Response to Reviewer #2

The authors presented a well-written paper about a high-quality study in which they increased CABLE LSM resolution and assimilated the GRACE data into the model to improve its accuracy. Overall, I am satisfied with the quality of the paper. Some minor revisions can help to increase the quality of the paper. Below you can find my suggestions in this regard.

We greatly appreciate Reviewer #2's positive feedbacks and constructive comments. We provided our responses (R, in blue) to each of the reviewer's comments (C, in black) below:

C1: 1) Line 45 and Line 115 LSM model: You used CABLE LSM in your study. What is the advantage of CABLE over other LSMs? Apart from being frequently used in Australia (your study area), what are other reasons/motivations for choosing CABLE as the main LSM in your study?

R1: We thank Reviewer #2 for addressing our unclear explanation. Please note that CABLE is a global model and has been used to simulate global storage and fluxes (e.g., Decker et al., 2015; Haverd et al., 2018). This study uses Australia as a case study due to the availability of in situ data for validation. Despite the small community size, CABLE has been updated regularly by the CABLE community to catch up with the state-of-the-art development (e.g., Decker et al., 2015; Heverd et al., 2018). However, the current 0.5°×0.5° model spatial resolution limits its application to TWS studies in large river basins, and the attempt to improve CABLE spatial details has not been considered thus far. This is in contrast to other models, in which a high-resolution version has already been developed. Our development of CABLE 0.05° aims to narrow this gap.

To clarity this, we will rewrite our opening statement (in the introduction section) as follows:

The Community Atmosphere Biosphere Land Exchange (CABLE; Kowalczyk et al., 2006) is an open-source global LSM developed and updated by the community. CABLE is a core LSM of the Australian Community Climate and Earth System Simulator (ACCESS, Bi et al., 2013; Kowalczyk et al., 2013) that can be used to simulate water storage and fluxes globally. The model has been regularly updated to incorporate the state-of-the-art model physics (e.g., Decker et al. 2015; Ukkola et al., 2016; Heverd et al. 2018). Despite its success, CABLE's spatial scale is currently limited to 0.5° (~50 km) due to the 0.5° resolution of its parameters and forcing datasets. This contrasts with other global model developments, where high-resolution versions have already been developed (e.g., van Dijk et al., 2013; Sutanudjaja et al., 2018). CABLE and its inputs must be reconfigured to increase the spatial detail of TWS estimates for smaller-scale studies (e.g., 0.01° – 0.05°). Our effort to increase the regional or local study's spatial resolution should narrow this development gap and has not previously been implemented.

C2: 2) Line 141 and section 3.2.1: Resampling coarse data to high resolution, bears extra uncertainty. How do you deal with this additional uncertainty?

R2: The effect of up/downscaling is also included in our DA process. In our perturbation process, when the data are resampled, their errors are also adjusted based on an error propagation approach. The relationship between coarse and fine-scale error can be expressed as:

$$\sigma_c = \frac{1}{M}\sum_{h=1}^{M}\sum_{l=1}^{M}\sqrt{\sigma_{f\ \ hl}^{2}\exp\left(\frac{-\phi_{hl}^{2}}{2\phi_0^{2}}\right)}, \tag{1}$$

where σ_c and σ_f represents a coarse and fine-scale error, (h, l) is the index of a grid cell, M is the number of fine-scale grid cells used in resampling, ϕ is a spherical distance between grid cells, and ϕ_0 is the considered correlation length (e.g., a coarse-scale's grid size).

We understand that this error size might not perfectly represent the truth (which is unknown), but it represents a more realistic error that changes with the increased/decreased spatial resolution. For clarity, we will add the above explanation to Sect. 3.1.

In Sect. 3.2.1, the resample is applied only to overlay the model grid cells with satellite products before comparison. We did not perform an error analysis here. Please note that remote sensing products may also contain bias and do not necessarily represent the truth. The inter-comparison is not used as validation. It is only used to assess the consistency between two independent estimates. The statement can be found in lines 353 – 356:

… the remote sensing products may contain biases (caused by, e.g., background model, processing algorithm) and do not necessarily represent the truth. (Ground truth validation is performed in Sect. 4.3.) The inter-comparison performed in this section is only to assess the consistency between two independent estimates: model and satellite…

C3: 3) Section 3.1: Why did you choose 3D Ensemble Kalman Smoother in your study and did not choose other DA/smoother methods? Please provide few lines about the benefits and potential limitations of this DA method. Also, for completeness, please discuss briefly why a Smoother can be a better choice than a Filter in your study?

