Reviewer #2 comments

General comments:

1. This paper sounds like a well written technical report and needs major changes before it can be published in GMD as a model description paper. A number of models, not all very complex, are mature enough to do the same job (and more) as the one proposed by the authors."

We appreciate the comment, and agree that there are many combined land surface water, energy and carbon cycle models in existence, and there are certainly models out there that are also mature enough. However, we are not aware of and could not ascertain that there a such models with a proven history of land surface data assimilation. A possible reason for referee #2's opinion here is that he or she refers to a much broader range of models and modelling frameworks, including weather forecasting and climate modelling.

For clarification, we have added two paragraphs discussing the suitability of other models for the task (see reply to Reviewer #1 Major Comment 1)

2. Why is this new model needed? In terms of process representation, I could not see any innovation in the proposed modeling framework, except for the simulation of SIF and VOD. I assume that the innovation is in the data assimilation part, but no example is shown.

The observation of referee #2 is correct insofar as the novelty of the described modelling systems is related to its data assimilation capabilities. These capabilities depend on two things: 1) a set of observation operators that translates model output into observables, and (2) the data assimilation framework. We have deliberately refrained from including (2) as this would have unnecessarily made an already long manuscript unwieldy. The innovation can thus be found in the observation operators for FAPAR, VOD, SIF and surface soil moisture. We would also like to point out that there is currently no VOD observation operator or even approach available in the literature that is able to exploit short-term (i.e. sub-annual) variations in VOD.

We believe that the text added comparing D&B to previously published models provides the necessary clarification (Reply to Reviewer #1 Major Comment 1).

3. Since data assimilation is not demonstrated here, "data assimilation" should be removed from the title. Similarly, data assimilation is mentioned frequently throughout the paper, but no example is provided. Data assimilation could be mentioned briefly in the Introduction and as a perspective in the Discussion section. No more. A clear definition of data assimilation is also lacking. Data assimilation can be done in many different ways. As far as I could understand, in this paper data assimilation is equivalent to "model parameter tuning". This is quite different from the variational or sequential Kalman filtering methods used in meteorology and in some land modeling frameworks to initialize initial conditions (e.g. of root-zone soil moisture) at a given time. This should be clearly explained.

As mentioned in the previous comment, the title includes the purpose of the model to be used within

data assimilation frameworks. If we drop a reference to data assimilation in the title, we run the risk of losing the main target group of this manuscript, which is the community interested in using various data assimilation methods to infer important parameters of land surface states. This was also the purpose of the European Space Agency project the work was funded by.

As far as a the specific method of data assimilation is concerned, we would like to repeat that this is specifically and intentionally not part of the manuscript, since the model could be used for various data assimilation setups. However, we can see that this was not clear from how we referenced data assimilation.

4. It should also be explained why the authors do not trust their default model parameter values. I assume that these default model parameter values come from the scientific literature. Why not trust them?

This is a valid question. Our data assimilation work rests on Bayesian statistical principals which specifically assume that no measurement, parameter value of model formulation should ever be fully trusted. We would like to reserve a full discussion of this for a later manuscript that demonstrates the data assimilation system built around D&B.

To clarify this, we have added the following text to Section 1:

"Data assimilation offers a valuable tool for automatically finding the optimal combination of model initial values, parameters and even input quantities given the observations assimilated, pertinent to certain assumptions about prior values and uncertainties of models and data within a Bayesian framework (Tarantola 2005)."

5. Finally, a critical risk of parameter tuning is that the tuned model might be good for bad reasons, which is unacceptable for a model that aims to explicitly represent the main biophysical processes ("process-based modeling system"). This is acknowledged by the authors on L. 546. But how do the authors ensure that this does not happen? This is not clear.

We can see what Reviewer #2 alludes to here, and, as mentioned, we discuss this point already. However, there is very clearly no set recipe to prevent that a model is "right for the wrong reasons". It is a matter of expert judgment and requires both knowledge of the model, experience with its behaviour, and an intimate knowledge of the processes represented, ideally underlain by field experience.

We have added the following statement to the end of Section 6.3:

"The potential advantage of D&B coupled to multiple observation operators is that it allows model testing via multiple data streams, thus providing are more comprehensive model evaluation which makes it less likely the model matches observations while misrepresenting important processes." (Also in reply to Major Comment 8 of Reviewer #1.)

6. Finally, the rationale for "parameter tuning" is that improved static model parameter values are needed for the surface component of climate models and for climate change impact models. I do not believe that the current model is designed for such applications. What is the real purpose of the model? For monitoring and reanalysis, sequential assimilation would be preferable to model tuning.

