Response to Reviewer #3

The manuscript presents a high resolution model (0.05°) for Terrestrial Water Storage (TWS). The model implemented (CABLE SubgridSoil GroundWater), that was previously used to estimated TWS at 0.5°, is 'upgraded' to 0.05° resolution and extended with GRACE satellite observations via Ensamble Kalman Smoother. The method is demonstrated on Australia, a complex case study, with different climatological regions. Processing of the 32-year time span on continental scale is an impressive test case.

We thank Reviewer #3 for the feedback and constructive comments. We provided our responses (R, in blue) to each of the reviewer's comments (C, in black) below:

C1: The introduction highlights the importance of high resolution TWS modeling for risk management (l. 31), however, the temporal range of 1981-2012 does not reflect this application. The conclusion states that this would have been possible with different data (l. 461), at even higher resolution (l. 463). Why wasn't this done?

R1: We agree with Reviewer #3 that our simulation period demonstrated here might not be realistic for the (real-time) risk management application. However, please note here that our statement in this context describes the importance of improved spatial resolution:

At a regional or local scale, the spatial resolution of the TWS estimate is vital, as most applications (e.g., risk management for drought or flood) require accurate information at the county or sub-county level (Quiring, 2009).

Our simulation was based on public global-datasets that are only available from 1980 to 2012. As our conclusion suggests, the new datasets can be used if they are available. Unfortunately, at the time we conducted our analysis, there were no comprehensive global forcing datasets with higher spatial resolution than 0.5°/0.05° AND time span longer than the 1981-2012 period. In conclusion, we suggest the possibility of using a slightly coarser spatial resolution forcing data (MERRA2) to extend the dataset to the near present. To clarify Reviewer #3 concerns, we will include the above explanation in our conclusion as follows:

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 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.

In the submitted manuscript, we also suggest a possible improvement of spatial resolution using higher spatial resolution data. However, we find that 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.

C2: The open character of the model (l. 81) is an important characteristic that deserves more attention, due to its high potential. (N.B. other reviewers signaled that the public code may not include the GRACE data assimilation, this is unclear to me from the text.)

R2: We thank Reviewer #3 for the comment. Please note that our statement here only describes the public datasets (e.g., forcing data, parameters, remote sensing data), not codes:

Our approach utilizes only publicly available global datasets, so resulting TWS estimates can be reproduced over any target region…

The GRACE DA approach developed in our study is thoroughly described in Sect. 3.1, and it can be simply reproduced using any computer language. However, we understand the need for GRACE DA 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.

C3: l. 56-64 list various models of comparable spatial resolution.

R3: We thank Reviewer #3 for pointing this out. We reported the native unit of models/products and did not convert them to avoid rounding errors. To clarify Reviewer #3 concern, we will report the native resolution of the models/products and include km or degree unit (or both) when they are available.

C4: l. 76 "GRACE DA has shown positive impacts […]", combined with l. 90 "[t]he objectives of this paper are […] 2) to assess the GRACE DA impact on […] CABLE […]". What does CABLE provide that previous studies did not?

R4: We thank Reviewer #3 for the suggestion. For clarity, we will revise our introduction as follows:

Recent studies have shown success in assimilating GRACE data into a coarse-scale CABLE version to improve TWS and groundwater storage (GWS) estimates in the Goulburn River catchment and in the North China Plain (Tangdamrongsub et al., 2020; Yin et al., 2020). In this study, GRACE observations (Luthcke et al., 2013) are also assimilated into CABLE 0.05° (and CABLE 0.5°) to improve the accuracy of TWS components between 2003 and 2012. Assimilating the coarse GRACE observations into a much higher-resolution model is performed using the 3-dimension ensemble Kalman smoother (EnKS 3D; Tangdamrongsub et al., 2017). This approach will reveal whether assimilating GRACE data can benefit a newly developed fine-scale CABLE configuration. Our study will perform a thorough investigation on this issue to address GRACE DA's benefit on CABLE 0.05°.

C5: The GRACE data set spans only a small part of the 1981-2012 time span of the study. How is GRACE data integrated, outside the periods of data assimilation?

R5: We thank Reviewer #3 for addressing this question. GRACE data is only assimilated between 2003 and 2012 as clearly explained in lines 381 – 383:

GRACE observations are assimilated into the CABLE 0.5° and CABLE 0.05° models (called GRACE DA 0.5° and GRACE DA 0.05°, respectively) between January 2003 and December 2012 (due to the availability of meteorological forcing and GRACE data).

GRACE DA is not performed when data are not available, e.g., prior to 2003. The DA evaluation is only performed in 2003-2012 period. To clarify this further, we will also include the GRACE assimilation period in Sect. 2.3:

In this study, GRACE data are assimilated into CABLE between January 2003 and December 2012 (due to the availability GRACE data).

In addition, our data processing diagram (Fig. 3) also clearly explains that GRACE data are assimilated only when they are available. Please note that the flowchart is already modified based on the Reviewer #3 suggestion (please see R8).

