Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-2-RC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "Description and evaluation of the process-based forest model 4C at four European forest sites" *by* Petra Lasch-Born et al.

Anonymous Referee #2

Received and published: 3 September 2019

General comments: In "Description and evaluation of the process-based forest model 4C at four European forest sites" the authors does exactly what the title states. Although GMD could be a suitable outlet for such a study, the current manuscript is not ready yet for publication: (1) the model description skips over most of the numerical details which are the key interest of GMD readers; (2) given the model has been extensively evaluated before, the objective of this study needs to be clarified; and (3) the discussion is superficial and largely speculative. Although the simulation set-up and evaluation are sound but routine, moving beyond the current level of speculation would require more demanding simulation experiments. These main points are detailed below:

Printer-friendly version

Discussion paper

The objectives of the manuscript need to be clarified. Given the absence of new model developments and extensive previous model evaluation at the site level, the objective of this study remains unclear. The model has already been evaluated against many more sites (including 100s if not 1000s of sites of BWI) with similar data so what was the hypothesis that justifies running the model against just four more sites. What do you expect to learn from 4C by using the PROFOUND data that hasn't been revealed by the previous model tests? L474-L476 may contain a hint but this sentence needs to be rephrased as it is not at all clear to me what is being proposed.

The writing is often imprecise and fails to satisfy the interest of a readership rooted in the modelling community. For many of the interested details the readers are referred to previous work or have to do with a vague description. I listed some examples in the detailed comments addressing the first 15 pages but the team of authors seems sufficiently experienced to revise the whole manuscript by adding the missing information either in the main text or an appendix without detailed editing from my side. The results section contains many vague classification such as: good, better, underestimated, less, smaller range, . . . the reader wants to see the numbers (some are given in the tables but they should be integrated in the text to substantiate such statements). If you insist on using qualitative language you will have to make a table showing that, for example, an R2 of 0.5 is considered average, R2 of 0.75 is considered good, . . . it would also be nice if there is a scientific ground for such a classification. Given that this is a stand-level model that comes with simplifications, what would the authors consider good enough? One cannot expect a 100% match but what would give you confidence in the model?

The figures contain some inconsistencies: in Fig 7 the obs are shown in blue and the sim in red, Fig 8 shows the obs in black and the sim in red, and Fig 9 shows the 25-75th percentile in black and the 10-90th percentile in red. Changing the meaning of the colors does not make it easy on the readers. Nine tables and 13 figures is a about two times too much display items for a research paper, especially because it looks like

several of the tables could be merged and several of the figures are rather trivial or they simply complement the table; they show how the ME, R2 and NRMSE reported in the table looks like in data/simulation space. Such a figure could be useful to give the reader an idea how to interpret the tables but it becomes a burden to read when there are too many.

The discussion (4.1, 4.2) refers to relationships, equations and parameters that are not presented in the model description. Given the readership of GMD consists of modelers, show the equations that are essential to understand the discussion. Much of the discussion in 4.1 should have been presented in the results as it compares the simulation with other data sources. The sentence "Collalti et al. (2016) also found a better performance for their 3D-CMCC-FEM model on a monthly scale for these sites." (also see L540-541). Provokes more questions than it answers: what is the probability that this happens by chance? Which other sites did Collalti analyze? If a different set of sites where analyzed, thus this render this comparison meaningless? If you stand to your position and consider this a good comparison, what makes these sites stand out or which site properties makes them more easy to model their behavior? The discussion remains speculative and the model experiments do not allow to isolate causes. Isolating causes would have been required to result in new insights. The model could have been forced to use observed soil water contents, when compared with the available simulations, such an approach could have isolated the role of the soil water simulations in the simulation of latent heat and would have enabled the authors to move beyond the current speculation. L641-642 demonstrate this point as this conclusion is not based on the results. As far as I could tell none of the simulations show the possible contribution of an understory. A literature search on GPP and transpiration of the understory could help the authors to make this point but at current there is no evidence in support of this claim.

The discussion on L598-594 is trivial (all observations come with uncertainties). The authors should use this measurement uncertainty to set a clear target for the model-

Printer-friendly version

Discussion paper

data comparison. Either the measurement errors are small enough to justify a comparison and then a “good match” (defined by the target) would give confidence to the model or the measurement errors are considered too large and then the data should not be used. The current approach of using the data but casting doubt over their quality when there is a mismatch between the model and the data is unfair. Same measurement errors exist at the sites where the model performs well. What does this tell you of the model given that the data come with considerable uncertainty? Note that on L629 the term “sufficient accuracy” is used. What is considered sufficient in this context?

Detailed comments: Title: version number. Give all the information that is required to link this study to a specific model code. Make sure that when you update the code, the code underlying this manuscript can still be downloaded. I thought GMD required at least a doi referenced code.

L26: abstract start with describing the objectives of the model, not its age.

L34: delete one “on”

L33-37: Split in two sentences to enhance readability.

L46: delete “and”

L50-58: No need to be complete but a more structure overview of what is currently available and what makes FORESEE a unique model would be informative. How does it differ from some relatively recent developments in land surface models (see Naudts et al in GMD (the ORCHIDEE model), Fisher et al in GMD (CLM model with cohorts based on ED)) and how does it differ from individual based models such as iLAND (Seidl et al). In this respect, the introduction lumps stand-level and individual-based models in one group. This seems a bit rough to me. Highlight some similarities and differences between all these models. I really hope there are some differences in functionality and/or underlying principles. If not, the community is wasting a huge amount of efforts and resources by repeating the same work over and over again.

