Answer to the Anonymous Reviewer #1

The reviewer comments are in bold, and the replies in regular font.

Thum et al describes a new terrestrial biosphere model, called QUINCHY, and presents a first evaluation of the carbon, nitrogen and phosphorous cycle against site-level data. Although the quality of what has been presented is good, I'm concerned about what has not been presented in the manuscript and its supplement: (1) the benefit(s) of starting a new terrestrial biosphere model, (2) the impact of the "consistent model formulation" (as called by the authors), (3) a clear overview of what makes QUINCHY stands out among the existing terrestrial biosphere models, (4) an evaluation of the energy and water balance, and (5) the target/criteria used to decide that the model's performance is "acceptable".

We thank the reviewer for the feedback on the quality of the paper and for pointing out things that need some improvement in this manuscript.

(1) The authors remind the readers that "Many process-based models of the terrestrial biosphere have been gradually extended from considering carbon-water interactions to also including nitrogen, and later, phosphorus dynamics." and state that "This evolutionary model development has hindered full integration of these biogeochemical cycles and the feedbacks amongst them". Although I fully agree with the first part of their assessment, models like CLM (10.5194/bg-11-1667-2014), CABLE (10.5194/bgd-6-9891-2009; 10.5194/gmd-2017-265), ORCHIDEE (10.5194/gmd-10-3745-2017) and JSBACH (10.5194/bg-9-3547-2012) show that the second part of the statement needs to be toned down unless the authors can provide evidence in support of their claim. The current presentation contains no elements that demonstrate that the technical and/or scientific performance of QUINCHY was only possible due to the fact that the group started their model developments from scratch. Most of the groups that maintain and develop a terrestrial biosphere model that has a history that goes back to over a decade are likely to have considered a rewrite of their model at one point. Most of these groups, however, decided to continue with "evolutionary developments". If this evolutionary approach really hinders scientific progress (as the authors seem to claim), this is an important message but it should be supported by evidence.

We agree with the reviewer that the statement can be read to say that science itself would be hindered due to evolutionary model development, for which there is likely no citable evidence. We will therefore clarify (tone down) the statement in the revised manuscript.

However, we would like to point out that adding nutrient dynamics to a model introduces dependencies between model compartments that are typically less tightly connected in carbon only land surface type models, e.g. through the dependence of soil organic matter decomposition on nutrient availability, and therefore directly on plant nutrient uptake and productivity. These additional dependencies are sometimes in conflict with the pre-existing model structure, such that including nutrient cycles either requires substantial code restructuring, or scientific compromises in the extent of nutrient effects represented by the model (e.g. the need to assume that certain processes have priority over others, because they are calculated sequentially in the model). It is this complication and limitation that we were referring to when we wrote that evolutionary model development prevents full integration of nutrient dynamics, because taking account of all the interactions between nutrient dynamics and carbon and water cycles suggests that a fresh implementation has advantages and allows for a full and consistent representation of nutrient effects. We do not mean to suggest that evolutionary development of models necessarily results in inconsistent or inferior model results, but rather to justify our own choice for developing a new model. We will clarify this statement in the text.

(2) It is mentioned several times that QUINCHY has "a consistent representation of element cycling in terrestrial ecosystems". It remained unclear to me what is meant by this. Towards the end of the manuscript I was under the impression that "consistent" referred to the fact that all processes in QUINCHY are calculated at the same half-hourly time step. Although I can appreciate that such an approach makes the code easier to read and maintain, I'm less sure this approach can be claimed to be "consistent" because the time step of the model itself is still arbitrary (1800 seconds) when compared to the actual time step of the processes. Moreover, the idea to use different time steps for different processes has been justified by a more efficient use of limited computer resources. This far most terrestrial biosphere models favored speed above accuracy for the calculation of the non-linear processes. The QUINCHY group choose to trade computer time for an expected increase in accuracy. Can you demonstrate that there was an increase in accuracy? Based on your experience and findings can you recommend other groups to make the same choice? Will you maintain this "consistency" in the near future when adding landscape-level processes to the model such as plant biogeography and disturbances?

With consistent representation we were referring to the representation of nutrient feedbacks with a common set of hypotheses, on how plant growth (through the nutrient effects on photosynthesis or respiration), short to long-term labile carbon and nutrient storage as well as the interaction between plant N uptake and soil organic matter decomposition are considered. Such consistency is not necessarily maintained in other models, in which for instance, nitrogen limitation operates on different timescales to affect soil processes or plant growth (e.g. Xu-Ri et al. 2008), or in which N affects photosynthesis (e.g. CLM5, which employs the FUN model, with the intrinsic assumption of constant leaf C:N, while the model actually simulates flexible C:N). We will clarify what we meant by consistency in the revised manuscript.

While we have the same timestep for all model processes, some processes respond on different timescales, through time-averaging of driving variables and therefore their influence on the ecosystem state is smoothed. We will demonstrate the effect of this averaging now in a new figure as a response to reviewer #2. We do not aim to imply that the half-hourly timestep is strictly required for this type of biosphere model, but we do expect that there are benefits from using such an approach, for instance, because it avoids the need for latent pools, in which carbon and other elements need to be stored temporarily to link processes on short-time scale (such as photosynthesis and respiration), with that operating at longer-time scales (vegetation growth and dynamics). This physical consistency of pools and fluxes reduces the need for numerical fixes to maintain mass balance, and is strictly necessary for the accurate calculation of changes in isotopic composition of the biosphere. We expect to maintain this time-scale consistency also in future model versions, but remain open to simplify model structure, if we can prove that the simplification does not entail any relevant loss in calculation accuracy.

