Reviewer 1

The manuscript Peano et al. compared results from seven land surface models (LSMs) to remote sensing estimated leaf area index (LAI) at global scale regarding the seasonal timing of LAI trough (lowest value) and peak (highest value) and amplitude (max. – min.). The results indicate that all studied LSMs tended to show delayed timing of LAI trough and peak, in particular for the Northern Hemisphere while the modeled amplitude is smaller than satellite estimates. While overall the work is pretty straightforward, I do have a few major comments.

We thank the reviewer for reading the manuscript and providing her/his comments.

1. Model description of phenology scheme (L85-125) and related discussions do not provide enough details for readers who are not familiar with these models or with the different phenology schemes or both to understand what might be contributing to the current variations in model simulated LAI and its seasonal dynamics. While most models tend to show delayed timing of LAI trough and peak, there's one model, LPG-GUESS did a better job capturing this (L170, L180). Why so? Similarly, when model differences, model-data discrepancy across different biomes/regions all present, it would be interesting to discuss the potential causes why we see all these differences. While the discussion tried to talk about the different sources of variability (L280) and potential observation biases (L295), it's quite thin and not well organized to address the observed spatiotemporal data-model discrepancy. I think expanding discussion and organize it in a similar way as how the results are organized, and maybe also add a little more detail on different phenology schemes in each LSM can be very helpful. It would also be nice to see some suggestions for future model development or priority areas for data collections for better model parameterization or better observations for model benchmarking.

We thank the reviewer for this comment. The companion paper (Peano et al., 2021) provides in Table 1 a description of the main differences in the phenology scheme used by the seven LSMs. The mentioned table is split into two sections: the first provides the main vegetation features of the LSMs, and the latter focuses on describing the drivers of the start and end of the growing season. We will add the following table in this manuscript which is similar to Table 1 in Peano et al (2021) but includes references on the computation of LAI, which we hope addresses the comment:

LSM	Original Resolution	PFT	CFT	Phenology schemes	LAI driver	LAI Reference
CLM4.5	1.25° x 0.9375°	15	1 C3	Evergreen; Seasonal Deciduous; Stress Deciduous	Leaf Carbon Specific Leaf Area	Thornton and Zimmermann (2007)
CLM5.0	0.5° x 0.5°	15	2 C3	Evergreen; Seasonal Deciduous; Stress Deciduous	Leaf Carbon Specific Leaf Area	Thornton and Zimmermann (2007)
JULES-ES	1.875° x 1.25°	13	1 C3, 1 C4	Deciduous Trees	Balanced LAI Temperature	Cox (2001)
JSBACH	1.9° x 1.9°	12	1 C3, 1 C4	Evergreen; summergreen; raingreen; grasses; tropical crops; extratropical crops	Maximum LAI Temperature Soil Moisture Net Primary Productivity	Dalmonech et al. (2015) Böttcher et al. (2016)
LPJ-GUESS	0.5° x 0.5°	25	3 C3, 2 C4	Evergreen; Seasonal Deciduous; Stress Deciduous	Specific Leaf Area Leaf Biomass	Reich et al. (1992) Smith et al. (2014)
ORCHIDEE	0.5° x 0.5°	15	1 C3, 1 C4	Deciduous; dry and semi-arid; grasses and crops	Temperature	Polcher (1994) Krinner et al. (2005)
ISBA-CTRIP	1° x 1°	16	1 C3, 1 C4	Leaf biomass	Leaf Biomass Specific Leaf Area Leaf Nitrogen	Delire et al. (2020) Gibelin et al. (2006)

Table 1: Grid spatial resolution used for each land surface model (LSM) and references for their principal features about Phenology and Leaf Area Index (LAI) computations. PFT stands for plant functional type, and CFT stands for crop functional type.

Following the reviewer's suggestion, we will increase the discussion on the differences between LSMs by expanding the discussion in the result section and section 4.3. Furthermore, we will add suggestions for model development and missing observations in the discussion sections.

2. How much of the data-model discrepancy in LAI seasonality can be contributed to model structure and parameterization respectively? I know this is not the main focus of this study, but I think it's very important to know this before we conclude that we need better phenology models (specifically refer to model structure improvement). Related, it is not clear to me how simulated distribution of the different PFTs (evergreen vs deciduous: they vary in phenology in addition to many other features that can influence plant growth and mortality thus phenology) can influence the modeled LAI seasonality both spatially

and temporally. Is it possible that the delayed timing of LAI trough and LAI peak is a mismatch between the observed and the simulated vegetation type that dominate a particular grid cell?

The comparison between the two versions of the Community Land Model, namely CLM4.5 and CLM5.0, supplies a “quality assessment” of the contribution of model structure versus model parameterization. The two LSMs apply the same phenology scheme (= model parameterization) but differ in representing other soil and vegetation processes (= model structure). In this case, it is possible to see differences in the simulated magnitude of the Leaf Area Index but limited differences in timings.

The difference in PFT distribution between models and observations is another potential source of discrepancies. This factor has a limited impact on the intra-models comparison since all the LSMs derive their PFT distribution from the same original dataset (i.e. Land Use Harmonization version 2, LUH2). However, differences in the amount of resolved PFTs can impact the final result. Similarly, LUH2 should account for observed PFT distribution, mitigating this source of discrepancy.