R3: The 3D EnKS is chosen for two reasons. One, it accounts for spatial correlations in model errors and observation errors. The latter are highly correlated at neighboring 0.5° × 0.5° or 0.05° × 0.05° grid cells. Two, EnKS does not require interpolation of the observations (as in Ensemble Kalman Filter (EnKF); Tangdamrongsub et al., 2015) and mitigates the spurious jump in water storage estimates caused by applying the updates at the end of the month only. The additional computational cost is small: handling large covariance matrices and running the model twice for each month. For clarity, we will add the above explanation to the revised paper.

C4: 4) Section 4.2: Please provide your reasoning/hypothesis on why GRACE DA impacts deep water storage more than other components of the utilized LSM. You stated that similar finding was reported in other studies. Do you see a similarity between those LSMs and CABLE that resulted in a similar impact of GRACE DA? Is this a location-specific finding (Australia) or can it be generalized to other regions?

R4: We thank Reviewer #2 for the suggestion. The gravity satellite (like GRACE) is sensitive to the Earth's mass variation, which is more significant in the deeper layer, e.g., groundwater. This explains the GRACE effectiveness in capturing low-frequency signals (e.g., long-term trend of GWS) seen in most GRACE literature (and this study). For clarity, we will include the above explanation in our revised paper:

… It dominates the entire water column in several basins (e.g., Indian Ocean, Lake Eyre, North West Plateau, South West Plateau). This behavior reflects the nature of GRACE: groundwater provides a majority of the seasonal changes to terrestrial water mass.

The similarity seen in the previous studies is the impact of GRACE DA on GWS estimates (e.g., a significant change in GWS caused by GRACE DA). We find that our statement might be unclear. As such, we revised our sentence to:

GRACE DA has been shown to significantly affect GWS in previous studies, e.g., Girotto et al. (2016), Tangdamrongsub et al. (2018), and Li et al. (2019).

The impact of GRACE DA on GWS was seen globally, e.g., Rhine River basin (Tangdamrongsub et al., 2015), Continental United States (e.g., Kumar et al., 2016), China (Tangdamrongsub et al., 2017), Australia (Tian et al., 2017), global (Li et al., 2019). Recent work by Yin et al. (2020) assimilated GRACE data into CABLE in North China Plain and also reported GRACE DA's positive impact of GWS. These studies confirmed a clear benefit of GRACE DA despite the different models used, DA configurations, or locations. However, in our case, it is difficult to make a conclusive comment on GRACE DA behavior globally based only on the results of our early development. For clarity, we will add a remark regarding the simulation performance in other regions. The following sentences will be added to the conclusion section.

This means TWS estimates can be reproduced with more spatial detail by CABLE 0.05 at locations outside the area studied here since high resolution forcing data and model parameters are available globally (or near globally). However, the performance of such simulations might differ from this study due to the uncertainty in model parameters and forcing data that vary with geolocations (e.g., Herold et al., 2017; Tifafi et al., 2018). This remark also applies to the performance of GRACE DA. Although the improvement of assimilating GRACE into CABLE is also seen in other regions, e.g., North-East China (Yin et al., 2020), it is still difficult to quantify the benefit of GRACE DA over global river basins based on these early developments of CABLE/GRACE DA. Validation is highly encouraged to ascertain the accuracy of TWS estimates when performing the simulation in other regions.

Additional reference:

Yin, W., Han, S.-C., Zheng, W., Yeo, I.-Y., Hu, L., Tangdamrongsub, N. and Ghobadi-Far, K.: Improved water storage estimates within the North China Plain by assimilating GRACE data into the CABLE model, J. Hydrol., 590, 125348, https://doi.org/10.1016/j.jhydrol.2020.125348, 2020.

C5: Earlier on line 78, you state " GRACE DA has shown positive impacts on most TWS components, including groundwater (e.g., Girotto et al., 2017; Nie et al., 2019), soil moisture (Jung et al., 2019), and snow (Kumar et al., 2016)." How do you reconcile this to what you found in your study.

R5: We thank Reviewer #2 for the suggestion. This discussion will be added to our results section as follows:

… Despite different study areas, LSMs, and validation data, our finding is in line with, e.g., Girotto et al. (2017) and Nie et al. (2019), who also found a significant impact of GRACE DA on GWS components.

