosen without empirical justification. There is a clear opportunity here and in other experiment-model intercomparisons to improve models by developing more evidence-based functions for the impact of drought stress (cf. De Kauwe *et al*., 2015; Smith *et al.,* 2015). The different drought response functions used in the models are distinguishable by their predicted time-courses for the progression of drought (Figure 5). Targeted EucFACE measurements of leaf-scale gas exchange and whole-tree sapflux, in conjunction with root biomass and time-courses of soil water content at a range of soil depths would allow us to discriminate among these competing hypotheses.

Possible dead-ends

In addition to highlighting processes where experimentation can really help to constrain models, the model intercomparison also flags some analyses that may not help to inform models. Firstly, it is common to test for an interaction of drought and eC_a by plotting the ratio of elevated to ambient NPP in a given year against the rainfall in that year, but the expected positive drought x eC_a interaction is often not found (e.g. Morgan *et al.,* 2004; Nowak *et al.,* 2004; McCarthy *et al.,* 2010). Our model results (Figure 7) suggest that this analysis is not a good test to identify interactions between eC_a and water availability. The models tested here all incorporate the standard theory for eC_a x drought interactions, but generally do not show a significant correlation between the eC_a effect and rainfall on NPP, due to the many feedbacks at play. Applying the same analysis to experimentally measured NPP data, which include random variability and measurement inaccuracy, is very unlikely to show any interaction.

Secondly, we compared predicted change in C storage over the 12-year simulation period due to eC_a, with the estimated initial C storage in each model. The predicted change in plant C was up to 10% of the initial plant C, a change which should be detectable. However, the predicted change in total soil C was never above 4% of the initial soil C, suggesting that changes in total soil C are unlikely to be detectable. It may be more useful to focus on changes in soil C fractions and isotopic signatures rather than total soil C (see Hofmockel *et al.,* 2011; Norby & Zak, 2011; Iversen *et al.,* 2012).

Good modelling practice

We strongly recommend that model intercomparisons, targeting a range of site-relevant models, be used in advance of large-scale ecosystem experiments to provide baseline expectations and to indicate key areas of model uncertainty. Some recommendations for this process can be made based on our experience here.

First, we requested that models output a large number of variables (>80), including the meteorological data they used, and all components of the carbon, water, N and P balances. These outputs were used to check model mass balances: for example, we checked whether NPP was equal to the sum of growth of all plant components, and whether the change in soil water content was equal to precipitation less evapotranspiration, runoff and drainage. These checks were invaluable for identifying errors in model outputs and ensuring that outputs were consistent across models. Mass balances ought to be carefully checked as a standard step in all modelling exercises, not just model intercomparisons.

Secondly, we recommend that *a priori* model intercomparisons use an "assumption-centred" approach to analysing the model outputs (Medlyn *et al.,* 2015). This approach recognises that each model is composed of a large number of assumptions, and focuses on identifying which model assumptions are chiefly responsible for differences in model predictions. This approach provides significantly more guidance for experimentalists than simply comparing summary model outputs, which often reflect interactions and feedbacks among multiple processes. An assumption-centred analysis identifies clearly the competing hypotheses and the data needed to test them, which is a key aim of such intercomparisons.

Finally, we plan to revisit these simulations when experimental data become available for evaluating and improving the models. It is important that we fully document the simulations now so that we can repeat them with actual meteorological data in the future. Model code, input parameter files and outputs have been archived, and the site information document, modelling protocols and mass balance check scripts are provided as supporting information to this article (S3 – S5). This material will not only allow us to recreate model runs in future, but should also enable other modelling groups to apply their models to the EucFACE experiment, which we encourage.

ACKNOWLEDGEMENTS

The National Climate Change Adaptation Research Facility (NCCARF), Primary Industries Adaptation Research Network (PIARN) supported this project and travel for the participants to Sydney, Australia. Additional support via EucFACE as an initiative supported by the Australian Government through the Education Investment Fund and the Department of Industry and Science, in partnership with the University of Western Sydney, is acknowledged. Research support from the Australian Research Council is also acknowledged. Contributions from APW, XJY, MDK, KL and RJN were supported by the US Department of Energy (DOE) Office of Science's Biological and Environmental Research (BER). The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7 2007-2013) under grant agreement n° 238366 (Greencycles II). This study is a contribution to MERGE, a strategic research area of Lund University.

REFERENCES

Abramopoulos F, Rosenzweig C, Choudhury B (1988) Improved Ground Hydrology Calculations for Global Climate Models (GCMs): Soil Water Movement and Evapotranspiration. Journal of Climate, **1**, 921-941.

Amthor JS. (2000) The McCree–de Wit–Penning de Vries–Thornley respiration paradigms: 30 years later. Annals of Botany **86**, 1-20.

Arora VK, Boer GJ, Friedlingstein P *et al.* (2013) Carbon-Concentration and Carbon-Climate Feedbacks in CMIP5 Earth System Models. Journal of Climate, **26**, 5289-5314.

Bolan NS (1991) A critical review on the role of mycorrhizal fungi in the uptake of phosphorus by plants. Plant and Soil, **134**, 189-207.

Cairney JWG (2011) Ectomycorrhizal fungi: the symbiotic route to the root for phosphorus in forest soils. Plant and Soil, **344**, 51-71.

Cleveland CC, Townsend AR, Taylor P *et al.* (2011) Relationships among net primary productivity, nutrients and climate in tropical rain forest: a pan-tropical analysis. Ecology Letters, **14**, 939-947.

Comins HN, McMurtrie RE (1993) Long-term response of nutrient-limited forests to CO_2 enrichment; equilibrium behavior of plant-soil models. Ecological Applications, **3**, 666-681.

Conroy JP, Milham PJ, Bevege DI, Barlow EWR (1990) Influence of phosphorus deficiency on the growth response of 4 families of *Pinus radiata* seedlings to CO_2-enriched atmospheres. Forest Ecology and Management, **30**, 175-188.

Conroy JP, Milham PJ, Barlow EWR (1992) Effect of nitrogen and phosphorus availability on the growth response of *Eucalyptus grandis* to high CO_2. Plant Cell and Environment, **15**, 843-847.

Crous KY, Osvaldsson A, Ellsworth DS (2015) Is phosphorus limiting in a mature Eucalyptus woodland? Phosphorus fertilisation stimulates stem growth. Plant and Soil, **391**, 293-305.

