Author’s response to Referee #3

This study models the error between the global TWS IAV observations of GRACE and two models, SINDBAD and H2M. The authors found that the global TWS IAV is mainly driven by humid tropical and semi-arid regions, and identified the hotspots of modeling errors of the global TWS IAV mainly in tropical regions that span across climatic regions. The study presents a novel way to attribute global variability to each pixel and focused on regions where hydrological cycle components in models may not be sufficiently well represented due to their complex hydrological and climatological processes.

The study in general is well-written and easy to follow.

AC: We would like to express our gratitude to the reviewer for positive feedback and suggestions on the manuscript. Of course, we have responded to the other reviewers and will incorporate all suggestions into the revised manuscript. Below, you will find the responses to each comment of reviewer 3.

Additional to comments made by the two Anonymous Referees, which I consider important to answer, my comments are as follows:

- **As the study identifies humid regions of northern South Americas as one of the main drivers of global TWS IAV, I suggest including these references in the discussion in which global models are compared with GRACE products in a very important instrumented tropical basin.**

Bolaños Chavarría, S., Werner, M., Salazar, J. F., & Betancur, T. (2022). Benchmarking global hydrological and land surface models against GRACE in a medium-sized tropical basin. Hydrology and Earth System Sciences, 26(16), 4323-4344.

Bolaños, S., Salazar, J. F., Betancur, T., & Werner, M. (2021). GRACE reveals depletion of water storage in northwestern South America between ENSO extremes. Journal of Hydrology, 596, 125687.

AC: Thank you for pointing to the relevant studies. We will include them in the discussion of revised manuscript as:

On the other hand, tropical regions come out as the dominant contributor to the variance of the global TWS IAV modeling errors (Fig. 4). Tropical regions were reported as one significant contributor to the global TWS IAV, but with a large disparity between the models and GRACE (Humphrey et al., 2018) <u>due possibly to characteristics of the regions that the tested models do not properly account for, for example, artificial reservoirs, complex topography, and wetlands (Bolaños et al., 2021, 2022)</u>. [...]

- **I am a bit confused with Equation 1, in figure 1 I think it is clear that TWS IAV is the result of detrending and deseasonalizing TWS, but in Equation 1, I understand that only TWS is deseasonalized.**

AC: Yes, as in Fig. 1 in the manuscript, Eq. (1) deseasonalizes and detrends TWS as the linear fitting includes the trend of the month of a year across years. Eq. (1) could detrend as well because each regression line of a month includes the trend as Fig. 1 illustrates below. We will include Fig. 1 in the revised manuscript as appendix.

