Dear Dr. Raoult,

The two reviewers are generally supportive of the research presented in your manuscript. Your responses to the reviewers' comments and suggested edits are mostly convincing. However, I have some reservation about the following points raised by the reviewers (and some of my own) and your replies:
Thank you for taking the time to read the manuscript, the reviews and our responses. We have addressed your following concerns below.

Some fundamentals of the methods are unclear or could be mis-understood:

- Eq. 1 is in a matrix-form, but covariances are not considered. Can Eq. 1 be provided in a simplified form that also clarifies the weighing of error terms from evaluating against FACE and FLUXNET data separately?
The equation as given in the text is the notation commonly used in our the field and although we do not consider covariances here, in future, this is something we would like to move towards. Nevertheless, we agree that in this case it is also helpful including a second equation showing the decomposition of terms obtained through a diagonal R matrix:

> … For **R**, we defined the observation error (variance) as the mean-squared difference between the observations and the prior model simulation so that this variance reflects not only the measurement errors but also the model errors. **Furthermore, since we do not consider error covariances, R is diagonal and therefore we can decompose the first term of Eq. 1 into different terms for each assimilated datastream:**
>
> $J(\mathbf{x}) = \mathbf{k_{Flx}}*(M_{Flx}(\mathbf{x}) - \mathbf{y}_{Flx})^T\mathbf{(1/\sigma}_{Flx}\mathbf{)}(M_{Flx}(\mathbf{x})-\mathbf{y}_{Flx}) + \mathbf{k_{FACE}}*(M_{FACE}(\mathbf{x}) - \mathbf{y}_{FACE})^T\mathbf{(1/\sigma}_{FACE}\mathbf{)}(M_{FACE}(\mathbf{x})-\mathbf{y}_{FACE}) + (\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)$
>
> **where Flx and FACE subscripts are used to denote the FLUXNET and FACE parts of the equation; k_i denotes the weighting using for each datastream, σ_i denotes the observational error, and M_i and y_i denote modelled and observed data streams.**

We add also added links to k_{Flx} and k_{FACE} in Sect. 2.4. (e.g. L232).

- Your use of the term 'data assimilation' is generous. I feel like 'parameter optimization' is a more appropriate term to be used here, since you are not rigorously estimating prediction uncertainties (you write that this manuscript is a "proof of concept study" in your reply to reviewer 2), and it is unclear how observation errors are treated.
The use of "Data Assimilation" for parameter optimisation is common our field of study, and the inclusion of the KSoil parameter allows us for perform a station optimisation of the initial soil C poosl. Nevertheless, we have softened the language and used parameter estimation instead of the data assimilation for a better clarity and accessibility.

- The calibrated parameters are insufficiently described, no equations where and how they are used are provided, and their units are not specified in Table 1.

We have added the units to Table 1 and equations in Appendix. We have added the following to the text to relax this:

> These parameters represent key parameters of the model controlling photosynthesis, carbon and nitrogen allocation, respiration and global nitrogen cycle behaviour **(full descriptions can be found in Appendix A)**.

In the Appendix:

> All processes and equations of ORCHIDEE can be found in the different documenting publications (e.g., Krinner et al., 2005), as well as on its website (https://forge.ipsl.jussieu.fr/orchidee/wiki/Documentation/, last accessed: 01/09/23). Here we only highlight the impacted modules, summarise the equations in which the optimised parameters feature and cite the relevant publication. Parameters optimised in the study, listed in Table 1, are coloured red in the following text.
>
> **Nitrogen-related processes**
>
> The nitrogen-related parameters and their equations are thoroughly described in Vuichard et al., (2019). In this version of ORCHIDEE-CN, we prescribe leaf nitrogen concentrations. This means that leaf C/N ratio is fixed within a prescribed range given by two of our parameters ([CNleaf, min; CNleaf, max]; $g_C[g_N]^{-1}$). To account for the nitrogen limitation on photosynthetic activity, VCmax (the maximum rate of Rubisco-activity-limited carboxylation) becomes a function of the leaf nitrogen content (N_l) as proposed by Kattge et al., (2009):
>
> $$VCmax = NUEopt*N_l$$
>
> where NUEopt (μmol $CO_2 s^{-1}$ $[g_{Nleaf}]^{-1}$) is the nitrogen-use efficiency.
>
> N_l decreases exponentially from the top to the bottom of the canopy with decreasing light levels or increasing canopy depth. The value of N_l at the top of the canopy, $N_l(0)$, is expressed as a function of total canopy N content, N_{tot}, and the LAI of the total canopy, L_{tot}:
>
> $$N_l(0) = (k_N N_{tot})/[1- \exp(-k_N L_{tot})]$$
>
> where k_N is a specific extinction coefficient. Note that this is different to the extinction coefficient k used to calculate the light profile within the canopy, although both are optimised. As we decrease through the canopy, the value of N_l at a cumulative LAI (L) is defined following Dewar et al., (2012):
>
> $$N_l(L) = N_l(0)*\exp(-k_N L).$$
>
> It is assumed that N_l varies through the canopy due to variations in specific leaf area (SLA; i.e., leaf area divided by leaf mass; $m^2 g^{-1}$), instead of variations in leaf nitrogen concentration which is kept constant. The SLA at the bottom of the canopy (SLAinit) is fixed and also optimised.