De Kauwe MG, Medlyn BE, Zaehle S *et al.* (2013) Forest water use and water use efficiency at elevated CO2: a model-data intercomparison at two contrasting temperate forest FACE sites. Global Change Biology, **19**, 1759-1779.

De Kauwe MG, Medlyn BE, Zaehle S *et al.* (2014) Where does the carbon go? A model-data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO2 enrichment sites. New Phytologist, **203**, 883-899.

De Kauwe MG, Zhou S-X, Medlyn BE, *et al.* (2015) Do land surface models need to include differential plant species responses to drought? Examining model predictions across a latitudinal gradient in Europe. Biogeosciences Discussions.

Denmead OT, Shaw RH (1961) Availability of soil water to plants as affected by soil moisture content and meteorological conditions. Agronomy Journal **54**, 385-390.

Dentener F, Drevet J, Lamarque JF *et al.* (2006) Nitrogen and sulfur deposition on regional and global scales: A multimodel evaluation. Global Biogeochemical Cycles, **20**.

Dietze MC, Serbin SP, Davidson C *et al.* (2014) A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. Journal of Geophysical Research-Biogeosciences, **119**, 286-300.

Dijkstra FA, Blumenthal D, Morgan JA, Pendall E, Carrillo Y, Follett RF (2010) Contrasting effects of elevated CO2 and warming on nitrogen cycling in a semiarid grassland. New Phytologist, **187**, 426-437.

Drake JE, Macdonald CA, Tjoelker MG *et al.* (2015) Short-term carbon cycling responses of a mature eucalypt woodland to gradual stepwise enrichment of atmospheric CO_2 concentration. Global Change Biology doi: 10.1111/gcb.13109

Dukes JS, Chiariello NR, Cleland EE *et al.* (2005) Responses of grassland production to single and multiple global environmental changes. Plos Biology, **3**, 1829-1837.

Dukes JS, Classen AT, Wan S, Langley JA (2014) Using results from global change experiments to inform land model development and calibration. New Phytologist, **204**, 744-746.

Duursma RA, Gimeno T, Boer M *et al.* (2015) Canopy leaf area of a mature evergreen *Eucalyptus* woodland does not respond to elevated atmospheric [CO_2] but tracks water availability. Global Change Biology 10.1111/gcb.13151

Edwards EJ, McCaffery S, Evans JR (2005) Phosphorus status determines biomass response to elevated CO_2 in a legume: C-4 grass community. Global Change Biology, **11**, 1968-1981.

Ellsworth DS, Crous KY, Lambers H, Cooke J (2015) Phosphorus recycling in photorespiration maintains high photosynthetic capacity in woody species. Plant Cell and Environment, **38**, 1142-1156.

Fransson P (2012) Elevated CO_2 impacts ectomycorrhiza-mediated forest soil carbon flow: fungal biomass production, respiration and exudation. Fungal Ecology, **5**, 85-98.

Friedlingstein P, Meinshausen M, Arora VK, Jones CD, Anav A, Liddicoat SK, Knutti R (2014) Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. Journal of Climate, **27**, 511-526.

Gimeno T, Crous KY, Cooke J, *et al.* (2015) Conserved stomatal behaviour under elevated CO_2 and varying water availability in a mature woodland. Functional Ecology 10.1111/1365-2435.12532

Gollan T, Passioura JB, Munns R (1986) Soil water status affects the stomatal conductance of fully turgid wheat and sunflower leaves. Australian Journal of Plant Physiology, **13**, 459-464.

Gollan T, Schurr U, Schulze ED (1992) Stomatal response to drying soil in relation to changes in the xylem sap composition of *Helianthus annuus* .1. The concentration of cations, anions, amino-acids in, and ph of, the xylem sap. Plant Cell and Environment, **15**, 551-559.

Haverd V, Raupach MR, Briggs PR *et al.* (2013) The Australian terrestrial carbon budget. Biogeosciences, **10**, 851-869.

Haxeltine A, Prentice IC (1996) A general model for the light-use efficiency of primary production. Functional Ecology, **10**, 551-561.

Hofmockel KS, Zak DR, Moran KK, Jastrow JD (2011) Changes in forest soil organic matter pools after a decade of elevated CO2 and O-3. Soil Biology & Biochemistry, **43**, 1518-1527.

Houlton BZ, Wang Y-P, Vitousek PM, Field CB (2008) A unifying framework for dinitrogen fixation in the terrestrial biosphere. Nature, **454**, 327-U334.

Iversen CM, Keller JK, Garten CT, Jr., Norby RJ (2012) Soil carbon and nitrogen cycling and storage throughout the soil profile in a sweetgum plantation after 11 years of CO2-enrichment. Global Change Biology, **18**, 1684-1697.

Kirschbaum MUF, Medlyn BE, King DA *et al.* (1998) Modelling forest-growth response to increasing CO_2 concentration in relation to various factors affecting nutrient supply. Global Change Biology, **4**, 23-41.

Lambers H, Szaniawski RK, de Visser R. (1983) Respiration for growth, maintenance and ion uptake: An evaluation of concept, methods, values and their significance. Physiologia Plantarum, **58**, 556–563.

Lambers H, Shane MW, Cramer MD, Pearse SJ, Veneklaas EJ (2006) Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. Annals of Botany, **98**, 693-713.

Lambers H, Raven JA, Shaver GR, Smith SE (2008) Plant nutrient-acquisition strategies change with soil age. Trends in Ecology & Evolution, **23**, 95-103.

Lambers H, Bishop JG, Hopper SD, Laliberte E, Zuniga-Feest A (2012) Phosphorus-mobilization ecosystem engineering: the roles of cluster roots and carboxylate exudation in young P-limited ecosystems. Annals of Botany, **110**, 329-348.

Landsberg JJ, Waring RH (1997) A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology & Management, **95**, 209-228.

Lewis JD, Ward JK, Tissue DT (2010) Phosphorus supply drives nonlinear responses of cottonwood (Populus deltoides) to increases in CO2 concentration from glacial to future concentrations. New Phytologist, **187**, 438-448.

Luo Y (2001) Transient ecosystem responses to free-air CO2 enrichment (FACE): experimental evidence and methods of analysis. New Phytologist, **152**, 3-8.

Luo Y, Melillo J, Niu S *et al.* (2011) Coordinated approaches to quantify long-term ecosystem dynamics in response to global change. Global Change Biology, **17**, 843-854.

