Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC2
https://doi.org/10.5194/bg-2021-355-RC2, 2022

Comment on bg-2021-355

Anonymous Referee #2

Referee comment on "Local scale evaluation of the simulated interactions between energy, water and vegetation in land surface models" by Jan De Pue et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-355-RC2, 2022

Review on "Local scale evaluation of the simulated interactions between energy, water and vegetation in land surface models" by Jan De Pue et al., submitted to Biogeosciences

Three models (prognostic models ISBA and ORCHIDEE and a diagnostic model rooted in satellite remote sensing product development) were run on the meteorological and land surface forcing data of 56 eddy-covariance sites to compare their results to each other and to the site observations. Results were compared with multiple strategies (including bias vs. RMSE, Taylor diagrams, sensitivities between state variables per land use type, error correlations, phenology and seasonal cycles) and results were used to identify (where possible) or hyothesize (otherwise) weaknesses of the different models, also considering uncertainties in the observation data. Main findings include a better performance of the diagnostic as compared to teh prognostic models, a convergence in strengths and weaknesses between both prognostic models partly due to latest updates of the ISBA model, but also remaining differences, and recommendations on most important future improvements (in particular related to drought stress response, phenology and biomass allocation). The manuscript is very well written, methods and results are presented in clarity, the subject is relevant to Biogeosciences, and original among others in the sense that the newest version of ISBA and a comparatively new data product of teh eddy-covariance network are used.

Notwithstanding a lack of expertise on my side when it comes to internal details of the used models, which would ideally be addressed by other reviewers, I recommend the manuscript for publication after minor revisions suggested below.

Detailed Comments:

Title: Consider adding "three" before land surface models, currently is lets readers easily think of a large multi-model study.

Figure 1: Relation and feedbacks in the caption is no distinction that makes it easy to understand for the reader - relation could also be something purely empirical but here apparently you mean it as "more direct / stronger / first order" than the feedbacks. Check if this distinction is really needed and if, which other words could stress it. Next line of the caption, from the arrow it would be better to write Soil moisture - LAI than vice versa. Figure itself: Would it make sense to add stomatal control somewhere in the middle? The way it is now it seems like LE and GPP are each controlled independently by soil moisture and LAI, such that the reader is almost wondering why there is not also an arrow between them.

L97: "corrected manually to represent the tower footprint area" is somewhat unclear, could you be more specific?

L99: "linearly interpolated" refers to the ERA5 data being hourly and the tower observations mostly half-hourly?

L140 & 159: Could the free drainage at the bottom explain the oversensitivity to drought stress? (To be discussed not here)

L161 & 203: Just mentioning "a selection of sites [...] to ensure adequate data quality" is a bit arbitrary

L165: Again of course not to be discussed here, could the non-specified management practices be an important explanation for the difference between diagnostic and prognostic model(s) in crop sites (e.g. Figure 3)?

L168: Were the PFT and vegetation type info derived from the IGBP metadata of the flux network, or from remote sensing, or other sources?

Table 2: Some sites (apparently especially crop sites, e.g. BE-Lon, DE-Kli and DE-RuS) are listed with very large LE corrections, that do not match what I thought I knew from past studies on their energy balance closure. I tried a detailed check on DE-RuS: The flux-weighted effective average factor between LE_corr and LE is 1.44, somewhat lower (why) than the 1.47 in Table 2, but still far too high compared to any energy balance analysis carried out for this site in the past (e.g. 1.18 would result from Eder et al. 2015, DOI: 10.1175/JAMC-D-14-0140.1 which focused on summer months and 1.23 from Graf et al. 2020, doi.org/10.1098/rstb.2019.0524 with a study period matching the drough2018

dataset). Note that the current OneFlux product does to my knowledge not use details such as heat flux plate depth important to compute the energy balance closure, however even considering this the difference seems far too large, so it seems that LE_corr from this dataset should be used with care.

L182: Actually ICOS does have a standardized setup for soil moisture; however, the used dataset (drought2018) still mostly consists of so-called "legacy data" (i.e. voluntarily provided measurements with pre-ICOS set-ups). No need to mention it, just avoid the misleading wording.

L196: From the way it is mentioned for LE and H and then a new paragraph starts, no EBC-based correction was assumed for NEE (and propagated to GPP and Reco)? Not that I would like to recommend it, just for clarity. Unfortunately even the correctability for LE and H is far from certain, but then depending on the assumed reasons it may or may not also apply to the CO2 flux (at least its turbulent part before WPL correction). It is nothing that can be done in a more certain way, but it is important to be aware of it later e.g. when LE and NEE show different model-observation biases. P.S.: It is nicely mentioned already in line 394, but may still leave the reader wondering here.

L221: Make clearer if the mean annual cycles are computed one per site across all its site-years (which implies that the deviations also include interannual variability, which is not a bad thing but one to be aware of)

L226-235 and Figures 5+6: Better explain for what the slope and for what the correlation was used. Comparing the text to the figure captions, I guess that the "Spearman slope" in the caption is wrong (slope yes but probably not between the rank-transformed variables, which is what "Spearman" would imply to me)?

L243: Maybe adding "independently" to the last sentence and mentioning it already at the start of subsection 2.3.2 would make it easier to understand.

