Reply to Comment by Anonymous Referee #2

April 3, 2023

We thank the anonymous reviewer for the helpful and detailed comment.

In the following, review comments will be styled in *blue italics*, our responses in black normal text, and proposed changes in green.

1.1

I think both the abstract and conclusion do not fairly reflect the results and discussion sections. Benefits and pitfalls have been discussed for both bias-blind and bias-aware approach and the bias-blind approach leads to greater improvements for most of the variables and most of the metrics. However, the abstract/conclusion leads to the recommendation of the bias-aware approach, which is partly inconsistent with the message sent from the result and discussion section. I would recommend the authors to reconsider making the key points that can better and fairly reflect the content.

In our conclusion we tried to point out the specific circumstances under which the bias-aware DA might have benefits over the current literature standard of bias-blind DA. Based on your and the other reviewers feedback we agree that this can give a false impression of disregarding the bias-blind DA. We will therefore add the following paragraph after l.549:

The bias-blind DA is most effective at reducing the disagreement in modelled and observed LAI, and leads to the largest improvements in GPP and runoff. For many applications it is therefore a suitable option, especially if large bias reductions are intended, even if bias-blind Kalman filtering is suboptimal. A temporal interpolation of the observation data, or even a direct insertion approach, could be even more efficient in this case. However, this approach does not necessarily improve other variables, e.g. if the model simulates biased LAI in conjunction with unbiased soil moisture. As alternative, we recommend to use observation rescaling techniques for LAI DA with Noah-MP if there are strong biases and if

- ... (same as in original manuscript)

Additionally, we will reformulate the last paragraph of the conclusion as follows:

A drawback of the observation rescaling approaches is that they result in estimates in the model climatology. If the observation-forecast bias is due to erroneous precipitation forcing or missing irrigation input, joint updates of LAI and RZSM in a bias-blind system can be considered instead. This might lead to large bias corrections while still retaining a stable model state even after large updates. However, if the bias is not only caused by bias in the precipitation/irrigation, this poses the risk of seriously degrading the soil moisture estimates.

Alternatively, updates to model parameters, either via joint parameter and state update DA, or via a priori model calibration can also lead to more stable and persistent updates and LAI estimates in the observational climatology. This is especially desirable for research on the carbon cycle, where absolute values of carbon fluxes are required. Parameters to consider for calibration are parameters related to model leaf growth, but potentially also photosynthesis or soil parameters.

To gain the most benefit from LAI data assimilation into Noah-MP, further research and model improvement of the coupling mechanisms between water and carbon cycle is necessary.

1.2

L48-49: "It is possible that other processes (e.g., transpiration) are only represented well for a biased model climatology". If that turns out to be true, isn't it right for the wrong reason? I think such side effect should be fixed by improving the model physics instead of regarding as the weakness in the "bias-blind" DA approach.

We agree that the best way to fix such effects would be by improving the model physics. However, this is out of scope of this study, where we only consider updates to the LAI state in the model. We discuss parameter update DA or model calibration as alternatives to updating only the LAI state, which might achieve the same effect as improving the model physics.

1.3

The spatial resolution of the simulation is coarse while there are evaluation reference datasets from in situ observations. The scale mismatch is not well considered in the metrics and comparisons in this paper, which may provide biased assessments. Would it be possible to conduct the simulation at a finer spatial resolution if most of the input datasets and the LAI observations are available at finer scale?

We used the ERA-5 reanalysis as forcing data, which are not available with a finer spatial resolution. We therefore do not believe that a simulation at a finer spatial resolution will significantly improve the model results, at least in terms of anomaly-based metrics, while it would at the same time strongly increase our computational demands.

Due to the scale mismatch, we limit the comparisons of in situ soil moisture data to anomaly-based comparisons, as mentioned in l. 249-250 in the original manuscript:

"Since soil moisture climatology and absolute values strongly depend on sub-grid scale factors like slope and soil texture, we only compared the in situ values in terms of anomaly correlation R_{anom}*."*

Additionally, we mentioned in our discussion that the scale mismatch hampers our assessment, l. 467-470: *"This can be due to assumptions and errors in the underlying satellite data and retrieval algorithms in the case of satellite-based data, or due to different spatial support in the case of in situ data. Hence, whether the bias-blind DA leads to estimates closer to the "truth" remains uncertain, and evaluations with different reference products might come to different conclusions"*

To increase the reliability of the comparisons, we propose to restrict the evaluation to ISMN stations that have been shown to be representative at the coarse scale in a triple collocation analysis including ISMN, ERA5-Land volumetric soil moisture layer 1, and ESA CCI soil moisture.

1.4

L102 - 105: What is the spatial resolution of these input datasets as well as the ERA5 forcing datasets? Please clarify in the text.