McCarthy HR, Oren R, Johnsen KH *et al.* (2010) Re-assessment of plant carbon dynamics at the Duke free-air CO2 enrichment site: interactions of atmospheric [CO2] with nitrogen and water availability over stand development. New Phytologist, **185**, 514-528.

McGroddy ME, Daufresne T, Hedin LO (2004) Scaling of C : N : P stoichiometry in forests worldwide: Implications of terrestrial redfield-type ratios. Ecology, **85**, 2390-2401.

McMurtrie R, Comins HN (1996) The temporal response of forest ecosystems to doubled atmospheric CO_2 concentration. Global Change Biology, **2**, 49-57.

Medlyn BE, Zaehle S, De Kauwe MG *et al.* (2015) Using ecosystem experiments to improve vegetation models. Nature Climate Change, **5**, 528-534.

Meinshausen M, Smith SJ, Calvin K *et al.* (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. Climatic Change, **109**, 213-241.

Millar AH, Whelan J, Soole KL, Day DA. (2011). Organization and regulation of mitochondrial respiration in plants. Annual Review of Plant Biology **62**, 79-104.

Morgan J, Pataki DE, Korner C *et al.* (2004) Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO_2. Oecologia, **140**, 11-25.

Nazeri NK, Lambers H, Tibbett M, Ryan MH (2013) Do arbuscular mycorrhizas or heterotrophic soil microbes contribute toward plant acquisition of a pulse of mineral phosphate? Plant and Soil, **373**, 699-710.

Nie M, Lu M, Bell J, Raut S, Pendall E (2013) Altered root traits due to elevated CO_2: a meta-analysis. Global Ecology and Biogeography, **22**, 1095-1105.

Norby RJ, Warren JM, Iversen CM, Medlyn BE, Mcmurtrie RE (2010) CO_2 enhancement of forest productivity constrained by limited nitrogen availability. Proceedings of the National Academy of Sciences of the United States of America, **107**, 19368-19373.

Norby RJ, Zak DR (2011) Ecological Lessons from Free-Air CO_2 Enrichment (FACE) Experiments. In: *Annual Review of Ecology, Evolution, and Systematics, Vol 42.* (eds Futuyma DJ, Shaffer HB, Simberloff D) pp 181-203.

Norby RJ, De Kauwe MG, Domingues T, *et al.* (2016) Model-data synthesis for the next generation of forest FACE experiments. New Phytologist, **209**, 17-28.

Nowak RS, Ellsworth DS, Smith SD (2004) Functional responses of plants to elevated atmospheric CO_2 - do photosynthetic and productivity data from FACE experiments support early predictions? New Phytologist, **162**, 253-280.

Oleson KW, Lawrence DM, Bonan GB, *et al*. (2010) Technical Description of version 4.0 of the Community Land Model (CLM). National Centre for Atmospheric Research, Boulder, CO, USA.

Parton WJ, Morgan JA, Wang G, Del Grosso S (2007) Projected ecosystem impact of the Prairie Heating and CO2 Enrichment experiment. New Phytologist, **174**, 823-834.

Piao S, Sitch S, Ciais P *et al.* (2013) Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO_2 trends. Global Change Biology, **19**, 2117-2132.

Poorter H, Van de Vijver CADM, Boot RGA, Lambers H. (1995). Growth and carbon economy of a fast-growing and slow-growing grass species as dependent on nitrate supply. Plant and Soil **171**, 217-227.

Reich PB, Tjoelker MG, Pregitzer KS, Wright IJ, Oleksyn J, Machado J-L (2008) Scaling of respiration to nitrogen in leaves, stems and roots of higher land plants. Ecology Letters, **11**, 793-801.

Reich PB, Oleksyn J, Wright IJ (2009) Leaf phosphorus influences the photosynthesis-nitrogen relation: a cross-biome analysis of 314 species. Oecologia, **160**, 207-212.

Reich PB (2012) Key canopy traits drive forest productivity. Proceedings of the Royal Society B-Biological Sciences, **279**, 2128-2134.

Reich PB, Hobbie SE, Lee TD (2014) Plant growth enhancement by elevated CO2 eliminated by joint water and nitrogen limitation. Nature Geoscience, **7**, 920-924.

Rogers A, Medlyn BE, Dukes JS (2014) Improving representation of photosynthesis in Earth System Models. New Phytologist, **204**, 12-14.

Ryan MG (1991) A simple method for estimating gross carbon budgets for vegetation in forest ecosystems. Tree Physiology, **9**, 255-266.

Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. Journal of Climate, **19**, 3088-3111.

Sitch S, Smith B, Prentice IC *et al.* (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Global Change Biology, **9**, 161-185.

Smith B, Warlind D, Arneth A, Hickler T, Leadley P, Siltberg J, Zaehle S (2014) Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. Biogeosciences, **11**, 2027-2054.

Smith NG, Rodgers VL, Brzostek ER *et al.* (2014) Toward a better integration of biological data from precipitation manipulation experiments into Earth system models. Reviews of Geophysics, **52**, 412-434.

Tjoelker MG, Reich PB, Oleksyn J (1999a). Changes in leaf nitrogen and carbohydrates underlie temperature and CO_2 acclimation of dark respiration in five boreal tree species. Plant, Cell and Environment, **22**, 767-778.

Tjoelker MG, Oleksyn J, Reich PB. (1999b) Acclimation of respiration to temperature and CO_2 in seedlings of boreal tree species in relation to plant size and relative growth rate. Global Change Biology **5**, 679-691.

Treseder KK (2004) A meta-analysis of mycorrhizal responses to nitrogen, phosphorus, and atmospheric CO2 in field studies. New Phytologist, **164**, 347-355.

Vetter M, Churkina G, Jung M *et al.* (2008) Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models. Biogeosciences, **5**, 561-583.

Walker AP, Hanson PJ, De Kauwe MG *et al.* (2014) Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO2 enrichment experiments: Model performance at ambient CO2 concentration. Journal of Geophysical Research-Biogeosciences, **119**, 937-964.

Walker AP, Zaehle S, Medlyn BE *et al.* (2015) Predicting long-term carbon sequestration in response to CO2 enrichment: How and why do current ecosystem models differ? Global Biogeochemical Cycles, **29**, 476-495.

Wang YP, Law RM, Pak B (2010) A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. Biogeosciences, **7**, 2261-2282.

Wang YP, Kowalczyk E, Leuning R *et al.* (2011) Diagnosing errors in a land surface model (CABLE) in the time and frequency domains. Journal of Geophysical Research-Biogeosciences, **116**.

