Reply rev 1

Dear reviewer,

We would like to thank you for the constructive comments on our manuscript. We have revised the manuscript to accommodate all your concerns. Below, we have replied to all your comments in detail providing concrete reference to the sections that have been changed. We have also marked the main revised parts of the manuscript in yellow.

We hope that this new version accommodates all your concerns.

Comments

(1) The language needs a thorough revision. There are numerous grammar errors, typos and spelling errors. Additionally, there are obvious flaws in several sentences which are either incomplete or include remnants of earlier versions.

We have thoroughly revised the paper. If we revised the paper according to the comments of the second reviewer we will check the language again.

(2) While the general structure of the model description part is adequate, its implementation is not consistent. The model structure section includes also processes. The water processes in the process section are weakly described.

We have cleaned the structure and renamed the chapter into ”model overview”. The water processes are described shortly in 2.1.2.5 as part of a quick model overview and in the detailed description (chapter 6).

(3) The model structure section needs a thorough revision in its own. The basic and essential structural features of the model needs to be described clearly and unambigously. For instance, the role of the tree cohorts is never clearly described. There are some statements about the inter-cohort competition. But nothing is said about intra-cohort relationships of trees. Area size of the forest that can be simulated is also not clear.
We have reformulated the section and added information about intra-cohort relationships. We now write:
“It can be applied for patches of various sizes (varying from 100 m² to several hectares) and mono- and mixed-species forests.”
and
“All trees within a cohort share the same characteristics which are species, age, tree dimensions (height, height of crown base (or bole height), and diameter at breast height), biomass differentiated into various compartments (foliage, fine roots, sapwood, and heartwood) and stage of phenological development. This allows simulating a representative tree of each cohort instead of each tree of the stand. The model is distance independent and so the trees within a cohort are horizontally evenly distributed and their position unknown. There are no differences in the growth behaviour of the trees of a cohort and there is no competition between the trees within a cohort. In contrast, the tree cohorts compete for light, water and nutrients.”

(4) In a manuscript like this one I expected an overview on past model applications and related literature. In Table 2 past evaluation experiments are listed, but no general overview about 4C-related literature is provided.

We added in chapter 2.2. (Previous model evaluations and applications) a short description of past applications and in the ESM a detailed table with an overview of past 4C applications.

(5) Main part of the manuscript is a new evaluation study that includes four sites across Europe. What has then been gained from these new experiments compared to earlier evaluation studies is never clearly presented.

We inserted a reason for the use of PROUFOUND data for our evaluation in chapter 2.3:

"Within PROFOUND, several other forest models are using the same data so that comparisons between those models are possible."
Therefore, we decided to use these data for this evaluation paper.

(6) The results section is quite long and includes too many graphs and tables. Some of the material could be moved to annex or supplementary material. In general there is the tendency, that without much context the results for individual stand and tree attributes are listed in no obvious order. Summarizing, it is very demanding to get the essential information from the vast amount of numbers.
Thank you for this advice. We shifted some figures to the supplement (the time series in figure 3-5, Figure 7), and some additional for the final revision. Former evaluations based on similar data /Reyer etal 2014) were not extensive or consider only fluxes. We think that the evaluation based on all available data allows identifying the deficits or benefits of the model.

(7) In Figures 3 and 4 the initial values for observed biomass and simulated biomass show a rather huge mismatch. Please explain why. Labels of y-axes are also not consistent. Please check.

The missing match for the initial values is explained by the initialization process. The model used averaged data of DBH, H and stem numbers to generate a stand with these characteristics. However, the initialization does not exactly match the DBH, H and biomass. In chapter 4.1 we tried to discuss this problem. Only if single tree data are available a better correspondence between data and model in the initialization is possible.

We used different scales for biomass and diameter for the two stands Peitz and Solling because of the very clear differences in the amounts of these values between the stand caused mainly by the species (pine versus spruce) and site conditions. Peitz is a very poor site with dry condition whereas Solling is a wet site with better soil conditions.

(8) Tables 7, 8 and 9 are not well structured. How sites are depicted is not consistent in formats.

We depicted the sites with available flux data (Hyytiälä and Soroe) and have checked that he formatting and layout of tables is consistent. We try to combine Tbale 6 and 7 but it leads to confusing and large table.

(9) Correlating time series data of simulated and observed stand level attributes will usually produce impressive R2 values without having too much meaning

Yes, we agree and yet it is a widely used model evaluation metric. In our examples you see clear difference between annual, daily and monthly correlation analysis. However, exactly because of the limitations of correlations in mind, R2 is not the only criteria for comparison we are using. We are also using other criteria, especially the model efficiency.

(10) The discussion deals mainly with the new evaluation experiment. No in-depth well-founded linkages to other earlier studies are provided.