The model calculates the nitrogen required (GN_{init}, $gm^{-2}d^{-1}$) to satisfy the new carbon (G_C ($gm^{-2}d^{-1}$) to the different reservoirs under the assumption that CN_{leaf} does not vary (Zaehle and Friend, 2010).

$$GN_{init} = (FCN_l/CN_{leaf} + FCN_{root}/CN_{root} + FCN_f/CN_{fruit} + FCN_{wood}/CN_{sap})G_C$$

where FCN_i represents the fractions (unitless) of carbon allocated to leaf (l), roots (root), fruit (f) and sapwood or stalks (wood) and CN_i represents the C/N ratios for the different biomass pools at the previous time step. FCN_{root} and FCN_{wood} are optimised in this study. R_{leaf} and R_{root} are the fractions of N retranslocated when shedding leaves and roots (f_{trans} parameter in Zaehle & Friend, 2010). CTE_{bact} is a parameter relating denitrifier bacteria activity to Soil Organic Matter.

Root density follows an exponential profile, with more roots located in the top soil layers. The root density profile parameter z defines the depth above which ~65% of roots are stored and use the calculate plant moisture availability (Krinner et al. (2005); Eq A18). Finally $VMAX_{UPTAKE}$ is used to calculate plant N uptake (Zaehle et al., 2010; supplementary material Eq 8)

Allocation

Allocation in ORCHIDEE-CN follows the formalisms of the OCN model (Zaehle and Friends, 2010), further described in Naudts et al., (2015), and respects the pipe model theory (Shinozaki et al., 1964). K_{latosa} (whose range $K_{latosa,min}$,$K_{latosa,max}$ is calibrated) is used to derive a scaling factor between leaf and sapwood mass:

$$d_1 = K_{latosa} \times m_w \times d_s$$

where d_l is the one-sided leaf area of an individual plant, d_s is the sapwood cross-section area of an individual plant and m_w is the water stress. Sapwood mass (M_s) and root mass (M_r) are related as follows (following Magnani et al., 2000):

$$M_s = k_{sar} \times d_h \times M_r$$

where the parameter k_{sar} is calculated:

$$k_{sar} \sqrt{(k_{root}/k_{sap})\times(k_{TS}/k_{TT})\times k_{ps}}$$

where k_{root} is the hydraulic conductivity of roots, k_{sap} is the hydraulic conductivity of sapwood, k_{TS} is the longevity of sapwood and k_{TT} is the root longevity, and k_{ps} is the sapwood density.

Phenology

For the phenology parameters we mostly refer to MacBean et al., (2015). The photosynthetic efficiency of leaves depends on their age L_{age}. Using four separate age classes, biomass newly allocated to leaves goes into the first age class and leaf biomass, B_l, is then transferred from one class to the next based on a PFT-specific critical leaf age value, $L_{agecrit}$. In temperate deciduous broadleaf forests, leaf senescence is triggered when the monthly air surface temperature goes below a threshold temperature:

$$T_{threshold} = T_{senes} + C_1T_l + C_2T_l^2$$

where T_l is the long-term mean annual air surface temperature and T_{senes}, C_1 and C_2 are PFT-dependent parameters. Once senescence has begun, a fixed turnover rate is applied, with trees losing their fine roots at the same rate as their leaves

$$\Delta B = B.\Delta t/L_{fall}$$

where Δt is the daily time step, B is the total biomass and L_{fall} is optimised.

