Response to Referee #3

We would like to thank Referee #3 for the review, and will improve on the points raised by the referee. Below, we address the comments of Referee #3, with the referee comments written in italics.

In general, I felt the authors could have been a bit more careful with their writing. For example, they confuse their hypotheses throughout the manuscript (see below specific comments) which sometimes makes it hard to follow their argumentation. The introduction is a bit too long and, in parts, not very well linked. I liked how the authors explicitly point out the four hypotheses they aim to explore, however, it is not quite clear to me how hypotheses 2 and 3 emerge from the introduction. It further would be nice to have more details on the model description – you could move table 2 to the supplement for example to make more room. Your discussion is detailed but I'd like to see more about possible future directions.

We are sorry that the numbering of hypothesis was confused a few times and thank the reviewer for picking this up. We will carefully check this in the revised manuscript. We will also improve the introduction by shortening it and improving the flow, as requested.

We tried to introduce the motivation for Hypotheses 2 and 3 in the introduction (Lines 122-129), but will be more explicit in the revised manuscript. Briefly, plant hydraulics are currently being implemented in TBMs as a major limitation for water use during drought, whereas in the VOM, it is only represented in the form of the water transport cost factor. For this reason, we test with this hypothesis whether a general water cost factor across sites leads to reasonable results or if site-specific cost factors would yield significantly improved results.

Regarding Hypothesis 3, we will emphasize more in the introduction that the prognostic simulation of phenology (LAI and vegetation cover dynamics) is a central concern of vegetation models and optimality theory. For this reason, a systematic evaluation of the prognostic vegetation cover in comparison with more conventional approaches (i.e. prescribing values), should show how the optimality theory, as applied in the VOM, extends the capability of conventional models and how it can be further improved. We will more explicitly discuss future directions in this context as well, as requested.

We included a detailed model description in the accompanying manuscript in GMD (https://doi.org/10.5194/gmd-2021-151), which is why there are not too many details in this manuscript. However, also based on the comments of the other referees, we will add more information here and refer more explicitly to this technical note for details.

Specific comments

Line 41-44 The sentence is very long and hard to follow

We will rephrase and shorten this sentence.

Line 61 Doesn't the default version of LPJ-GUESS have more than five plant functional types?

Thank you for pointing this out, we will correct this.

Line 91-92 The contents of the sentence are not linked very well

We will rephrase this sentence.

Line 98 Increase or decrease in annual rainfall?

We will add here that it is a decrease towards the south.

Line 118 'optimizing vegetation properties to maximize the NCP'?

Will be changed accordingly.

Line 136 Do timescales of precipitation matter? I.e. is annual PPT driving the rooting depth or are seasonal timescales more important?

We will add more discussion about this topic. In the VOM, rooting depths are a result of the long-term optimization of the roots, and do not only depend on climate but also on hydrology, i.e. the water storage capacity of the soil and the distance to groundwater.

Line 136 Therefore instead of but? '[...] therefore is likely to change over […]'

We will change this to "and is therefore likely to change over…".

Line 162 In table 1 it's AU-How

We will correct this.

Line 173 In table 1 it's AU-DaS

We will correct this.

Line 199 '[...] is defined by maintenance respiration, projected cover to the turnover and maintenance of leaf area' – I find this sentence a bit unclear

We will split this sentence into two sentences: " *is defined by maintenance respiration. At the same time, the cover is linked to the turnover and maintenance of leaf area, while...*"

Line 208 How can seasonal vegetation cover vary on a daily basis? Maybe rephrase

We will clarify that in the VOM, seasonal vegetation cover is allowed to vary slightly from day to day, resulting in a seasonality with a maximum during the wet season, and a minimum during the dry season.

Line 225-230 Does SILO provide point data or are the site met data derived from a spatial dataset (if yes which resolution?) I understand the argument that a longer timeseries helps to run the model, but it would be nice to see any sort of comparison between the observed met data at the site and the SILO dataset. I guess in general I would just like to have more information about the input forcing to get an idea about the uncertainty. Do the models from the Whitley paper run with the same meteorological forcing or do they use the data collected at the site?

We will add more details about the SILO-data. We will also refer to our accompanying technical paper in GMD (https://doi.org/10.5194/gmd-2021-151, Supplement S4), where we replaced the daily meteorological data from SILO with aggregated daily data from the flux towers. Eventually, this did not lead to strong differences in the results. The models in Whitley et al. (2016) were generally run with the flux tower data, except for BIOS2 that also used a gridded product. We will explain this in the text.