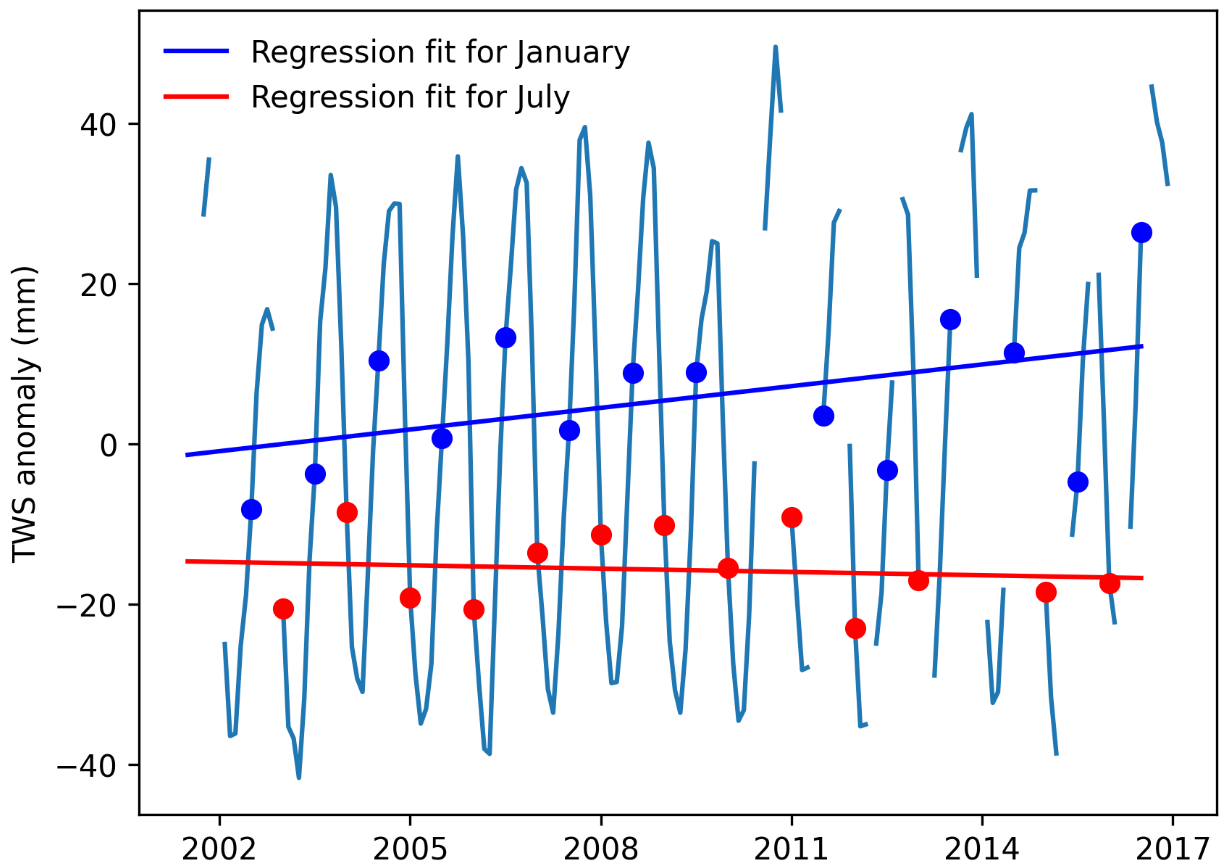

Figure 1. Illustration of the calculation of interannual variability for the global terrestrial water storage (TWS) anomalies.

- **I think is necessary to define what is the meaning of SREX Regions, I don't identify what is.**

AC: We will add the meaning of SREX in the manuscript as follows:

After the error hotspots are identified, we compare the time series of TWS and precipitation IAVs at the regional scale for error hotspots within selected *<u>Intergovernmental Panel on Climate Change (IPCC) Special Report on Extremes (SREX) regions</u>* *(Sect. 3.4; see Fig. B1 for the SREX regions) to diagnose TWS IAV errors. [...]. Note that SREX regions include different regions of the world, and they have been used extensively to diagnose regional variation of climate model simulations (e.g., Seneviratne et al., 2012; Pokhrel et al., 2021).*

and in the caption of Fig. B2 of the manuscript as follows:

Figure B1. *<u>The Intergovernmental Panel on Climate Change (IPCC) Special Report on Extremes (SREX) regions</u>* *(Seneviratne et al., 2012), which were used in figures (Figs. 6, B4, B6, B7, and B8) to spatially average global terrestrial water storage interannual variability time series into four selected regions: the Laurentian Great Lakes (SREX regions 5), Amazon (SREX regions 7), Eastern and Western Africa (SREX regions 15 and 16), and South Asia (SREX regions 23).*

- **Why the preference for the JPL mascon if there is another mascon product like the mascon CSR that has the same resolution?**

AC: The purpose of this study is to qualitatively diagnose the hotspots of the global TWS IAV and its modeling error. For this purpose, either JPL mass concentration (mascon) product or CSR mascon product can be used, as JPL mascon and CSR mascon are qualitatively comparable to each other across global basins and at the interannual scale as well as longer-term temporal scales (Scanlon et al., 2016).

Nevertheless, to further clarify this, we analyzed the spatial contribution to the global TWS IAV for RL06 version 1 mascon GRACE products by JPL (i.e., one used in this study) and CSR (i.e., a comparable one by CSR). We find that the two GRACE products are largely consistent with each other in terms of hotspots of the dynamics of global TWS IAV (Fig. 2). This suggests that the main findings of our study would not change even if a different GRACE data product was used.

Scanlon, B. R., Zhang, Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., Longuevergne, L., and Chen, J. (2016), Global evaluation of new GRACE mascon products for hydrologic applications, *Water Resour. Res.*, 52, 9412– 9429, doi:10.1002/2016WR019494.