… GRACE is sensitive to the low-frequency variation (originated from deeper stores) and cannot effectively capture SSM, which is dominated by a high-frequency signal (e.g., precipitation). As a result, GRACE DA is found to have a minor (or negative) impact on the top soil component in most GRACE DA studies (e.g., Li et al., 2012; Tian et al., 2017; Tangdamrongsub et al., 2020). The small impact on SSM estimates also agrees with Jung et al. (2019), who observed GRACE DA's small (or negative) impact over dry regions in West Africa.

C6: 5) Section Conclusion: What is the role of uncertainty of the CABLE model inputs on the DA results? In another word, if you used other public resources as the CABLE inputs, would you get different results out of DA-based models? You mentioned 250-m resolution SoilGrids data for future use. Based on my personal experience, I found SoilGrids data not very accurate in many locations. Why do you think by using this data you can improve your model? I would suggest that you discuss about it in the paper.

R6: We thank Reviewer #2 for the comment. The role of uncertainty in model inputs will be discussed (please see below). We also thank Reviewer #2 for addressing the soil map accuracy. We found that our wording might be unclear. We present the improved spatial details (not accuracy) in this context. However, we agree with Reviewer #2 that the impact of forcing data and parameter accuracy on model performance needs to be discussed. As such, we rewrite our conclusion as follows:

The enhanced CABLE model resolution developed in this study relies on improved parameter and forcing data. The land surface physics remains unchanged. The workflow can be adopted for other CABLE repositories or different LSM with only slight modifications, e.g., number of soil or vegetation types. This means TWS estimates can be reproduced with more spatial detail by CABLE 0.05° at locations outside the area studied here, since high resolution forcing data and model parameters are available globally (or near globally). However, the performance of such simulations might differ from this study due to the uncertainty in model parameters and forcing data that vary with geolocations (e.g., Herold et al., 2017; Tifafi et al., 2018). This remark also applies to the performance of GRACE DA. Although the improvement of assimilating GRACE into CABLE is also seen in other regions, e.g., North-East China (Yin et al., 2020), it is still difficult to quantify the benefit of GRACE DA over global river basins based on these early developments of CABLE/GRACE DA. Validation is highly encouraged to ascertain the accuracy of TWS estimates when performing the simulation in other regions.

Our development is only demonstrated between 1981 – 2012 due to the availability of the Princeton forcing data. Future development can consider extending the temporal record or further increasing the spatial resolution of TWS estimates. The timespan extension is feasible using reanalysis forcing data from the Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017). Despite a slightly coarser spatial resolution than the Princeton data, MERRA2 datasets would allow TWS simulations to be extended to the near present.

Note that we suggested a possible improvement of spatial resolution (not accuracy) using higher spatial-resolution data in our submitted manuscript. However, our suggestion might be too optimistic because sub-kilometer global forcing data needed for model simulation are not currently available. As such, we remove the statement regarding a sub-kilometer resolution to avoid confusion.

Additional references:

Herold, N., Behrangi, A. and Alexander, L. V.: Large uncertainties in observed daily precipitation extremes over land, J. Geophys. Res.: Atmospheres, 122(2), 668–681, https://doi.org/10.1002/2016JD025842, 2017.

Tifafi, M., Guenet, B. and Hatté, C.: Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France, Global Biogeochem. Cycles, 32(1), 42–56, https://doi.org/10.1002/2017GB005678, 2018.

C7: 6) Code availability: It would be more useful to the readers if you could share your code for the DA framework.

R7: We thank Reviewer #2 for the suggestion and particular interest in the software. Software development is already on the list of our research plans. Despite our very limited resources, we are trying hard to make the software available as soon as we can.

C8: 7) Title of the paper: GRACE Data Assimilation implies a DA method that is called GRACE. To avoid confusion for readers who are not familiar with GRACE mission, I would suggest using "Assimilation of GRACE data" instead.

R8: We agree with Reviewer #2. We modify the title to "Development and evaluation of 0.05° terrestrial water storage estimates using CABLE land surface model and assimilation of GRACE data" as Reviewer #2 suggested.

