

Dear Prof. Neale,

We would like to thank you for the time invested in handling our manuscript and for your constructive attitude towards it. We have replied point-by-point to all concerns raised by the reviewers (see below) and all changes are made visible in the manuscript with yellow. To summarize, the main changes are:

- 1) We have revised the figures along the lines suggested by you and the reviewers.
- 2) We have added a new section (2.3) explaining in which way 4C is different from other models
- 3) We have clarified the target audience early in the text (abstract L30, L43, introduction L95-101, and in each subsection of the discussion we now also provide detailed recommendations on how to best use the model in the form of recommendations (see e.g. section 4.1, 4.2, 4.3). We have also added a FAQ table (Table 7) for prospective model users. The table is based on the comments/questions of reviewer 2 and added a few more. This FAQ table is rather long and although we think it is well placed within the manuscript, we leave the decision whether it should be part of the manuscript or being moved to the Supplement to you.
- 4) We have completely rewritten the conclusion.

We think that this manuscript version has greatly improved and a really grateful for the constructive criticism from you and the reviewers. We hope you agree and find this version of the manuscript acceptable for publication in GMD.

On behalf of all co-authors

Petra Lasch-Born

Detailed replies to reviewers:

Dear reviewers,

We would like to thank you for the constructive comments on our manuscript and the possibility to revise our paper again. We have now uploaded a revised manuscript to accommodate all concerns the reviewers 2 and 3 have articulated. Below, we reply to all 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.

On behalf of all co-authors

Petra Lasch-Born

### Anonymous Referee #3

#### **Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)**

I was invited as additional referee for the manuscript “Description and evaluation of the process-based forest model 4C v2.2 at four European forest sites”. The previous manuscript has been revised extensively, and, in my opinion, the authors have incorporated most of the referees' feedback satisfactorily, justifying publication of the manuscript in Geoscientific Model Development after another minor revision.

**Reply: We are grateful for the time and effort invested by the reviewer and his/her constructive attitude towards our manuscript.**

One issue that has already been pointed out by a different referee, and yet remains vague, is the novelty statement of the manuscript. A study that claims to improve our ability to model forest ecosystems should be very precise about its novelty. I would expect a table or, at least, longer paragraph comparing the new 4C version with similar models.

For instance, you should answer the question: What can be simulated with this model that cannot or just to a limited extent be simulated by other models?

**Reply: We agree with the reviewer that the novelty and especially uniqueness of 4C should be stressed further. Therefore, as suggested by the reviewer, we have introduced a new section 2.3 called “In which ways is 4C different from other forest models?”. This section picks up the very useful (although still subjective) classification of models used for climate change projections by Medlyn et al. 2011 and compares 4C with some of the main models of the different classes explaining in which ways the model is different. Please note that we are also working on a follow-up study, comparing 10-15 of commonly used forest models within the framework of the Inter-sectoral Impact Model Intercomparison Project (ISIMIP) which will shed further light on how different these models actually are.**

Moreover, I wondered why there was no simulation and evaluation of species mixtures? Monospecific stands are relatively simple to simulate in comparison to mixed stands where different species traits need to interact in a meaningful way.

**Reply: We fully agree with the reviewer that species mixtures are more challenging to simulate and in Table 1 we present the results of former evaluation experiments for mixed forests. However, although there is a mixed stand available in the PROFOUND Database, unfortunately, there are no carbon/water flux observations available for this site and the growth and yield data are also only available for a rather short period. However, our criteria for including sites where to have long time series and/or data from flux towers available. Since both are not available for the mixed stand in the PROFOUND database, we excluded it from our analysis. We have also clarified that improving the representation of mixed forests beyond 2-3 species mixtures etc. is a key area for improvement in the model (e.g. in the conclusions and the newly added FAQ table Table 7, see above).**

A few minor points:

L 64 omit one parenthesis

**Corrected**

L 78 impact of biotic disturbance agents on what?

**Corrected**

L 80 “partly” what parts?

Corrected

L 123 isn't this a weekly time step that is just divided by seven?

Corrected

L 176-177 doesn't read well. Further, citing the new TRY database paper (Kattge et al. 2020) would be more useful here

R: Thank you, we use the new reference

L 179 one or several studies?

Corrected

L 332 parameters were derived

Corrected

L 464-466 does that mean trees grow better if disturbance is disabled as more structural C is available? That wouldn't make much sense.

Reply: we thank the reviewer for spotting this point. The sentence was misleading so we have rewritten the section in the paper. Because the "disturbance mode" is not used in the simulations presented in this paper, we provide only a general overview of how disturbances are included and the role of the NSC pool. We also referenced here the full model description which gives a more detailed explanation.

Table 1: I agree with the previous referee that quantitative figures would greatly improve interpretation of the results column. We always under- or overestimate something to some degree. Maybe the authors can at least provide ranges where quantifications are still missing.

Reply: We agree with the reviewer and have included quantifications wherever available. Unfortunately (and maybe not understandable from today's point of view anymore) this is not possible for all papers because some of the older papers do not give these quantities. We have clearly indicated in the Table that for some earlier papers no quantitative evaluation is available.

L 517 you mean "we did not re-calibrate any species parameter"?