Woodward FI, Smith TM, Emanuel WR (1995) A global land primary productivity and phytogeography model. Global Biogeochemical Cycles, **9**, 471-490.

Woodward FI, Lomas MR (2004) Vegetation dynamics - simulating responses to climatic change. Biological Reviews, **79**, 643-670.

Yang X, Thornton PE, Ricciuto DM, Post WM (2014) The role of phosphorus dynamics in tropical forests - a modeling study using CLM-CNP. Biogeosciences, **11**, 1667-1681.

Zaehle S, Friend AD (2010) Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. Global Biogeochemical Cycles, **24**.

Zaehle S, Medlyn BE, De Kauwe MG *et al.* (2014) Evaluation of 11 terrestrial carbon-nitrogen cycle models against observations from two temperate Free-Air CO2 Enrichment studies. New Phytologist, **202**, 803-822.

SUPPORTING INFORMATION CAPTIONS

S1: Document: Site information and model parameters supplied to modellers.

S2: Figure: Simulated allocation patterns at aC_a and changes in allocation due to eC_a in the variable climate simulations. Values at left are % of NPP used in growth of the plant component (a_{fol} : foliage; a_{wood}: wood; a_{root} : fine + coarse root; a_{repro} : reproduction). Values at right are the change in %NPP allocated to different components.

S3: Document: Modelling protocol.

S4: Document: Output protocol.

S5: Document: Additional information, including list of mass balance checks applied to model outputs, and location of archived model code, inputs and outputs.

1046 **TABLES**

1047 Table 1: Summary of key assumptions in models used in this paper. See text for further model details & references.

Model	Nutrient limitation mechanism	Autotrophic respiration	Water use efficiency	Allocation	Soil moisture content effect on gas exchange
P & N limitation					
CABLE	If N or P uptake is insufficient to support potential NPP, NPP is reduced, and excess C is allocated to labile C pool	Maintenance respiration is proportional to biomass, plus extra respiration from labile carbon pool. Growth respiration is proportional to leaf tissue N:P ratio	Stomatal conductance is proportional to GPP. Moderate coupling of transpiration to stomatal conductance.	Fixed allocation	β calculated from root-weighted SWC. Applied to g_1 and respiration.
CLM4-P	If N or P uptake is insufficient to support potential GPP, GPP is reduced	Proportional to biomass	Stomatal conductance is proportional to potential (non-nutrient limited) GPP.	Fixed allocation	β calculated from root-weighted SWP. Applied to V_{cmax}.
N limitation					
CLM4	If N uptake is insufficient to support potential GPP, GPP is reduced	Proportional to biomass	Stomatal conductance is proportional to potential (non-nutrient limited) GPP.	Fixed allocation	β calculated from root-weighted SWP. Applied to V_{cmax}.
GDAY	Progressive nitrogen limitation: as C uptake increases, foliage nitrogen content decreases, causing	50% of total GPP	Stomatal conductance is proportional to GPP. Strong coupling.	Pipe model for leaf vs stem allocation; functional balance for leaf vs root allocation; constrained by	β calculated from total SWC. Applied to g_1, V_{cmax} and J_{max}.

	nitrogen immobilization and further reducing N availability			min and max values	
LPJ-GUESS	Progressive nitrogen limitation as in GDAY, but with lower bound for leaf N:C ratio. Wood and root N:C vary as well as foliage.	Proportional to tissue N content	Stomatal conductance is proportional to GPP. Weak coupling.	Allocation to roots increases with nitrogen and water stress	Plant transpiration is the minimum of supply and demand, where supply is the product of plant root-weighted soil moisture availability and maximum transpiration rate
O-CN	Progressive nitrogen limitation as in GDAY, but with lower bound for leaf N:C ratio. Wood and root N:C vary as well as foliage.	Proportional to biomass, plus extra respiration if low N uptake would cause leaf N:C to fall below lower bound.	Stomatal conductance is proportional to GPP. Strong coupling.	Allocation to roots increases with soil moisture stress, following a functional balance approach	β calculated from root-weighted SWC. Applied to g_1 and V_{cmax}
No stoichiometric nutrient limitation					
SDGVM	Leaf N is a monotonically decreasing function of soil C. V_{cmax} and J_{max} are a function of leaf N.	Proportional to biomass and leaf N	Stomatal conductance is proportional to GPP. Strong coupling.	Leaf biomass optimised such that lowest canopy layer has zero C balance.	β calculated from total SWC. Applied to g_1, V_{cmax}, J_{max} and respiration

Table 2: Predicted carbon cycle components for each model, presented as values at aC_a and the change between eC_a and aC_a, for the fixed climate simulations. All values are in g C m^{-2}, either as a standing amount or a cumulative total over the 12 years of the simulations. Initial values of plant carbon (C_{plant}), soil carbon (C_{soil}) and litter carbon (C_{litter}) are the same for both ambient and elevated simulations. Models are ordered according to the change in total NPP with elevated C_a. Abbreviations: NPP, Net Primary Productivity; R_{het}, Heterotrophic Respiration; NEP, Net Ecosystem Productivity.

Ambient C_a	Initial C_{plant}	Initial C_{soil}	Initial C_{litter}	12-yr NPP	ΔC_{plant}	12-yr Litterfall	12-yr R_{het}	ΔC_{soil}	ΔC_{litter}	12-yr NEP
CABLE	12101	8496	1442	10710	742	9969	10133	-92	-71	577
CLM4-P	5064	3267	627	3800	67	3730	3674	40	15	126
OCN	24616	7918	3444	10441	-2064	12506	11675	47	784	-1233
CLM4	10604	8184	1605	7323	-11	7335	7424	-48	-42	-101
LPJ-GUESS	11418	1573	1999	4348	868	3475	3274	48	121	1073
GDAY	18088	4392	643	8350	2770	7207	7302	-71	-23	1355
SDGVM	21211	7732	-	14001	2443	11382	11812	-246	-184	2184

Change with elevated C_a	12-yr NPP	ΔC_{plant}	12-yr Litterfall	12-yr R_{het}	ΔC_{soil}	ΔC_{litter}	12-yr NEP
CABLE	55	23	33	39	-1	-6	16
CLM4-P	104	49	63	59	5	-2	45
OCN	679	-623	1302	1050	33	219	-371
CLM4	765	285	480	316	122	42	449
LPJ-GUESS	1113	682	440	259	60	121	853
GDAY	1695	1265	431	308	63	59	1387
SDGVM	1978	2002	21	-86	38	69	2064

FIGURE CAPTIONS

Figure 1. (a)Annual precipitation in the fixed and variable climate scenarios. (b) Atmospheric CO_2 concentrations applied in the ambient and elevated CO_2 scenarios. Note that years indicate the southern hemisphere growing year, assumed to commence 1st July. For example, 2012 indicates 1/7/12 – 30/6/13.