Line 235-236 Can you provide a bit more detail about the water retention model? It was never mentioned before

We will add more details about the water retention model. There are also more model details in the accompanying technical note in GMD (https://doi.org/10.5194/gmd-2021-151).

Line 247 Can you specify what the 'usual energy fluxes' are

We referred here to observations of incoming and reflected solar radiation. This will be clarified.

Line 249 LE has already been introduced two sentences earlier

We will correct this.

Line 264 Isn't the last hypothesis about rooting depth?

We are sorry for the confusion, we changed the order in a previous version of the manuscript. We will carefully check this and correct it. Thank you for pointing this out.

Line 265-276 I might have just overlooked in your submission – but can you describe in more detail what the model set up is for the model intercomparison you use from the Whitley et al paper? Surely there will be more detail in the Whitley paper to help understand but while reading your submission I was for example wondering whether there are changes to some of the parameters in the models to capture the site specifics better or whether they 'just' ran in their original configuration with the meteorological forcing from the sites [...]

We will add more specifics about the models here.

Line 277 Third and fourth hypotheses?

Thank you, we will correct this.

Line 294 In the introduction it says -0.1-0.1 for the cost factor for water transport (but I might have understood?)

In the introduction, we discussed the variation, instead of the absolute value. We will clarify this.

Line 297 second hypothesis (also Line 298)

Will be changed accordingly.

Line 300 'Regardless of the result here' can you explain why you make this decision?

We used this value, at it was the outcome of the sensitivity analysis by Schymanski et al. (2015). The assumption was that this cost parameter is valid with this value for all sites, while here we assess the original assumption using a sensitivity analysis. We will clarify this a bit more in the revised.

Figure 2: Maybe include shaded areas for dry and wet season, but also include dry and wet season months in caption. Can you include what the ensemble years are too? Panel e says Daly Uncleared but it was referred to as Daly River before

Thank you for these suggestions, we will make corrections accordingly. Daly River and Daly Uncleared refer to the same site, we will make this more consistent throughout the manuscript.

Line 306 Can you define dry/ wet season (which months)?

The wet season is from December-March, and the dry season from June-September. We will add this.

Line 327 Not sure, it looks like the minimum is quite similar for the models but for the maximum values, LPJ-GUESS and MAESPA seem to be too low

Thank you for pointing this out, we will more carefully formulate this sentence.

Line 400 I'm not sure I agree with this. It may be true when you look at the annual values in figure 3 but based on figure 2 you can't really reach this conclusion.

We will rephrase this, our main point is that the VOM does not perform substantially worse than the other models.

Figure 3: Why are the data points connected in panel a and b? It already is hard to distinguish the data points, the lines make it even harder. You also do not connect them in the other panels – it might be nicer to be consistent. Further, it might be helpful to offset the points in a and b (like in c-f). You could also increase the size of the observation marker, it gets lost in all the other points. Lastly, it could be helpful to include an arrow below the lower x-axes indicating whether the sites go from dry to wet or the other way around (but a lot of this is personal reference of course)

We tried to assess the patterns over the transect, which is why we added the line. But we agree that it may be clearer without the lines, so we will re-plot the figures without lines and then decide which is clearer. We also like the other suggestions, and will add an arrow indicating the dryness and increase the size of the observation markers. Thank you for the suggestions.

Figure 4 and 7: Maybe use a global legend and remove obs legend from panels. Are you ever using the information of Qflag in the results or discussion? If it never comes up you might as well delete it.

We will make a global legend, but believe it is good to be transparent about the data quality. We will add some discussion about the data quality in the main manuscript.

Figure 6 the text on the x- and y-axis is too small

We will increase the font size.

Figure 8 It would be nice to stay consistent in the color choice for the models

We will change the colors accordingly.

Line 569 Aren't some of the models (all?) processed based and not empirical?

We agree that our formulation here seems to indicate this, but we referred only to the more empirical components in these models, such as prescribed vegetation cover and rooting depths. We will try to remove this confusion in the revised manuscript.

Figure S2.1, S2.4, S2.7, S2.10, S2.13, panel h – can you maybe add padding between
100% projected cover and the figure edge, the way it is now it looks like you're cutting off
at 100%

We will change this accordingly.

In general, you have a lot of supplementary figures but don't refer to all of them in the
manuscript.

This is indeed true, we will add some more references to the supplementary material. We added relatively many, in order to provide full transparency and background information. Our supplementary information is also the output of the fully reproducible and re-usable modelling workflow we employed here, so it contains a bit more material than strictly needed for the paper. We will make this a bit clearer in the main paper and the introduction to the SI.