Reply: Yes, we only used the parameters of 4C as they are, no tuning or whatsoever parameter changes happened.

L 600-602 and elsewhere: provide values for under- and overestimates and correlations in the main text even if they are shown in tables

Reply: we added required values in the result part

L 645 wrong order!? Should read  $0.887 < ME < 0.974$

Corrected

L 723 omit ".".

Corrected

## Anonymous Referee #2

### Suggestions for revision or reasons for rejection (will be published if the paper is accepted for final publication)

I was pleased to see that the authors made a considerable effort to revise the manuscript “Description and evaluation of the process-based forest model 1 4C v2.2 at four European forest sites”. The introduction (including the aim of the manuscript), and the methods (including the model description), are now clear and to the point. In my opinion, the results, discussion and conclusions still lack focus.

**Reply:** We are grateful to the reviewer for reviewing the manuscript again and for acknowledging the improvements. Along with the comments of the Editor and the other reviewer, we have also tried to provide more focus to the discussion and conclusion sections, especially by clarifying the target audience and by providing clear recommendations for prospective target users (see details below) as well as adding a FAQ table for prospective model users.

When revising these sections it could help the authors to think of a target audience, for example, prospective model users and write for them, i.e., what does a prospective model user need to know to choose for 4C for their research instead of another model? If prospective users are targeted the discussion should answer questions like for which forest, soils, climate regions, etc is 4C a good choice? For which applications can 4C be used? Which simulations should be carefully interpreted given known model deficiencies? Which models with similar capabilities (but different strengths and deficiencies) are available? If the target audience are the 4C model developers and the study is seen as a benchmark then the authors should start from the objectives of the model (thus its reason of existence, why was this model developed?) and evaluate all process needed to demonstrate that 4C meets these objectives. This evaluation should be systematic and quantitative such that future versions of the model can be tested against the same data and the evolution in model performance can be monitored.

**Reply:** We agree with the reviewer that defining the target audience is key. Our intention is to write this paper for prospective scientific model users and we have made this now very clear in the text (abstract L30, L43, introduction L95-101, in each subsection of section 4).. We have also introduced recommendations for this user group in the discussion and partly rewritten/amended section 4.4. which also serves this purpose. Moreover, we have added a FAQ table for prospective scientific model users that builds upon the questions you have outlined above as table 7.

The figures are still plagued by inconsistent color coding (see previous review) making it difficult for the reader to understand the figures at a glance. Fig 8 for example, what is the difference between the black and red data? Simulations vs observations (as in fig 7)? If not (i.e. simply 25-75 quantile vs outliers) use a single color.

**Reply:** We revised Figure 7 and 8 regarding color coding and better comparability of the sites.

Fig 9 has very little value from a benchmarking point of view. If a model uses climate drivers with prescribed seasonality (as is the case for 4C) the results are expected to show a strong seasonality as well (as shown in Fig 9). To become meaningful from a benchmarking point of view the residuals should be shown, hence, the observations should be subtracted from the simulations.

Fig 10, is this the best figure for the message? A scatter graph would be better. Do you really need to show so many different soil depths? One figure + statistics for the other soil layers would be more informative.

Reply: Despite the expected seasonality, a comprehensive model validation should show the expected result if the corresponding data is available. This seasonality cannot be seen from the time course of the residuals. The residuals do not show the time lag in the course of the water content. The statistics for the residuals are given for all results and all soil depths in table 6.

For both sides, we show the results for soil temperature and water content for only one layer (Figure 9 and 10). We put the old figures 10 and 11 into the supplement.

The top half of the table with previous evaluation results still contains qualitative language. I.e., “good correspondence of simulated yield with experimental data”. Such assessment is useless from a benchmarking point of view because whether model performance is “good” or not depends on the assessor and the intended application of the model. Neither of these are reported.

Reply: We agree with the reviewer and have included quantifications wherever available. Unfortunately (and maybe not understandable from today's point of view anymore) this is not possible for all papers because some of the older papers do not give these quantities.

The conclusion starts with an outlook that has not been discussed previously. Parameters are mentioned in general but should be specified to become informative (which parameters do the authors have in mind and many records of that parameter are available in TRY for the species parameterized in 4C?)

Reply: We agree with the reviewer that the conclusion was lacking focus and have completely rewritten the conclusion which now focuses on achievement of the paper aims and future model improvements. The conclusion now reads:

The model 4C belongs to a family of process-based forest models operating at stand scale and has been widely applied to simulate climate change impacts on European forests in the last 20 years. The model has been evaluated in a wide range of situations and we here describe the structure and the processes of the recently open-sourced version 2.2 of 4C (Lasch-Born et al., 2019). The evaluation of the model's main processes relating to forest growth, carbon, water and heat fluxes against a comprehensive set of observational data at different time scales shows that, despite shortcomings, 4C is widely applicable and reliable. Particular areas for future improvement of 4C are (in no order of importance): refining species-specific model parameters (especially phenological parameters and respiration to enable organ specific respiration and allocation) using recently available data and methods, inclusion of abiotic disturbances especially fire and storm and better representation of ground vegetation, complex mixed forests with more than 2-3 species and their management. We conclude that 4C v2.2 is ready to be released to the scientific community to use and further develop the model in a wide range of climate impact studies on forests.