Figure 2. Annual ambient C_a values of key model outputs in the fixed climate experiment. Green indicates N & P limited models, blue indicates N limited, red indicates no stoichiometric nutrient limitation. Abbreviations: GPP, gross primary productivity; NPP, net primary production; CUE, carbon use efficiency (=NPP/GPP); LAI, leaf area index; E, transpiration; N:C ratio, foliar nitrogen:carbon ratio; N_{up}, nitrogen uptake; P_{up}, phosphorus uptake. Dashed horizontal lines indicate site estimates given to modellers (see Supplementary S1).

Figure 3. Simulated responses of key model outputs to elevated C_a (eC_a) in the fixed climate experiment. Abbreviations as in Figure 2.

Figure 4. Time series of key model outputs in variable climate experiment, expressed as anomalies from outputs in fixed climate experiment. The three low-rainfall years are indicated by grey lines. Abbreviations as in Figure 2, plus SWC, soil water content.

Figure 5: Reduction in gas exchange (β) with soil moisture content in the fixed climate experiment. See Table 1 for explanation of how β was applied in each model.

Figure 6: Summertime (Dec-Feb) average SWC vs. LAI in the (a) wettest and (b) driest years. Ambient C_a values are shown as squares, elevated C_a as circles. The change from ambient to elevated C_a for each model indicates whether that model simulates a decrease in SWC, an increase in LAI, or neither, in response to eC_a.

Figure 7: Percentage increase in (a) GPP and (b) NPP as a function of annual rainfall in the variable climate simulations. Lines indicate statistically significant regressions

($p < 0.05$). Note that two very high values (one for CLM4, one for LPJ-GUESS) have been clipped from (b), but were included in regressions.

Figure 8: Comparison of drought and eC_a effects for the three low-rainfall years, for GPP (top row) and NPP (bottom row). Grey points: Ratio of ambient variable (AV) to ambient fixed (AF) simulations for that year. Black points: Ratio of elevated variable (EV) to ambient fixed (AF) simulations.

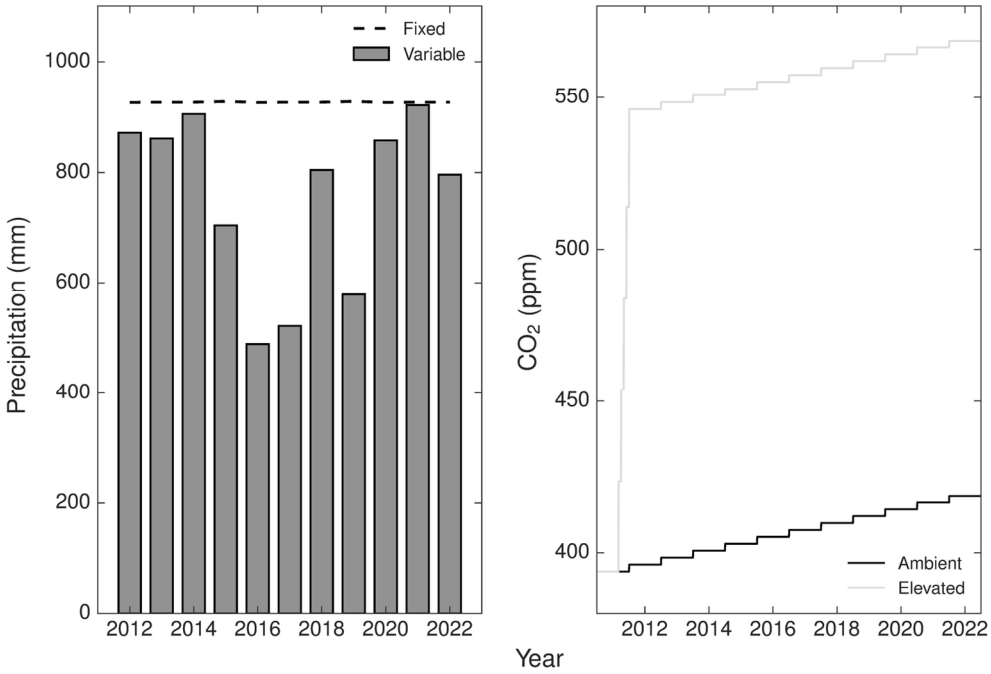

Figure 1. (a)Annual precipitation in the fixed and variable climate scenarios. (b) Atmospheric CO2 concentrations applied in the ambient and elevated CO2 scenarios. Note that years indicate the southern hemisphere growing year, assumed to commence 1st July. For example, 2012 indicates 1/7/12 – 30/6/13.
138x95mm (300 x 300 DPI)

