In this study, Metzler and his colleagues provide a very important and interesting approach to benchmark global land carbon-cycle models based on radiocarbon data. This approach could be a great contribution to connect the research communities of Earth system modeling and radiocarbon. I have carefully studied their mathematic assumptions and derivation. I found their work is solid, though the writing is very technical for both Earth system modelers and empirical scientists who use radiocarbon as their major research tools. Thus, I only have some broad suggestions on discussion and a few specific suggestions on the details. Overall, I recommend JAMES to publish this work with some minor revisions.

Page 13, Lines 278-292: The authors used the accelerated-decomposition approach to spin up their model. The "1000yr+400yr" spinup time seems common for a global model, but it might be careful for a global model with 10 vertical layers. The criterion of <0.01% yr^{-1} change usually cannot constrain the steady state of passive SOC pool in deep soils, which has a large pool size but very low turnover rate. This issue must has a small impact on the results in this study, but may generate some errors when apply their approach for long-term approximation.

Page 15, Lines above 312: It is good to gradually introduce the approach from single to two compartment system. How about to give one or two real examples of natural ecosystems for the "two compartment system" here? For example, vegetation-soil could be a nice "two compartment system". I feel this may make these equations more understandable for an empirical scientist.

Page 24, Line 401-402: The authors suggest that using their proposed approach can avoid rewriting the model in matrix form. I agree that rewriting the models into matrix form is not necessary for the benchmarking analysis based on radiocarbon data. However, as I know, one goal of the traceability framework in Luo et al. (2017) and Xia et al. (2013) was to enhance the understandability of the spread results among different models. As also highlighted by some recent papers (e.g., Bonan et al. 2019GBC), understanding the difference between models itself becomes very important for model improvements. In my opinion, the approach in Metzler et al.'s paper has shown the great potential to use matrix approach to improve the Earth system models, even without the information of model structure. However, this approach could be more efficient if some additional outputs or information could be provided by the community of Earth system model. I highly recommend the authors to add some detailed suggestions here for the CMIP6 community to promote the application of the matrix approach.

Application of the Metzler et al.'s approach to CMIP6 models: the timestep of model output is usually monthly, and only a few variables have daily outputs. The bias of the matrix approximation is relatively larger in the surface soil layers as shown in Fig. 4, even with a 10-day output timestep. So, it would be great if the authors can show how their approach performs with the monthly outputs. If the bias is large, then a

recommendation of weekly output could make for the ESM community, at least for the historical runs.