The parameter datasets are available on a 0.01° regular grid. The ERA5 forcing dataset has an original resolution of 31 km and has been rescaled to a 0.25° regular grid. We will add the following sentence in l. 114:

The soil texture and land cover maps are available on a 0.01° regular grid and have been upscaled to a 0.25° grid using the largest fraction within a model grid cell. The ERA5 forcings are have an original resolution of 31 km and have been rescaled to a 0.25° regular grid.

1.5

L106-108: I didn't quite understand why the model interprets the evergreen broadleaf forests as tropical rainforests. And what land cover type does the UMD data assign to for these locations? Are there any references that can justify the substitution of UMD data is more realistic?

The Noah-MP vegetation model allocates GPP to leafs according to a parametric allocation function. This function makes a distinction between evergreen broadleaf forests (EBF) and all other land cover types such that the maximum LAI is approximately 8 for EBF and approximately 6 for all other land cover types (see Figure 1 in this document). This is done to account for the much higher productivity of tropical rain forests than other land cover types. However, we found that for the pixels in Europe assigned to EBF this approach produced even more strongly biased LAI estimates than we showed in our manuscript. Replacing the land cover of these pixels with UMD as fallback option reduced this bias.

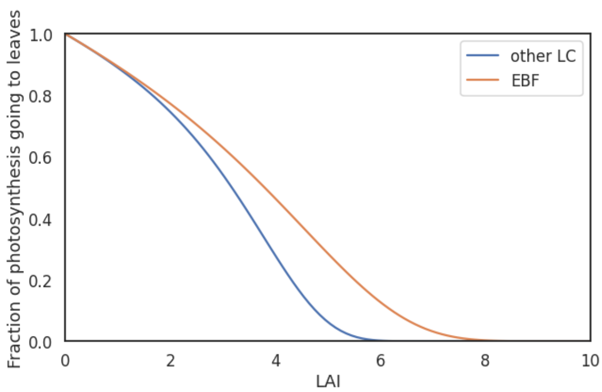

Figure 1: Allocation of total photosynthesis (GPP) to leaf mass balance

1.6

L116: Is there a specific reason to use a spatial resolution of 0.25 degree? Why not bring it to a finer resolution since your CGLS LAI product is at such a fine resolution.

As mentioned above, a simulation on a much finer scale would be a lot more computationally demanding, while at the same time the benefit is questionable, since the forcing data is only available on a coarse resolution.

1.7

L125: Is there a particular reason not using the temporal interpolation method? If this can better deal with the sawtooth issue, why not apply it?

We do believe that the interpolation technique might be a viable solution for some of the issues associated with the bias-blind approach, but for this study we had 3 particular reasons not to perform interpolation:

- Interpolation can only be reasonably used if the assimilated LAI product is smooth enough, so that we can assume that the pseudo-observations introduced by the interpolation are close to the true observations. This is the case for the CGLS LAI product, but not for other products. By not interpolating the data, we firstly stay closer to how the data is provided, and secondly we also obtain more insights into how observation frequency (and possibly changes therein) affect the data assimilation. This could be especially useful if microwave vegetation optical depth is used as LAI proxy, similar to Mucia et al. (2021).
- Interpolation of data will introduce a strong auto-correlation of observation errors. However, our data assimilation framework assumes that observation errors are uncorrelated.
- The main effect that interpolation has is to pull the model closer to the observations and hide the drifts between DA updates better. If this is the intended goal, a direct insertion technique might be even more suitable than a Kalman filter.

To make our reasoning more transparent, we will modify l.125 as follows:

In contrast to Kumar et al. (2019b), we did not interpolate the LAI to daily values. This way we (i) do not introduce observation error auto-correlations, (ii) allow our results to be generalizable to LAI data sets (or proxy data sets as used in S. V. Kumar et al. (2020) and Mucia et al. (2021)) with less frequent observations or changes in observation frequency, and (iii) can investigate if the filter efficiently interpolates and operates as intended (or assumed). We assimilated the aggregated data every 10 days at 0:00 UTC, where and when they are available.

Additionally, we will add the following paragraph in the discussion, after l. 485:

The sawtooth pattern can be reduced by interpolating the observations or applying time series smoothing methods to obtain pseudo-observations at a daily frequency. This will keep the analysis closer to the observations and prevent model drift over multiple days. However, in this case direct insertion approaches or using observed LAI directly as model parameter could achieve even better results than an EnKF.

1.8

L163: For the seasonal scaling approach, how does the phase of seasonality look like between the LAI observation and the model simulation for the study domain? For instance, what does the spatial map of the peak month in LAI observation compared to the OL simulation? The vegetation scheme in Noah-MP has weakness even in reasonably estimating the magnitude and phase of the seasonal cycle of vegetation growth. It may introduce additional bias if rescaling observation based on the modelled climatology. Any comments on this?