This is very good point. We agree with the argument or Reviewer #2 that, when it comes to the purpose of data assimilation, deriving static parameters lends itself for the improvement of process representation (e.g. in climate or climate impact models, as mentioned), and adjusting initial conditions lends itself naturally when it comes to monitoring or short-term forecasting. However, there are many more possible configurations, including adjusting input data, or a combination of adjusting all those simultaneously, i.e. input data including land type classifications, initial conditions. In fact, in a probabilistic Bayesian methodological framework, not adjusting any part of the model is always a compromise between mathematical accuracy and computation feasibility.

See reply to Comment 4 of Reviewer #2 for text we have added to Section 1. The text should clarify that the D&B framework is open to all of such application.

7. This work combines the BETHY and DALEC models. A comparison of the model simulations is presented over two contrasting sites (Spain and Finland). Why these two sites in particular? Results from the comparison are not good, which tends to show that this new model is not a good model. Or maybe these sites are particularly difficult to represent? Could you indicate score values from other models over these sites?"

The reason for choosing those two sites are that, while both are located in Europe, they cover a large variation in ecosystem properties. The parameters chosen were not site-specific and therefore we do not expect a perfect match.

At the start of Section 4.1, we have added "The D&B model is run for two study sites ***with widely varying climate****."*

Particular comments:

8. L. 54-59 (list of requirements): Is it something that other models could not do?

See the added discussion of alternative models (reply to Reviewer #1 Major Comment 1).

9. L. 97 (daily time step): I believe that the daily time step is not sufficient to represent snow processes. Especially when snow melt occurs.

We have included a comparison between modelled and observed snow height an find a rather good agreement.

10. L. 154 (potential photosynthesis): It is not clear whether potential photosynthesis varies from one day to another according to solar radiation and leaf temperature. Could you clarify?

The answer is that it does. We have added: "This rate of demand is determined by the potential rate of photosynthesis without water stress ***computed previously at each time step****."*

11. L. 275: The ICOS data portal contains a large number of sites. Why have you selected these two sites in particular?

See answer to Comment 7 of Reviewer #2.

12. L. 324 (overestimate of GPP): Over the Boreal site, GPP is much more than "overestimated". There is nearly a factor of two at summertime. Why is the model that bad?

A factor of two can easily obtained within the uncertainty of model parameters, such as Vmax, or state variables, such as LAI or fractional cover.

13. L. 433: On L. 214, Ssif = 1 and on L. 434 Ssif = 10. Could you explain why?

We have added the following sentence to the second paragraph of Section 5.2:

"While the prior value of s_SIF was 1, this change reflects the high uncertainty regarding the absolute magnitude of the measured SIF."

14. L. 481: "in the simulations, soil moisture decreases to near zero": why? Could this be caused by the overestimation of soil evaporation (a classical modeling problem)?

Thank you for pointing this out. In D&B, soil evaporation when approaching zero soil moisture decreases linearly with soil water content (see Eq. 83 of the SI). Judging from the observations, it may be possible that the relationship is much more non-linear than expected. Unfortunately we lack measurements of soil evaporation and are therefore unable to systemtically explore alternative formulations.

We added the sentence

"We also find that the model may overestimate soil evaporation for very dry soils."

to Section 6.3 (3rd paragraph).

- L. 486-487 (carbon fluxes [...] are simulated reasonably well): You cannot say that for the Boreal site.

*Agreed. We have added the following qualifier: "… reasonable job at representing energy and carbon fluxes between the atmosphere and terrestrial vegetation****, albeit with the seasonal amplitude of the net carbon exchange overestimated at the boreal site****".*

We do not refer to either GPP or TER, because these are derived quantities and not direct observations.

15. L. 529: The large uncertainty on Ssif shows that the biophysical basis for SIF is very weak in the proposed model. Why not using machine learning to build an observation operator for SIF?

The link between electron transport and SIF is well established in the literature. D&B explicitly models the electron transport, and so it makes sense to couple the Gu et al. (2019) leaf level SIF model in this manner. However, all SIF models rely on some form of internal parameterisation and it is clear that some work is required on how we couple D&B and the Gu model. We see this as an opportunity. The more we are able to strengthen that link between D&B and the leaf level SIF model, the better we are able to interrogate the model performance at a process level.

In principle, we have no objection to building machine learning emulators to serve as observation operators, but we argue that this is best done once the biophysical mechanisms are well understood, otherwise the emulator becomes a black box where we may struggle to interpret the results. Understanding of the biophysical mechanisms is best served, in our opinion, by building process level models.

We have also added the following paragraph to Section 5.2:

"The difference in magnitude between the modelled and observed SIF is likely due to the choice of prior parameters for the SIF model, taken from (Gu et al. 2019). Although it has not been done here, there is scope within D\&B to adjust these parameters in the assimilation. We believe, however, that tt is more important, in the first instance, that we have a model that can track the seasonal and diurnal cycle of the observations, and this appears to work reasonably well."