Photosynthesis

Stomatal conductance (g_s) is coupled to leaf photosynthesis by the following equation:

$$g_s = g_0 + \frac{A + R_d}{C_i - C_{i^*}} f_{VPD}$$

where g_0 is the residual stomatal conductance when irradiance approaches zero, A (μ mol CO_2 $m^{-2}s^{-1}$) is the net assimilation rate, C_i (mol $CO_2 m^{-2}$) is the intercellular CO_2 partial pressure, C_{i^*} is the C_i-based CO_2 compensation point in the absence of respiration (R_d) and f_{VPD} is the function for the approximal effect of leaf-to-air vapour pressure difference (VPD, kPa):

$$f_{VPD} = \frac{1}{1/(A_1 - B_1 VPD) - 1}$$

The empirical factors A_1 (unitless) and B_1 (kPa^{-1}) are optimised in this work.

Respiration

Q_{10} (unitless) used to calculate the temperature control of heterotrophic respiration:

cT=min(1 ,Q10^((T-30)/10)))

where T is the surface/soil temperature for the above/below-ground pools.

The growth respiration is calculated as a fraction of the remaining total biomass:

Rg = $FRAC_{growthresp}$ ·max (B_a −\delta t ·\sum R_m,i 0.2·B_a)

where B_a is the total biomass, 1t the time step (one day), and $FRAC_{growthresp}$ a fraction to be optimised.

- Is the model calibration done starting always with the same spinup?

Yes. This has been clarified in the text:

Once spun up, we performed two main sets of optimisation **always starting from this spinup**

- Revise terminology: You write in several places that “parameters showing sensitivity to outputs” (l. 128). However, the Morris sensitivity analysis quantifies the sensitivity of an output variable to a certain parameter (not vice versa), given that the parameter varies within the constrained range.

We agree that this is somehow misleading and we have revised the line as follows:

we removed all parameters **to which** the different modelled outputs tested (i.e., net primary product (NPP) and leaf-area index (LAI)) **showed no sensitivity**. All remaining parameters were optimised in this study

- It is unclear what has changed in the model since Vuichard et al., 2019.

The version of ORCHIDEE-CN (r4999) described and evaluated in Vuichard et al. (2019) was based on a version of ORCHIDEE without the nitrogen cycle (trunk version r3977) which was anterior (by ~one year) to the one used for the CMIP6 exercise (ORCHIDEE_2.0, trunk version r5107). The version of ORCHIDEE_3 (branch ORCHIDEE_3 r6863) used in this study is based on ORCHIDEE-CN (r4999) but has been updated for latest developments of ORCHIDEE_2.0 regarding small ug corrections. In addition, few specific N-related process modelling has been

updated in ORCIDEE_3 (r6863) in particular growth and maintenance respiration modelling. The following has been added to L121:

> The version of ORCHIDEE used in our study (ORCHIDEEv3, r6863) is more recent than the one used Vuichard et al. (2019**, r4999**). **ORCHIDEEv3 (r6863) includes the latest developments of the main ORCHIDEE model (mainly small bug fixes). Furthermore, it includes updates to a few specific N-related processes, notably growth and maintenance respiration. Although this version** has been used in the multi-model ensemble for the Global Carbon Budget 2020 (Friedlingstein et al., 2022), **it** has not yet been optimised against independent data. As such, the initial fit of the model to the Fluxnet data is different than that shown in Vuichard et al. (2019).

Interpretation of the calibration setup in the context of model improvement, applicability, and generalisability:

- You write that “By optimising the model against a number of different constraints, we gain confidence in our parameter and hence in the projections”. Note that a parameter optimisation, by design, always improves the model-observation fit, yet the model could be overfitted and perform poor on data that was not used for calibration. Hence, interpreting the value of this exercise as a means to “gain in confidence” is not generally justified. Indeed, the optimisation deteriorated some aspects of the fit, as documented here. Therefore, the discussion should address the generalisability of the calibrated model and potential of overfitting, in particular in view of site/experiment peculiarities (see comments by Martin DeKauwe).

