

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
<https://doi.org/10.5194/hess-2022-284-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2022-284

Anonymous Referee #2

Referee comment on "Diagnosing modeling errors in global terrestrial water storage interannual variability" by Hoontaek Lee et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-284-RC2>, 2022

The authors presented a study to diagnose the modeling errors by comparing GRACE and model TWSA based on IAV. The motivation of this study is nice, since Scanlon's PNAS study revealed an interesting question on the discrepancy between GRACE and models. The focus on interannual is a good complementary to the focus on trend by Scanlon et al.. Generally, this study is interesting. However, I have some critical questions related to the methods used for analysis, which may largely affect the reliability of the findings.

- Generally, WGHM, PCR-GLOBWB, and maybe some other LSMs that include GW module, are more popularly used than the two models used in this study. I am not going to say the two models used here is not good enough, but I guess many researchers would be more interested on what will it like if we use WGHM, or PCR-GLOBWB, or CLSM. Besides, it is not clear how the including of GRACE in model parameter estimation and evaluation (Line 77) will impact the comparison between GRACE and the two models.
- It is not clear why using Equation (1) to derive the IAV for analysis. I cannot understand the physical meaning of subtracting long-term trend (fit ()) from monthly values. So, the question comes that what is interannual variability, and how to define it? Can we just subtracting long-term average from monthly values? I am not sure my understanding is correct or not. Please verify it.
- Since GRACE Level-3 data has been already processed by subtracting the mean of a period (2004-2009?) from monthly TWS to get TWSA. If the authors again do subtracting (2002-2017) for GRACE and models, it may lead to mismatch between GRACE and model, because different subtracting were done for GRACE (subtracting 2004-2009, and then subtracting 2002-2017) and models (subtracting 2002-2017).
- Line 128: I am not sure it is the best way to evaluate model performance by comparing the IAV derived from GRACE and models. How about compare TWSA?
- Before Figure 2, people may be interested on seeing spatial distribution map of TWSA from GRACE and models, as well as the distribution map of IAV, which both can help we better understand the difference and consistence between GRACE and models.
- Figure 3: Sorry, but I do feel difficult to understand what the exact meanings of the spatial maps are. Maybe more information can be added to the figure showing who minus who, something like that. Besides, I guess the white blank areas here are the

grid cells with positive covariances, is it true?