Printer-friendly version

Discussion paper

L68: “respects the principle of parsimony”. Several modelling groups embrace this principle but how was it applied. How is it reflected in the code? Which parameters were removed after test runs because they did not helped explaining additional variation? What are the thresholds for parsimony. Please detail how the parsimony principle has been applied throughout the 20 years of developments.

L82: The introduction does not describe a problem with the previous version of 4C or a hypothesis for which new model functionality is required to test it. Why did you believe it was worth to make the effort to conduct and write up this study? It could be that in the past the data used for model evaluation were spatially and/or temporarily incoherent (for example, NPP from MODIS but biomass from NFI or NPP from 1970 through IGBP and biomass from 2010 through NFI) and that the PROFOUNF database is the first consistent database for forest models in Europe. In that case you put the model to a new test that is more challenging than any of the other evaluations done with 4C. Likewise it could be that new model code was developed since the last manuscript and that this code is now evaluated.

L115: Add a caption to figure 1. What do grey boxes represent? What is the difference between a grey and a white box? What is shown by an arrow?

L118-124: List the four methods. The readers want to learn about 4C without having to read other papers/manuals/websites. This manuscript can be concise but it should be complete. The target audience of GMD are modelers, they want to learn how processes were modelled. It is frustrating to read a manuscript without getting the information one is expecting. These different approaches for the same model is a rather unique feature of 4C. It should be detailed so that the reader can better appreciate it.

L125-130: List the different methods. See previous comment.

L139: How is plant water supply formalized? Does it account for root water conductivity? State the assumption that is being made on root distribution and how it is calculated. If I understand it correctly, the plants take up water in proportion to the root mass

in that soil layer. If the top soil layers dry, the plant will thus experience water stress. How does this assumption holds against data?

L147-153: How is N calculated in the plant? Is the C/N ratio fixed? Is the C/N ratio dynamic? How does the plant C/N ratio interact with the soil C/N. How does it interact with (long term) environmental changes? Which nitrogen-species are distinguished (more meaningful for the description of the soil)?

L205-206: First time specific species are listed. Given the model aims to be generic, I assume that different species relate to different parameter sets. List the different species in the general description if you want to keep line 205-206. Once you mention one species, the reader wants to know for which other species the model can be used.

L217-218: Figure 2 repeats Fig 1. Remove fig. 1. Add a caption. Boxes, arrows and colors need to be explained.

L220: the typical disturbances are fire, wind, drought and pests. Following the description 4C only deals with pests. The title of this paragraph should be “pests”. Based on the description, the implementation of pests seems very empirical. A mechanistic approach should probably simulate its own pest dynamics as a function of the environmental conditions (which is key in the climate change discussion and the question of whether forest will suffer from more frequent and more intense pest outbreaks).

L256: the least that should be reported is the minimum and maximum number of parameters to give the reader some idea of the number of parameters.

L263-268: Rephrase “Calibration of the parameters is therefore not usually carried out when setting up the model for a new site.” It is not clear when the parameters are calibrated. I fail to see “Therefore, in recent studies, 4C has also been calibrated using a Bayesian framework (van Oijen et al., 2013; Reyer et al., 2016)” fixes the problems described above. Bayesian uses a prior and an uncertainty of the prior but it cannot distinguish acclimation from ecotypes. Rewrite this section as the logic seems

to be flawed. Is there a atypical sequence that is followed (first photosynthesis, then mortality, ...)? How many parameters are calibrated at once? Where do the priors come from? How was the uncertainty on the priors determined? Was heterogeneity used a proxy for uncertainty of the prior? This section lacks lots of basic information that is required to understand the model parameterization procedure.

L277-278: "From these data tree cohorts are generated using distribution functions." Setting the initial conditions is an essential step in a forest model. This description is too vague. Again, the readership are persons who want to know exactly that kind of details. If you feel this may overload the manuscript, add it into an appendix. Table 2. is nice as gives the reader a concise overview of which aspects of the model has already been evaluated but the wording in the results column is too vague. What is good? What is called "underestimation" and when is an underestimation so large that it becomes unacceptable or just bad? Those statements should be quantified. Percentage deviations, RMSE or even correlation coefficients could do the job here.

Table 2. Add version numbers to the different tests. Given the range of publication years, I assume this table refers to different model versions.

L300: the manuscript does not describe any changes to the model. Without model version numbers, table 2 suggest all these test have been performed on the current model version. This provokes the question what you expect to learn from the new sites that you haven't learned from the extensive previous test. The objective of this study and manuscript should be better described.

L305. Sorø needs to be introduced first. The reader has no idea yet that Sorø was one of the test sites.

L308. I suppose you mean "recalibrate" because the parameters were calibrated earlier for other sites. Now it reads as if the parameters were never calibrated which sounds unrealistic for the totality of the parameter set (I expect the most sensitive parameters were calibrated, others were just based on literature values and others were

just guessed and/or tuned because they cannot be observed). I stopped making detailed comments as I trust that the authors have enough experience in scientific writing to revise the remainder of the manuscript along the lines suggested by the above comments. Most of these comments require careful editing rather than an expert review.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-2, 2019.

Printer-friendly version

Discussion paper