C6: I would suggest to report both input and validation/evaluation data sets (§ 2.4) in similar fashion. For example, include the evaluation (satellite) data in a similar fashion in Table 2.

R6: We thank Reviewer #3 for the suggestion. The characteristics of evaluation (satellite) data will be included in Table 2 as Reviewer #3 suggested:

	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)
TWS	GRACE NASA GSFC Mascons	Irregular	~1 month	Luthcke et al. (2013)
Soil moisture	European Space Agency - Climate Change Initiative program (ESA-CCI)	0.25°	1 day	Dorigo et al. (2017)
Evapotranspiration	Global Land Evaporation Amsterdam Model (GLEAM)	0.25°	1 day	Martens et al. (2017)

C7: Upsampling of precipitation (l. 144) may not reflect natural precipitation patterns. Likewise, nearest neighbor interpolation (l. 142) may introduce strong gradients.

R7: We agree with Reviewer #3 that the resampled precipitation might not reflect its natural patterns. We are aware of this additional error and treated it statistically. 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:

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

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 is changed with the increased/decreased spatial resolution. For clarity, we will add the above explanation to Sect. 3.1.

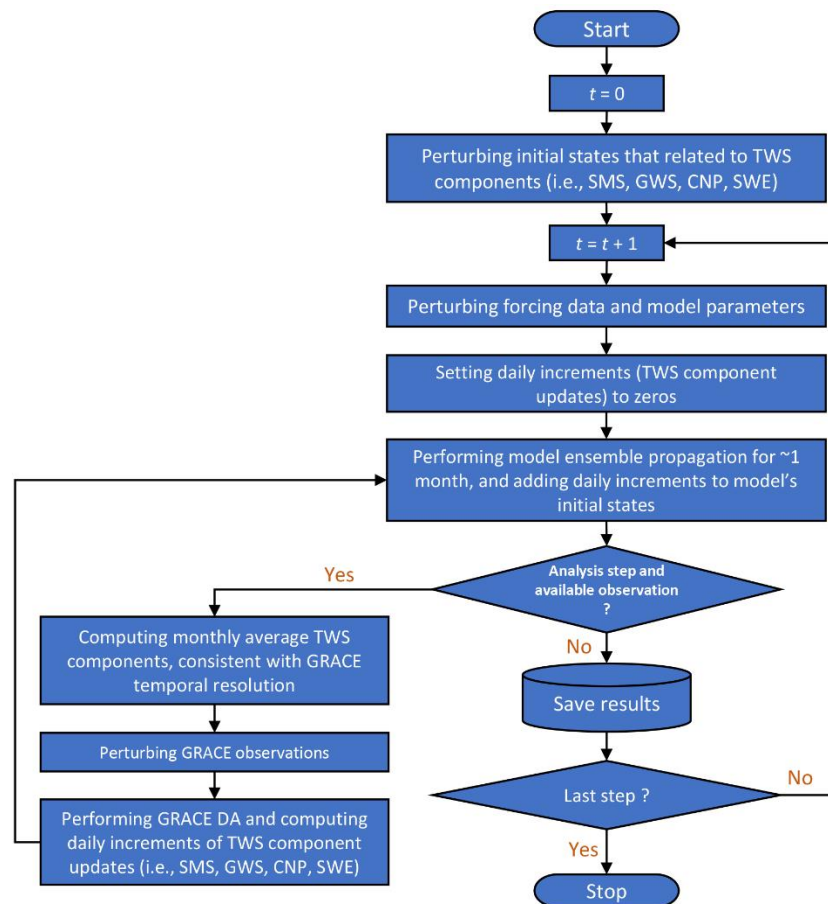
In addition, the impact of the resampled forcing data is also discussed in lines 376 – 378:

...The use of coarse resolution forcing data (e.g., precipitation) could also explain the small TWS amplitude observed in CABLE 0.5°. Coarse scale forcing data averages local precipitation signals over a larger area than does the finer resolution forcing data, resulting in a smaller amplitude.

C8: What is provided to CABLE, and what is optimized in the Ensemble Kalman Smoother. A 'data flowchart' would be a helpful extension to Figure 3.

R8: We thank Reviewer #3 for the suggestion. As described in lines 212 – 213, the updated state variables are six soil moisture layers, canopy storage, snow water equivalent, and groundwater storage:

The state vector consists of nine model states (n = 9): six soil moisture layers, canopy storage, snow water equivalent, and groundwater storage.

As Reviewer #3 suggested, we will update our processing diagram to include more processing details:

C9: l. 243 "[…], the daily increment (ΔA_d) of the update is computed by dividing ΔA by the total number of days in that month." How does this influence high frequency signals (e.g. precipitation spikes)?

R9: Because GRACE by nature provides lower frequency observations, it is unlikely to capture signals at surface layers (e.g., top soil) that is governed by high-frequency signal (e.g., from precipitation). The applied increment in EnKS is to distribute the GRACE monthly update throughout the month. The limitation of GRACE DA on high-frequency signals is already discussed in the manuscript lines 434 – 436:

... The small impact is attributed to limited GRACE sensitivity. 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 ...