The bias-aware LAI DA will lead to LAI estimates in the model climatology, so in fact any biases in magnitude and phase of the seasonal cycle in the model will, by design, still be present in the analysis. We agree that estimates in the observation climatology, as obtained with the bias-blind approach, might be more suitable for some applications, but this comes with the drawbacks presented in our study (model drift, biased flux estimates, large side effects). As discussed in the manuscript, a better calibration of model parameters might combine the respective advantages of the two approaches. In the revised section, we will also add a discussion of joint LAI and RZSM updates as further alternative, which could also help to obtain stable estimates in the observation climatology.

1.9

L247: Considering the coarse spatial resolution of the model set up, the scale mismatch much be an outstanding issue when comparing the gridded soil moisture value to the in-situ observations. A simple nearest neighbor matching between ISMN stations and model grid might be troublesome. Have the authors considered the representativeness of the ISMN data for a model grid? Please comment on it and potentially discuss the uncertainties. I think simulation performed on a much finer scale might be better if one were to directly compare the in-situ observation to simulated values for a model grid.

Comparisons to in situ soil moisture are commonly done also for coarse model simulations, Sujay V. Kumar et al. (e.g., 2014), Sujay V Kumar et al. (2019), Albergel et al. (2020), Rahman et al. (2022a), Rahman et al. (2022b), and Heyvaert et al. (2023, in review). To account for bias due to the different spatial support/representativeness, e.g. due to different soil texture we only compare soil moisture anomalies.

Additionally, as mentioned above, we propose to restrict our analysis to ISMN stations considered representative at a coarse scale.

1.10

L260 -262: Why not use the dam and irrigation specific datasets to mask out the basins that are heavily affected by these factors? I think directly masking out the basins based on the correlation threshold of 0.4 is biased because it is impossible to justify that such low correlation is surely due to the unmodeled processes. Such selective results may mislead audience in terms of the DA performance. I would recommend the authors reconsider the approach.

This threshold removes only very few outlier stations, while most other stations have correlation values larger than 0.6. For reference, we provide the original correlations of all considered stations in Figure 2 (this document).

1.11

L308 & L402 -403: I wonder if the case of Nile delta may not simply because of the lack of irrigation representation in the model. How does the LAI time series look like for the full study period compared to the time series of precipitation? Did you see consistently low LAI regardless of dry and wet years? I wonder if the representation of vegetation in response to water stress factor or the mismatch of soil types may play a role in the dynamic vegetation model that limits the growth of vegetation for this area? There are multiple reasons that the model may underestimate/overestimate vegetation growth just as the author discussed in section 4.4, I strongly recommend the authors take a closer look at the data and elaborate more on the discussion in terms of the factors at play. The analysis presented here does not justify that the lack of irrigation forcing is the main reason.

OL LAI is consistently lower than the observations, but shows strong interannual variability

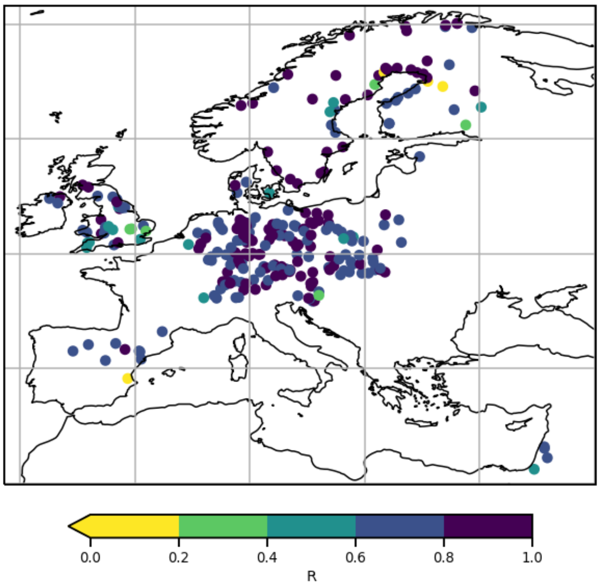

Figure 2: Correlation of GRDC runoff with OL.

related to water availability. An additional analysis of the sensitivity of the model equilibrium LAI on RZSM showed that in the Nile delta LAI is highly sensitive to RZSM.

We will add the following paragraph in the discussion:

In the Nile delta, a known source of bias is the missing irrigation input in the model. Additionally, we found a very strong water limitation on vegetation growth, and a strong sensitivity of the equilibrium LAI on changes in RZSM, implying a strong model coupling of LAI and RZSM. Therefore, joint updates of LAI and RZSM are likely to improve the LAI DA results here, because RZSM is temporarily adjusted, but changes to soil parameters might also be necessary to sustain the increased moisture values.

1.12

L315 - 317: This explains places where bias-blind DA leads to LAI decrease and soil moisture increase. I wonder how would the authors interpret places where bias-blind DA indicates an LAI increase, while there is no soil moisture decrease?