Given the modular design of QUINCY, we will be able to test the importance of the detailed versus lumped representation of processes when scaled to larger scale. This is something that will hopefully be valuable also to the other groups in the community.

(3) The authors claim that QUINCHY is a new model. Although I have no doubt that this assessment is correct from a technical point of view, it is less clear whether this is also true from a scientific point of view. It would be interesting to present the family tree of QUINCHY as it seems to be strongly inspired by O-CN (10.1029/2009GB003521). When thinking about weighting models in the IPCC context (10.1038/s41558-018-0355-y), would you argue that QUINCHY is independent or do you expect similarities with for example ORCHIDEE (10.5194/gmd-10-3745-2017) in which the C and N-cycle seems to be very similar to the one used in QUINCHY. If I understood the model legacies correctly, O-CN partly relied on ORCHIDEE and subsequent versions of ORCHIDEE (10.5194/gmd-10-3745-2017 and 10.5194/gmd-2018-261) relied on O-CN. Given that QUINCHY adopted many approaches from O-CN is it fair to assume that both models are likely to have some similar behavior? As a reader it is not clear at all what makes QUINCHY unique. After reading the current manuscript and its supplement, I expect that prospective model user will still have no idea when they should choose QUINCHY over CABLE, CLM, ORCHIDEE, JULES, JSBACH,. . .

While it is true that the QUINCY model has some commonality with the O-CN model (e.g. the photosynthesis schemes), the two models differ in fundamental aspects (e.g. the representation of labile pool dynamics and the competition of plants and soil organisms for nutrients, representation of vertical soil profiles, which affect the response of soil processes to perturbations). We will revise the manuscript to be clearer about which aspects of QUINCY derive from O-CN, and which are new (see also our response to reviewer #3). Given these differences, we do expect that the QUINCY model results can be considered as independent from O-CN. We would like to highlight that there are a number of important differences between the ORCHIDEE and O-CN models (in terms of the photosynthesis and allocation schemes and the representation of stand-level vegetation dynamics), such that these models should also be considered as independent.

We have highlighted in the manuscript the processes that are considered novel in the context of large-scale biosphere modelling. In this manuscript, we provide a model description and first evaluation to lay the groundwork for future studies evaluating the novel aspects and features of this model, and only together with these studies (which as reviewer #3 points out merit a scientific paper on their own) it will be possible for the wider community to decide as to whether the QUINCY model is an interesting and valid contribution to the ensemble of terrestrial biosphere models.

(4) Although the SI presents the formalisms used to simulate the water and energy budgets, these processes are not at all discussed in the manuscript. The whole point of having a terrestrial biosphere model (especially when it will be coupled to a general circulation model which is the case for QUINCHY) is that the terrestrial biosphere model links carbon, nutrients, water and energy cycles in a quantitative way. In my opinion, the most telling evaluation targets for a terrestrial biosphere model are those showing the skill of the model in jointly reproducing two or more cycles. Such analyses has not been presented.

We agree that it is important to show two or more cycles jointly, which is why we show the model behaviour with the carbon only version alongside the carbon and nitrogen as well as the carbon, nitrogen and phosphorus versions, and provide metrics of carbon, nitrogen and phosphorus cycles at selected sites. We choose to not show a detailed evaluation of the water and energy cycles in this paper, as the representation of these cycles will very likely be replaced after the coupling with ICON. Nevertheless, we will add an evaluation of the latent heat flux prediction at FLUXNET sites to provide also an evaluation of the water cycle simulation of QUINCY.

(5) The evaluation is sound but routine meaning that no clear effort was made to go beyond the typical "acceptable performance" where "acceptable" remains undefined and "performance" is limited to a RMSE or a correlation. I do realize that this represents a common modus operandi in the community but the tools and data exist to do better. Hence, there is no excuse for a leading journal as GMD not to raise the bar by insisting on more ambitious evaluation practices. Could you, for example, set quantitative targets, i.e., reproducing 95% of the seasonal cycles in addition to 50% of the residuals data structure (i.e. observations minus the seasonal cycle)? Or using a simple purely climate driven statistical model as the reference to beat? Subsequently, quantify whether these targets were met or not. The statistical methods for such an approach are available and have even been proposed for spatially explicit analysis (see SI of 10.1038/nature02771). Furthermore, the study somewhat overlooks the concerns of the community who wants to learn about the performance of QUINCHY who presents itself as "the new kid in town". From a community point of view it would make sense to run the model through the ilamb benchmarks (10.1029/2018MS001354) and compare QUINCHY's performance relative to what is considered state of the art within the community (in addition to the evaluation shown by the authors).

In this paper our aim is to introduce the model and present its functionality and in the revised version of the paper we will aim to better illustrate the underlying novel processes so that the high level model evaluation can be better understood. We agree with the reviewer that the model evaluation is very important, but we also believe that rigorous benchmarking would be a study in its own.

The reviewers #2 and #3 requested more emphasis on illustrating the new model characteristics, and we do find these points important, as this is what is novel in this study and these sides of the model were not shown properly in the first version of the manuscript. These new figures will show more functional dependencies, i.e., how do some variables change according to the meteorological conditions and nutrient deposition. This is a more qualitative way to look into the model's behaviour than rigorous numerical benchmarking, but this is of utmost importance when considering the potential applications of the QUINCY model. Therefore we will now put more emphasis at this point to these analysis, but we agree that the numerical model performance against benchmarks is also important and should be addressed in future work.