We will refer to these points in the revised version of the manuscript.

Specific comments

L20: add relevant citations.

We will add the reference to Running and Coughlan (1988), Bonan et al. (2012), Murray-Tortarolo et al. (2013), Fang et al. (2019).

L40: maybe can expand to add a little more detail here.

We thank the reviewer for the suggestion. We will expand the introduction to satellite datasets and their applications.

L140: how the different domain resolution might influence model results? Also how coarse is the model resolution?

The original resolution used by each LSM will be added to the novel table 1 (see reply to point 1 above). The seven LSMs have an original resolution spanning from half-degree to almost two degrees in horizontal resolution. This discrepancy in the original resolution may influence the representation of vegetation heterogeneity and will be part of the results' discussion.

L168: multi-model ensemble mean

We thank the reviewer for pointing at this typo.

References

Bonan, G. B., Oleson, K. W., Fisher, R. A., Lasslop, G., and Reichstein, M.: Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4, J. Geophys. Res., 117, G02026, doi:10.1029/2011JG001913, 2012.

Böttcher, K., Markkanen, T., Thum, T., Aalto, T., Aurela, M., Reick, C., Kolari, P., Arslan, A., and Pulliainen, J.: Evaluating biosphere model estimates of the start of the vegetation active season in boreal forests by satellite observations, Remote Sensing, 8, 580, https://doi.org/10.3390/rs8070580, 2016.

Cox, P. M.: Description of the TRIFFID Dynamic Global Vegetation Model, Hadley Centre Technical Note 24, Hadley Centre, Met Office, Bracknell, UK, 2001.

Dalmonech, D., Zaehle, S., Schürmann, G., Brovkin, V., Reick, C. H., and Schnur, R.: Separation of the effects of land and cli- mate model errors on simulated contemporary land carbon cycle trends in the MPI Earth System Model version 1, J. Climate, 28, 272–291, https://doi.org/10.1175/JCLI-D-13-00593.1, 2015.

Delire C., Séférian, R., Decharme, B., Alkama, R., Cal- vet, J.-C., Carrer, D., Gibelin, A.-L., Joetzjer, E., Morel, X., Rocher, M., and Tzanos, D.: The global land carbon cycle simulated with ISBA: improvements over the last decade, J. Adv. Model. Earth Sy., 12, e2019MS001886, https://doi.org/10.1029/2019MS001886, 2020.

Fang, H., Baret, F., Plummer, S., and Schaepman-Strub, G.: An overview of global leaf area index (LAI): methods, products, validation, and applications, Reviews of Geophysics, 57, 739–799, https://doi.org/10.1029/2018RG000608, 2019.

Gibelin, A.‐L., Calvet, J.‐C., Roujean, J.‐L., Jarlan, L., & Los, S. (2006). Ability of the land surface model ISBA‐A‐gs to simulate leaf area index at the global scale: Comparison with satellites products. Journal of Geophysical Research, 111, D18102. https://doi.org/10.1029/2005JD006691

Krinner, G. and Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, GB1015, https://doi.org/10.1029/2003GB002199, 2005.

Murray-Tortarolo, G., Anav, A., Friedlingstein, P., Sitch, S., Piao, S., Zhu, Z., et al. (2013). Evaluation of land surface models in reproducing satellite-derived LAI, over the high-latitude Northern Hemisphere. part I: Uncoupled DGVMs. Remote Sensing, 5, 4819–4838. https:// doi.org/10.3390/rs5104819

Peano, D., Hemming, D., Materia, S., Delire, C., Fan, Y., Joetzjer, E., Lee, H., Nabel, J. E. M. S., Park, T., Peylin, P., Wårlind, D., Wiltshire, A., and Zaehle, S.: Plant phenology evaluation of CRESCENDO land surface models – Part 1: Start and end of the growing season, Biogeosciences, 18, 2405–2428, https://doi.org/10.5194/bg-18-2405-2021, 2021.

Polcher, J. (1994), Etude de la sensibilite´ du climat tropical a` la de´forestation, The`se de doctorat, Univ. Pierre et Marie Curie, Paris.

Reich, P. B., Walters, M. B., and Ellsworth, D. S.: Leaf life-span in relation to leaf, plant and stand characteristics among diverse ecosystems, Ecol. Monogr., 62, 365–392, 1992.

Running, S.W.; Coughlan, J.C. A general model of forest ecosystem processes for regional applications I. Hydrologic balance, canopy gas exchange and primary production processes. Ecol. Model. 1988, 42, 125–154.

Smith, B., Wårlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cy- cling and N limitations on primary production in an individual- based dynamic vegetation model, Biogeosciences, 11, 2027– 2054, https://doi.org/10.5194/bg-11-2027-2014, 2014.

Thornton, P. E., and N. E. Zimmermann, 2007: An Improved Canopy Integration Scheme for a Land Surface Model with Prognostic Canopy Structure. J. Climate, 20, 3902–3923, https://doi.org/10.1175/JCLI4222.1.