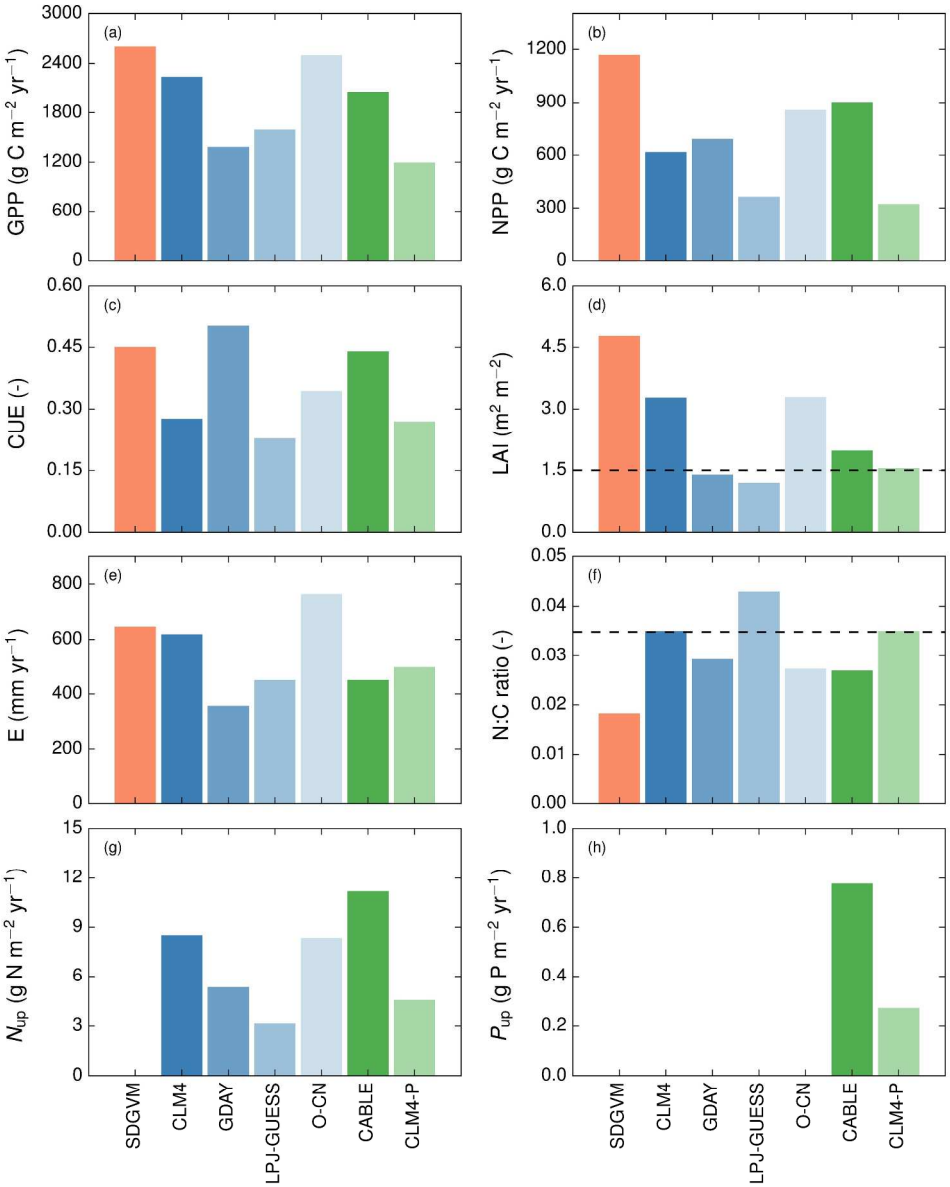

Figure 2. Average annual ambient Ca values of key model outputs in the fixed climate experiment. Green indicates N & P limited models, blue indicates N limited, red indicates no stoichiometric nutrient limitation. Abbreviations: GPP, gross primary productivity; NPP, net primary production; CUE, carbon use efficiency (=NPP/GPP); LAI, leaf area index; E, transpiration; N:C ratio, foliar nitrogen:carbon ratio; Nup, nitrogen uptake; Pup, phosphorus uptake. Dashed horizontal lines indicate site estimates given to modellers (see Supplementary S1).
275x344mm (300 x 300 DPI)

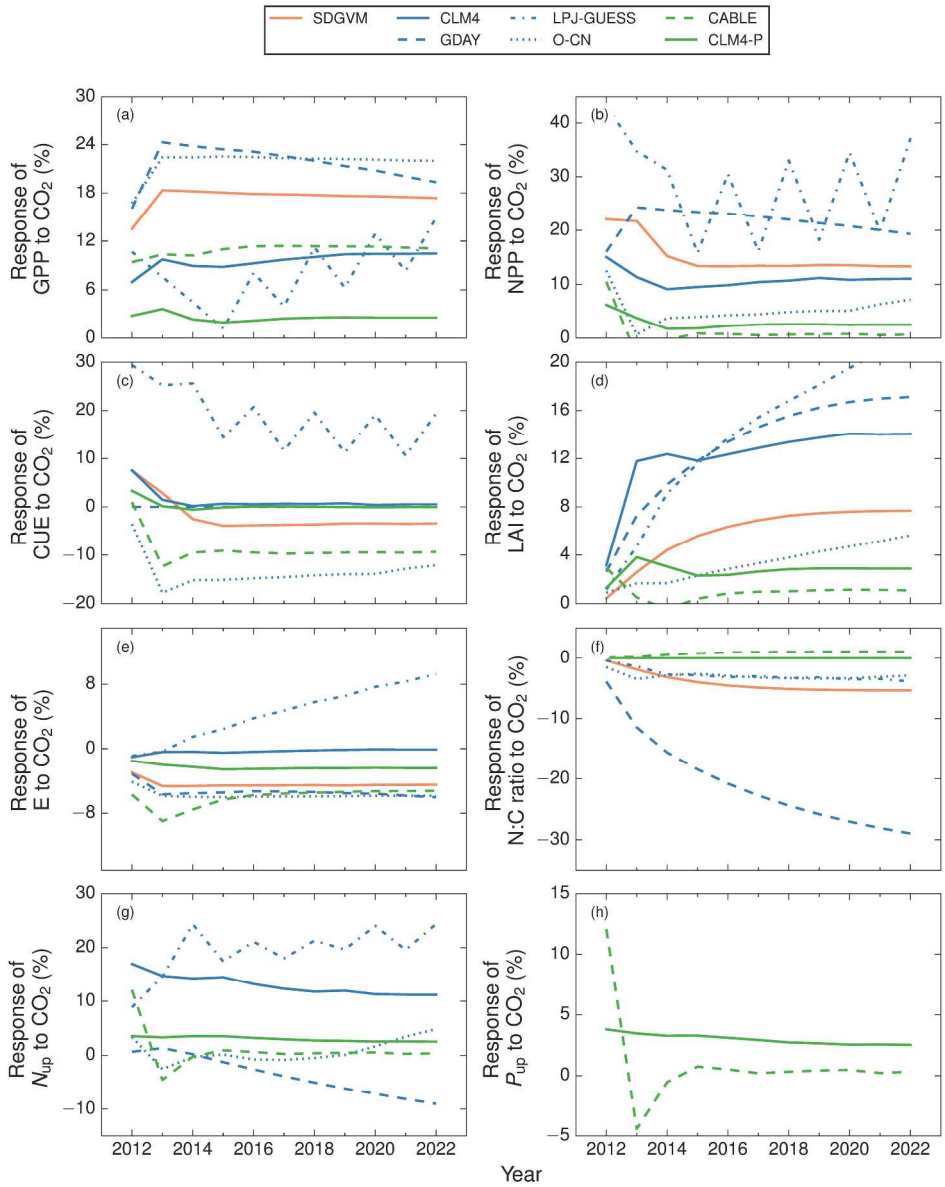

Figure 3. Simulated responses of key model outputs to elevated Ca (eCa) in the fixed climate experiment. Abbreviations as in Figure 2.
281x355mm (300 x 300 DPI)