L247: Here it is unclear whether the LE partitioning methods are just mentioned out of interest, or were applied in this study (which seems not to be the case according to the result section).

L255: Shouldn't this be visible in Fig. 2a? If accuracy corresponds with the bias (x axis) and precision with random errors (y axis), it would be more accurate to state that both models have the same accuracy but ISBA a slightly better precision. Sometimes accuracy is also used as a combined name corresponding to both, systematic and random errors; then the statement is true but imprecise and the "significantly" seems a bit overstated (unless it refers to a successful statistical significance test of course).

Around L350, Figure 12: Are differences in the partitioning between drainage and runoff really interesting to discuss for models which were all run in uncoupled 1D mode? My (maybe wrong) expectation would be that it is a quite arbitrary function of model physics that only converges between models if horizontal neighbours with given slopes get a chance to communicate with each other.

L389: "caused by surface" looks a bit as if something was missed out here, maybe "... surface heterogeneities"?

L396: Mentioning GPP and LE alongside each other with parenthesis does not fully capture the extend of the problem (see also comment on L196): Since LE was corrected for EBC non-closure (at least tried to, given the open questions correctly mentioned by the authors) while GPP was not, it could somewhat be expected that the mean difference model vs. obs is smaller (or more negative) for LE and larger (or zero or less negative) for GPP. This is exactly what we see in Fig. 2 (if the x axis is model - observation). Which might indicate (among other ways to explain more associated with model shortcomings of course) that the LE is overcorrected even by the current EBC correction. Note that to my knowledge (if I didn't overlook something) Gebler et al. (2015) do not report a better EC-lysimeter match by putting the whole deficit into LE, but by a correction conserving the evaporative fraction, which is similar to the Bowen ratio conserving correction by Pastorello et al. Others even suggest that most or all of the deficit might be related to sensible heat (Ingwersen et al. 2011, https://doi.org/10.1016/j.agrformet.2010.11.010), or found a good match with independent reference data without LE correction (e.g. Graf et al. 2014, https://doi.org/10.1002/2013WR014516 for the catchment water budget method). In general, a problem with the body of existing comparisons of eddy-covariance fluxes to independent reference methods is that the latter can have their own systematic errors (e.g. island effect in case of lysimeters, or different footprints of both systems) on a similar order of magnitude as the eddy-covariance energy balance closure gap, and that the (often quite definite) answers of the single studies are in conflict when comparing these studies with each other. Maybe (especially given the risk of a too large energy balance gap seen by the flux product as discussed in comment on table 2) it would even be interesting to see how the model-observation match without the energy balance correction is. Of course, the results would not completely reliably indicate an overcorrection / different source of the closure gap, but could also point to an unintended adaption of the models towards uncorrected eddy-covariance data during past validations.

L399-407 (4.1.2): Almost all differences discussed here could also be due to the different management intensity between forests and herbaceous, the latter including the crop sites (I guess?) and also intensively managed grassland. The prognostic models were not informed about management (e.g. which crop, when bare soil), while for the diagnostic model some of teh effects of management may have been implicit in the data provided.

L436: Do not understand why this (slow buildup of LAI in early season) should be a consequence of the sentence before (assimilated carbon invested into leaves first). Do you maybe mean the same thing you describe more understandably in the next paragraph, i.e. that a process is missing in ISBA which can grow leaves from stored biomass?

L516-519: Maybe for readers jumping to the conclusions section it would be helpful to give a brief hint on the most important process(es) underrepresented, e.g. as discussed around L446-452 (where does the biomass for new leaves come from at season start).

Acknowledgement: despite much praise (“This work stands on the shoulders…”) and correctly citing the DOI of the drought2018 data product, the attribution of the work at least of the flux site PIs offering the flux data is a bit awkward. The explanation attached to data policy states that PIs should be contacted before publication (to learn about possible acknowledgement requirements, or in extreme cases offer the possibility to scientifically contribute to the study) at least in case the data play a very substantial role for the publication. It might be argued that the latter is the case here. I am well aware that the current situation is unsatisfying for both sides (study authors cannot continue forever to ask hundreds of data authors for each multi-site synthesis, which would delay scientific progress and encourage bagatelle coauthorships; but the latter still often feel incompletely compensated for their voluntary work, given that in most countries they are unfortunately not paid for the site servicing and raw data processing the way weather service employees are, but for science, often on non-permanent contracts, from which they divert worktime for the data production), and do not suggest to revise the communication workflow for this study, but would like to remind the authors and community of it for future studies - at least until either DOI citations have become a highly valued measure of recognition, or data providers are mainly employed to provide free data and most of the data processing including raw data to flux processing has been taken over by the central facilities of the next-generation networks.

Purely Technical Comments:

L23: not sure "to better" is good English, maybe "improve"?

L65 & 105: Check if "remote sensed" works, maybe "remotely sensed" or "remote-sensing based" (could also be satellite based if it is exclusively satellites)

L307: "In" missing before "Fig. 6"

L330: error*s*?

L362: frequent*ly*

L367: check usage of "in/on(?) the one / other hand"

L467: Blank missing at start of new sentence.

L470: Maybe replace "Which" by "This"

L471: "wears many hats", just like the above, maybe a bit too "oral" style.

L479: "shows" instead of "learns"?