This is correct because the focus of our paper was the evaluation of the recent model version with the PROFOUND data. However, we mentioned the paper by Reyer et al. (2014), which provide a detailed model evaluation) and in 4.2 (Evaluation of carbon and water fluxes and 4.3 (Evaluation of soil water content and soil temperature) we inserted some remarks about how these 4C results compare to earlier evaluation studies..

(11) There are not that many models available that are applicable over a broad range of tree species and site conditions that could be used for climate change impact studies. Thus, provision of the context to other similar models would definitely improve the manuscript.

We appreciate the reviewer noticing the assets of 4C being and available and applicable over a wide range of environmental conditions
The main focus of the paper was evaluation of 4C. The comparison or description in the context to other similar models is planned in other ongoing papers.

(12) The conclusions are generally positive despite partly rather weak results of the new evaluation experiments. This may leave an interested reader also quite puzzled.

We revised the conclusion

(13) The abstract should be revised, too, based on an improved version of the manuscript.

We revised the abstract.

Further comments:

L 45: corrected

l 77: features described in the above mentioned papers

l 86: in the very extensive model description

l 88: model structure → Overview

l 96: all elements in a cohort are identical, there nothing happens in a cohort, but the number of elements can be reduces by mortality

l 105: see new title of this chapter

l 111: replaced

l 112: revised

l 115: the top line is not clear! Please explain. We explained it

l 151: revised

l 152: heartwood is the correct name, the basal area of a tree is divided into sapwood area and heartwood area

l 161: reference inserted

L 181: revised

L 188: water content and soil temperature

L 195: revised in chapter 2.1.1.

L 213: revised

L 217: revised

L 224: reviewer: take care when using this form:

Nature recommends the followings https://www.nature.com/nature-research/for-authors/write

"Nature journals prefer authors to write in the active voice ("we performed the experiment...") as experience has shown that readers find concepts and results to be conveyed more clearly if written directly. We have also found that use of several adjectives to qualify one noun in highly technical language can be confusing to readers. We encourage authors to "unpackage" concepts and to present their findings and conclusions in simply constructed sentences."

L 227: revised

L 270 reviewer: not necessarily. that the smallest model time step is daily doesnt mean that you rely on daily climate input.
We revised it
The references indicate the sources of the CO_2 scenarios for the RCP climate scenarios, which are not used in this study but available in the model.

L 296: revised

L320 revised with explanation of PROFOUND

L 331: revised

L 334: arithmetic mean/ average and geometric mean are clear mathematical terms; stem biomass is sapwood plus heartwood

L 347: reviewer: why cant you say monthly? i.e. the monthly variation.????
We think that inter-monthly is more comprehensible.

L 359 reviewer: what is the relevance of this statement. there will always be a site where the fit is better than at another site.
Here, we compared Peitz and Solling.

L 360 reviewer: worse; not as good as, we revised it.

L 365: revised

L 367: the ME values do not indicate poor results. the results are actually poor.
The statistical measures indicate the quality of the results.

L 368 the quality of model results is measured by ME, which indicate poor or good model results in comparison to measurements

L 369 reviewer: not a good idea to regress time series data of accumulating variables. will always show high R2 values. you should better use differences (i.e. growth).

Annual differences in biomass or diameter from simulation results are influenced by mortality, which differs clearly from the observation data. We intended analyzing the times series of diameter and biomass and not the increments.

Similar analyses were shown in:

Seidl, R., Lexer, M. J., Jager, D., and Honninger, K.: Evaluating the accuracy and generality of a hybrid patch model, Tree Physiology, 25, 939-951, 2005.
Miehle, P., Battaglia, M., Sands, P. J., Forrester, D. I., Feikema, P. M., Livesley, S. J., Morris, J. D., and Arndt, S. K.: A comparison of four process-based models and a statistical regression model to predict growth of Eucalyptus globulus plantations, Ecological Modelling, 220, 734-746, 2009.
Grote, R., Kiese, R., Grünwald, T., Ourcival, J.-M., and Granier, A.: Modelling forest carbon balances considering tree mortality and removal, Agricultural and Forest Meteorology, 151, 179-190, 10.1016/j.agrformet.2010.10.002, 2011.

L 391: revised

L472: revised

L483: well, not very convincing model test.

This is not a model test, it should only show, that the initialization is reasonable

L 492: revised

L 496: revised

L 500: revised

L 629 reviewer: what is sufficient accuracy? you need to define and explain your definition.
We deleted this expression.

L 665 reviewer: in this form this sentence is not very usefull.

We revised it:

Using the PROFOUND database enables the comparison with simulation results of other process-based models using the same database. This includes the potential to gain new insights into the process understanding of forest growth on the base of such model-intercomparisons.