The areas where the bias-blind DA increases growing-season LAI but does not decrease soil moisture are high-altitude regions in the Alps and Norway. These areas have some specific attributes that limit the impact of LAI change on soil moisture:

- Precipitation is frequent and keeps the soil moisture high. Any decrease in soil moisture would quickly be refilled. The same happens during the rainy season at the Majadas pixel, see Figure 9 in the manuscript.
- These mountaineous areas have steep slopes, the surface water balance is therefore dominated by runoff and ET has a relatively lower role.

- The areas are cold and therefore feature a low atmospheric water demand. ET therefore is also low, especially compared to runoff.

We will add the following in l. 319:

In areas where growing-season LAI is decreased by the bias-blind DA the response differs. In the Nile delta, the increase in LAI leads to a reduction in soil moisture via transpiration. In the Alps and Scandinavia, soil moisture is not affected systematically, since the water balance is dominated by runoff, and transpiration changes therefore have a relatively lower impact.

1.13

L324-325: The anomaly correlation of GPP improves a lot for places where LAI bias is large. Isn't it persuasive that the raw LAI observation plays an essential role in constraining the inter-annual changes in vegetation growth? The bias-aware approach may limit such benefit.

Agreed. As indicated above, we will reformulate our conclusions to point this out.

1.14

L334: It is not clear to me what area has been covered in the "southern part of the domain". I would suggest mark it out in Figure 3 or simply provide a mask map along with Figure 4 to highlight area of analysis.

This refers to "the high-bias regions in the southern part of the domain", which we define as "all model grid cells south of 42° where the relative LAI difference is lower than -30%" (see caption to Figure 4 in the manuscript). We can add a figure marking these areas in a supplement.

1.15

ET comparison: Considering large uncertainty in ET products, would it be valuable to include a few other ET datasets for comparison besides GLEAM ET?

We agree that more reference data sets are always better, but (i) we did not want to extend the already quite long manuscript, and (ii) the amount of information that can be obtained from the comparisons to satellite or model derived ET datasets is limited, especially if we want to evaluate the effects of bias, since the bias of all of these with respect to the truth is unknown. Since changes in ET are lower than in GPP, we decided to do a more detailed evaluation with GPP instead. If deemed useful by the editor, we can include a more detailed comparison.

1.16

Figure 9: The seasonal cycle of the LAI observation would be very different if scaled by CDF or seasonal factors. I highly doubt whether such scaling reduces the value of using LAI observation to constrain the model performance. Scaling the observation to a model climatology that is further away from the truth is another form of bias. This might explain partly that the two bias-aware DA does not lead to better improvements for GPP as compared to the bias-blind DA.

We fully agree with these sentiments, and will adapt the discussion & conclusion as detailed above.

1.17

L387: About the Majadas site: I'd suggest informing where the site is located in Figure 5.

We will update Figure 5 with locations of the Majadas site and the selected Nile delta example pixel.

1.18

L375-378: Repeated sentence. And I think such sawtooth pattern could be taken care of if enables the temporal interpolation to the observations.

The temporal interpolation would at least visually remove the sawtooth pattern with daily output, but effects like the unphysical short-term dynamics of LAI and GPP would still be at play. However, we will also mention this in the revised discussion giving more space to the bias-blind DA.

1.19

Section 3.4: Again I have concerns in the scale mismatch in the modeling space vs. the observation. Without consider such effect, the conclusion may be biased. Would it be possible to conduct a sensitivity analysis regarding the spatial resolution of the model? Alternatively, is it possible to collect more in situ observation sites that can better jointly represent the condition for a 0.25 degree grid cell?

Unfortunately we do not have more in situ observation available, but we will instead restrict the analysis to stations that are deemed representative at the coarse scale.

1.20

L405-408: Good point! How does the other two bias-aware DA perform in this case?

We did not add the bias-aware DA results in these plots because they are very similar to the OL. If requested, we can add them here or in a supplement.

1.21

L426-427: True but the seasonal scaling may introduce additional bias to the scaled observations if the modelled climatology is further away from the ground truth. I think this is probably a larger flaw as compared to the bias-blind DA approach because at least the later remains the true temporal variation of what has been observed.

We agree that there is a trade-off between staying close to the observations and having a well-behaved filter algorithm with a stable model. We discuss this in more detail in the proposed revisions above.

1.22

Section 4.2 - 4.3: I agree that if large bias exists then the bias-blind DA may lead to misuse of Kalman filter; but I also think that the rescaling of observation based on a biased model is a big issue, it may create even bigger problem if the seasonality is deteriorated. I think the benefit and pitfalls for both approaches should be elaborated and discussed in a more comprehensive manner before it gets to section 4.4.

We again refer to the changes to the discussion and conclusion mentioned above.