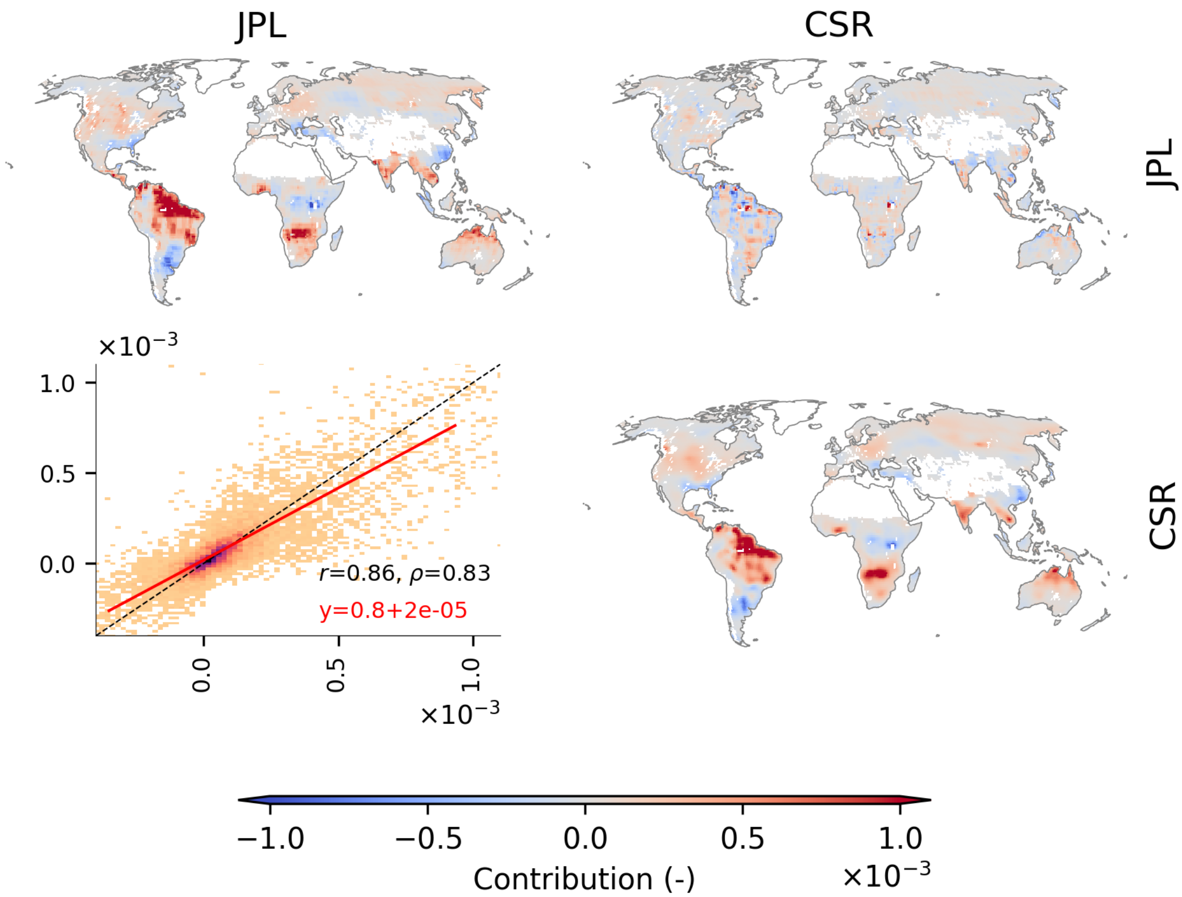

Figure 2. **The same as Fig. 3 in the manuscript, but using two mascon GRACE products only.** Global distribution of pixel-wise contributions to the variance of the global terrestrial water storage (TWS) IAV. Along the diagonal, maps of the pixel-wise contribution in GRACE by JPL and GRACE by CSR are shown. Above the diagonal, maps of the difference (i.e., column - row) are shown. Below the diagonal, scatter plots comparing the corresponding column (x-axis) versus row (y-axis) are shown. In the scatter plots, colors indicate the density of points, r is the Pearson correlation coefficient, and ρ is the Spearman correlation coefficient. Red lines are linear regression fit and red texts are corresponding equations.

Save, H., S. Bettadpur, and B.D. Tapley (2016), High resolution CSR GRACE RL05 mascons, J. Geophys. Res. Solid Earth, 121, doi:10.1002/2016JB013007.

- **Figure 2 a) describes a "NSE is the Nash-Sutcliffe Efficiency", but it does not appear in the figure**

AC: We are sorry for the error. We will remove the wrong reference from the caption as follows:

Figure 2. [...] R^2 statistics in the bottomleft is calculated as the square of the Pearson correlation coefficient~~; NSE is the Nash-Sutcliffe Efficiency~~. [...].

We will correct other errors as well, for example, the second text of Fig. 2a, from R^2(GRACE, SINDBAD) to R^2(GRACE, H2M).