We agree that this sentence can be a bit misleading. The point we were trying to make was that by using different constraints (data streams) we decrease the risk of overfitting and thus increase the chance of gaining confidence in the projections. Indeed, there will always be a risk of overfitting. Our study provides a way to expand the assimilation of classic observations (Fluxnet, Satellite phenology, temperature, soil moisture,...) with some observations under different CO_2 levels, which should hopefully decrease the risk of overfitting. We have expanded the text in the following manner:

> “However, we do need to be cautious in assessing these results since we are only using one FACE site for each PFT meaning we are likely tuning to the specificities of that site. For example, ORNL shows a progressive nitrogen limitation but this is not expected over all sites. Ideally, we would include a lot more FACE sites to capture different conditions. Especially, if we could optimise by grouping sites based on different levels of nitrogen limitation, then if the posterior parameters were found to be similar then the model processes allow for these differences.
>
> **In any optimisation, there is always a danger of overfitting to data limiting the generalisability of the calibrated model.** By optimising the model against a number of different constraints **(i.e. more than one data stream)**, we **decrease the risk of overfitting and therefore,** gain **some** confidence in our parameters and hence in the projections.”

Interpretation of the main results:

- The finding “strength of the CO2 fertilization effect changes depending on the type of forest considered” is difficult to understand. What makes the two forest types investigated here (TeNE and TeBS) different in their C-N coupling and response to CO2? To what extent are specific information about experimental sites and setups underlying your results? Are different simulated responses to eCO2 at the two sites due to the parameter values or initial conditions? If ORNL parameters were used for the Duke simulations, would we also get a declining response ratio?
The CO_2 fertilisation is slightly different after the optimisation between the sites. This is probably due to the differences in site history, especially the soil fertility. Furthermore, the types of forests have very different functions (broadleaf vs needleleaf) most likely impacting their CO2 sensitivity. This difference in functions (both for the leaf type (broadleaf and needleleaf) and the leaf seasonality (summergreen vs evergreen)) means that testing the ORNL parameters at the DUKE site won’t make much sense.However, since it is hard to disentangle the different responses here, we have decided to drop the sentence to avoid confusion.

- I feel like there is a somewhat selective reporting of results that leads readers to conclude that previously unconsidered FACE observations imply a smaller response of GPP to CO2 than when only FLUXNET data is used (You write: “we find that the rate of CO2 fertilisation is much lower when Free Air CO2 Enrichment data has been in the optimisation.”). However, the response of GPP to CO2 that considered FACE observations in the calibration yields the same response as the model with prior parameter values (which excluded FACE data from the calibration). Please adjust the text to reflect this aspect.
We apologise for this confusing statement. This was alluding to the fact that the Fluxnet optimisation (blue) and the Fluxnet+FACE optimisation (orange) gave different GPP responses in the Figure 6. The prior model was not included in this statement, since this new version of the ORCHIDEE model had never been formally (Bayesian approach) calibrated. Indeed, the prior values changed a lot during our study period based on different manual tuning experiments. Nevertheless, we agree that the statement is too strong and can be misleading, especially in the abstract. Therefore the sentence has expanded as follows:

> “Using an idealised simulation experiment based on increasing atmospheric CO2 by 1% per year over 100 years, we find that **optimising against only FluxNet data tends to imply a large fertilisation effect whereas optimising against FluxNet and FACE data (with all nutrients limitation and acclimation of plant) decrease it significantly**.”

- Negative response ratio of NPP simulated for Duke: The explanation provided doesn’t seem to fit. An increasing autotrophic respiration (decreasing biomass production efficiency) should be expected if leaf N increases, not decreases.
We have changed autotrophic respiration to maintenance respiration in the text to be more correct, since, in the model, maintenance respiration is a function of leaf N but not growth respiration.

Furthermore, the manuscript needs careful editing to resolve, among others, the following points:

- Make sure that the use of tenses is consistent throughout the manuscript.
Done

- Since you don't have a Discussion section and since "discussion points" are wrapped within the Results section, I suggest to re-name Section 3 as "3. Results and Discussion".
Done

- l. 54 "One such experiment, the Free Air CO2 Enrichment experiment" - 'FACE' refers to an experimental setup and technique, not a particular experiment per se. There have been a number of FACE experiments conducted.
Changed to:

> One such **type of manipulation** experiment, the Free Air CO$_2$ Enrichment experiment**s** provide

- l. 55: "FACE experiments are conducted across several 55 vegetation types and typically consist of two plots" - Usually, there are more then one plot under each, the treatment and control.
Changed to:

> FACE experiments are conducted across several vegetation types and typically consist of two **types** of plots: one where CO is fumigated to high concentrations and one left as a control.

I am therefore returning the paper to you so that you can make the necessary (major) revisions and I am looking forward to receiving your revised manuscript.

Beni Stocker