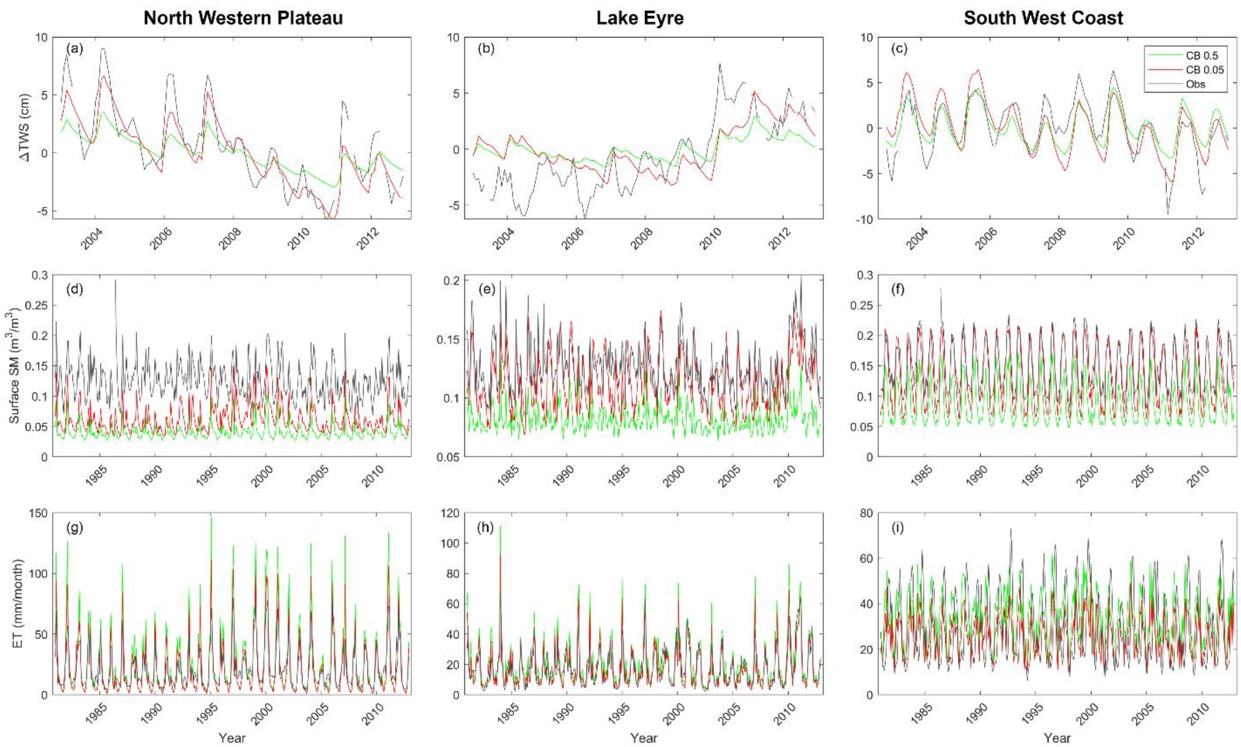
C9: 8) Table 2: grid size of Harmonized World Soil Data base is 30 arc-second ~ 0.0083 deg Please carefully check the rest of the data in this table.

R9: We greatly appreciate Reviewer #2 for pointing this out. This is a typo. We correct the error and carefully recheck the entire table. In our revised version, we express the grid size using native resolutions of the products for consistency with product descriptions.

	Products	*Grid size*	*Time interval*	*References*
Meteorological Forcing data	*Princeton forcing data version 2*	*0.5°*	*3 hours*	*Sheffield et al. (2006)*
Precipitation	*Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)*	*0.05°*	*1 day*	*Funk et al. (2015)*
Soil type	*Harmonized World Soil Database version 1.2*	*30 arc-second*	*n/a*	*Nachtergaele et al. (2009)*
Vegetation type	*MODIS Land Cover Maps*	*500 m*	*n/a*	*Broxton et al. (2014)*
LAI	*Global Land Surface Satellite (GLASS)*	*0.05°*	*~8 days*	*Xiao et al. (2013)*
GRACE	*NASA GSFC Mascons*	*Irregular*	*~1 month*	*Luthcke et al. (2013)*

C10: 9) Figure 11 & 12: Please specify the label for x axes.

R10: We modify our x-axis to a full year. The label of the x-axis is also given. We show the modified version of Fig. 11 below, and the same correction is also applied to Fig. 12.



Finally, we would like to thank Reviewer #2 for taking the valuable time to review our manuscript. We hope that our responses clarify Reviewer #2 questions. All suggestions will be implemented in our revised manuscript.