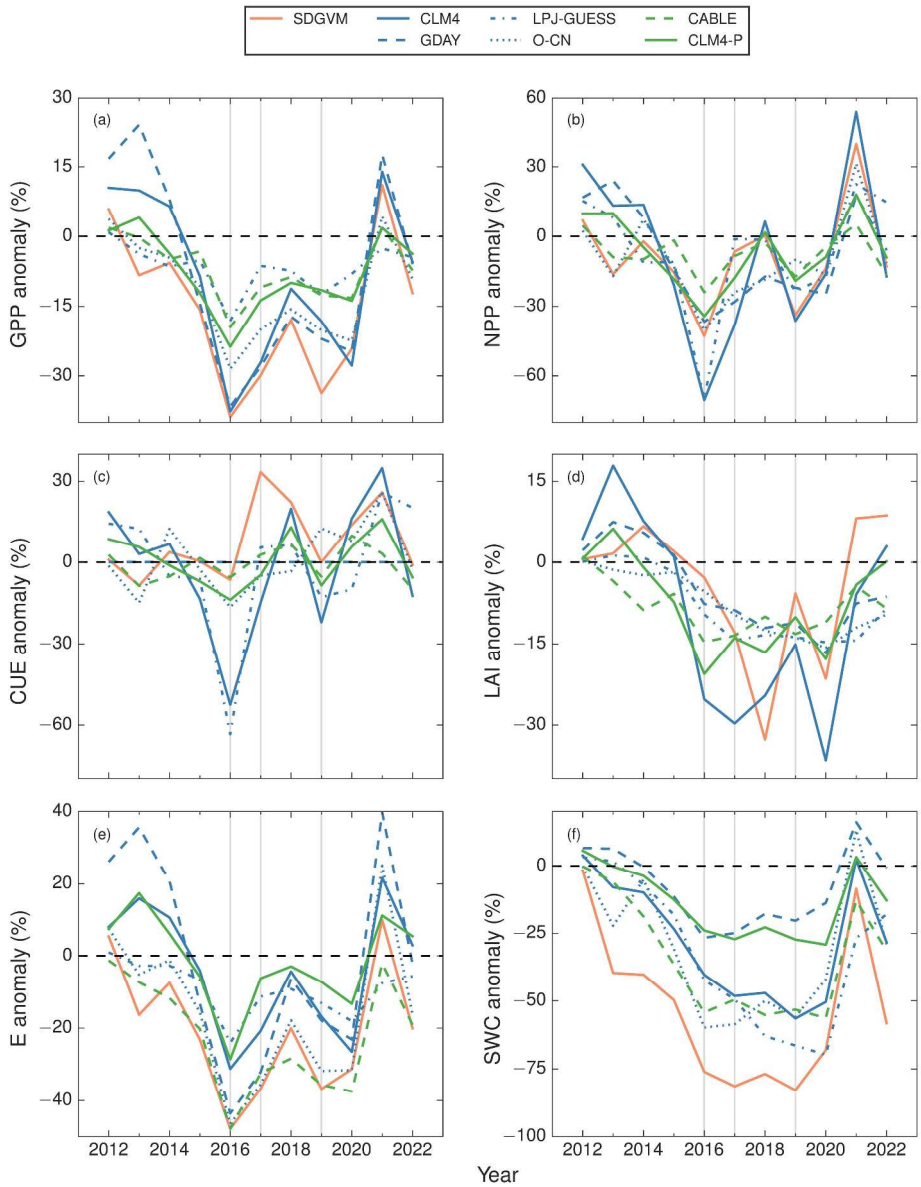

Figure 4. Time series of key model outputs in variable climate experiment, expressed as anomalies from outputs in fixed climate experiment. The three low-rainfall years are indicated by grey lines. Abbreviations as in Figure 2, plus SWC, soil water content.
281x365mm (300 x 300 DPI)

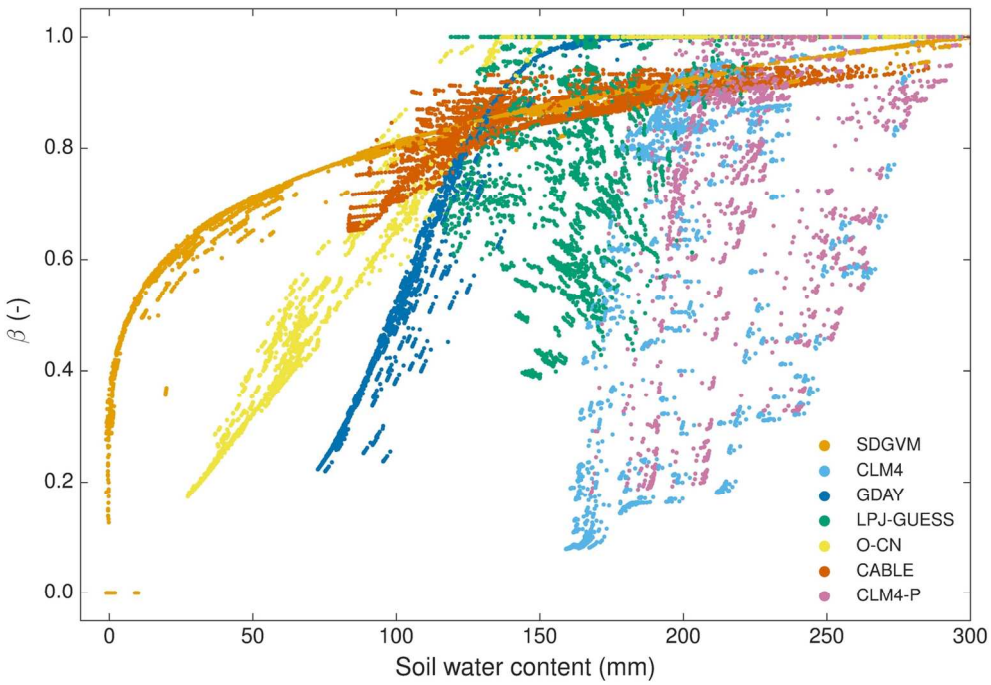

Figure 5: Reduction in gas exchange (β) with soil moisture content in the fixed climate experiment. See Table 1 for explanation of how β was applied in each model.
138x96mm (300 x 300 DPI)

