Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC1
https://doi.org/10.5194/gmd-2022-291-RC1, 2023

Comment on gmd-2022-291

Jinfeng Chang (Referee)

Referee comment on "LandInG 1.0: a toolbox to derive input datasets for terrestrial ecosystem modelling at variable resolutions from heterogeneous sources" by Sebastian Ostberg et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-291-RC1, 2023

This is a comprehensive manuscript that described a toolbox for generating commonly used input datasets for terrestrial ecosystem modelling at two spatial resolution 5’ and 30’. The generated datasets include static inputs like land-sea mask, country and region mask, soil texture and pH, river routing, grid locations of lakes, rivers, dams and reservoirs, and dynamic inputs, a harmonized gridded annual land use and land management (irrigation and fertilization) for the historical period 1500-2017. The application of this toolbox for generating input datasets for LPJmL was presented as an example. The manuscript is well structed and very well written. I would think it is a valuable effort to facilitate the input generation. I only have a few suggestions as follows.

- Given the fact that 1) most of the source datasets existed or used in this toolbox has the highest resolution of 5 arc minutes, 2) the spatial resolution of the TEM simulation usually (if not all) depend on the coarse resolution of all input datasets, and 3) in many cases of this toolbox, the aggregation can only be done with an integer multiple of the source resolution, it could be better to give the possible resolutions for each of the input datasets.
- For all the datasets, it is essential to provide not only the reference, but also the link to the source datasets, the access date (as datasets can be updated), the original data format, and the data content (e.g., exact variable name used by the toolbox). Otherwise, it makes the toolbox much more difficult and less useful for users.
- It is understandable that the authors did not provide results on the created gridded maps as it might contain source datasets that require a license to publish. But for those datasets that were publicly available and has been licensed to distribute, it would be better to provide the resulted maps in addition to the code. For the resulted gridded land use and land management dataset, in particular, the strategy described in this manuscript is in some sense novel (or at least comprehensively described for the first time). Though the authors claimed that this manuscript is solely a description of the toolbox, putting the gridded land use and land management dataset into a public repository could be useful for the community.