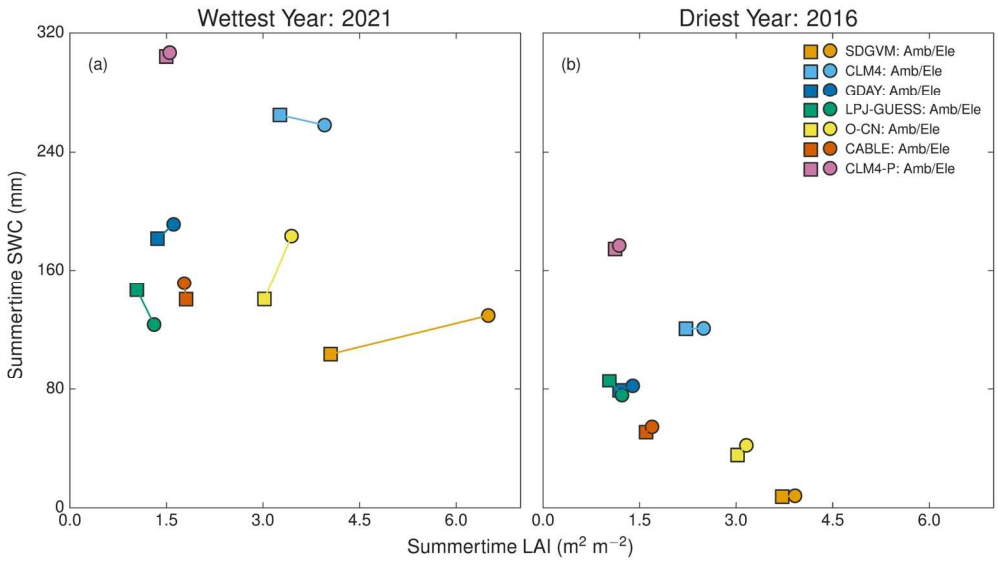

Figure 6: Summertime (Dec-Feb) average SWC vs. LAI in the (a) wettest and (b) driest years. Ambient Ca values are shown as squares, elevated Ca as circles. The change from ambient to elevated Ca for each model indicates whether that model simulates a decrease in SWC, an increase in LAI, or neither, in response to eCa.
145x82mm (300 x 300 DPI)

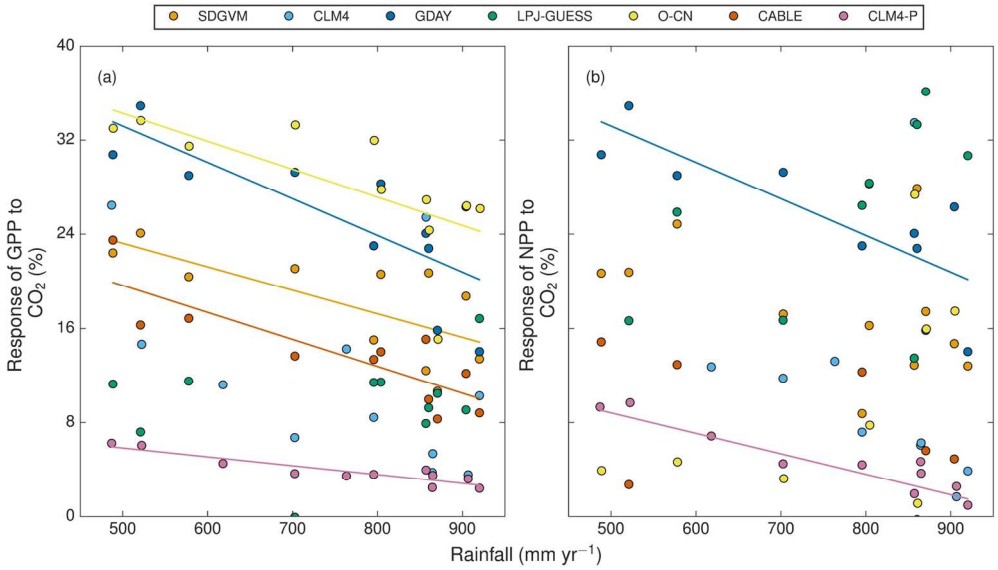

Figure 7: Percentage increase in (a) GPP and (b) NPP as a function of annual rainfall in the variable climate simulations. Lines indicate statistically significant regressions ($p < 0.05$). Note that two very high values (one for CLM4, one for LPJ-GUESS) have been clipped from (b), but were included in regressions.
149x86mm (300 x 300 DPI)

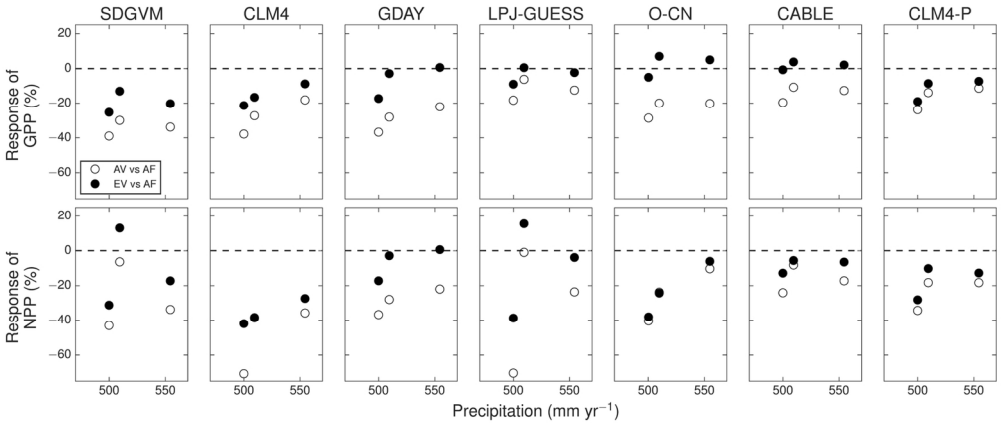

Figure 8: Comparison of drought and eCa effects for the three low-rainfall years, for GPP (top row) and NPP (bottom row). Grey points: Ratio of ambient variable (AV) to ambient fixed (AF) simulations for that year. Black points: Ratio of elevated variable (EV) to ambient fixed (AF) simulations.
146x62mm (300 x 300 DPI)