C10: § 3.2.1 How is in-situ (point) data (§ 2.4.2) handled?

R10: The in situ data is not resampled. Our sensitivity analysis (not shown) reveals that performing the validation at in situ data location or at model grid cells leads to the same outcome. This is due to the fact that there are not many in situ data in the same model grid cell.

C11: § 3.2.3 What will this metric express?

R11: The regression is used to estimate the long-term trend and seasonal amplitude of TWS components. The results are discussed in Sect. 4.1.1, 4.1.2, and 4.2.

C12: Is there any indicator for the timing of the signal (e.g. systematically late detection)?

R12: The timing indicator is not used in our study. Our analysis is only based on correlation coefficient and RMS, consistent with all GRACE DA studies, e.g., Zaitchik et al. (2008), Tian et al. (2017), Kumar et al. (2016), Girotto et al. (2016).

C13: § 3.2.4 The term 'spatial resolution' is very confusing in the context of the study, see also l. 289. Also l. 450 "the 0.05° model also improves the spatial resolution by a factor of two to three over the 0.5° version," is counter-intuitive.

R13: We thank Reviewer #3 for raising this concern. Please note that a spatial resolution is defined as a minimum distance at which two signals of equal magnitude can be separated. As it is well understood, the spatial resolution is not necessarily equal to the model grid size. In other words, the 0.05° estimated variable may not have 0.05° resolution. We understand that the terms spatial resolution and model grid size can be confusing. As such, we carefully explain how the spatial resolution can be determined in Sect. 3.2.4 and provide its schematic in Fig. 4. We also clearly explain the difference between spatial resolution and grid size in lines 289 – 290:

It is noteworthy that the 0.5° or 0.05° represents the CABLE grid size, which may differ from the spatial resolution. The term "spatial resolution" used in this paper refers to the determined resolution computed from Sect. 3.2.4.

For clarity, we will add the definition of spatial resolution in the revised version: "*Spatial resolution is defined as the minimum distance at which two signals of equal magnitude can be separated*".

C14: There are no comments on the computational resources required by the model. Given the open character of the model, some hints would be welcome.

R14: The details on model source code, installation, and computational resource requirement can be found in the CABLE webpage (internet links are provided in Data availability section). CABLE is developed using Fortran and can be executed in a Unix environment. The input/output file format follows

NetCDF Climate and Forecast (CF) convention. We run the model in high-performance computing (HPC) environment. For clarity, we will include the details of CABLE software in our revised version:

CABLE is developed using Fortran and can be executed in a Unix environment. The input/output file format follows NetCDF Climate and Forecast (CF) convention…

C15: More overarching conclusions/summaries, for each comparison, would make the section more readable. What is the 'take home message' from each comparison.

R15: We thank Reviewer #3 for the comment. Please note that each analysis contains a summary sentence. For clarity, we will add an additional summary statement to the last sentence of our analysis (if it is not already there). Please also note the summary from each comparison can also be found in the conclusion section (in chronological order):

This study enhances the spatial resolution and timespan (> 30 years) of regional TWS estimates using the CABLE LSM, high-resolution land cover maps and forcing data, and GRACE DA application. By improving the model parameter and forcing data (without GRACE DA), the developed CABLE 0.05° model shows clear improvements in the accuracy of water balance component estimates (e.g., soil moisture, groundwater, evapotranspiration) compared with in situ and independent satellite data. The 0.05° model also improves the spatial resolution by a factor of two to three over the 0.5° version. The extended timespan provides insightful information for long-term assessment of regional water resources and climate variability. The enhanced model parameterization is found to play a significant role in the improved TWS estimates. Incorporating GRACE DA into the model leads to further improvements of TWS component estimates. The positive impact of GRACE DA is found in the deep storage component (e.g., GWS), while the impact on the surface components and flux estimates (i.e., SSM and ET) is trivial. Of the four case studies investigated here, the most accurate simulation uses CABLE 0.05° with GRACE DA…

C16: Various spatial units are mixed together, making difficult to compare between models and sources. The resolution of this model is reported in degrees, while the resolution of relevant models are mentioned in kilometers (l. 56-64). A single unit would be best, or report both.

R16: We understand that the mixed unit in the same paper might be confusing. Our paper reports the native unit of models/products and did not convert them to avoid rounding errors. To clarify Reviewer #3 concern, we will report the native resolution of the models/products and include km or degree unit (or both) when they are available.

C17: Ambiguous terminology, e.g.:

R17: Changes will be made. Please see below:

meaningful, l. 8, l. 33;

We will change meaningful resolution to high resolution.

sufficiently, l. 43, l. 52;

sufficiently will be removed from the sentence.

can, l. 44 (instead, provide a figure such as the longest available time series).

We thank Reviewer #3 for the suggestion. We will include two good references (e.g., Flechtner et al., 2014; Karthikeyan et al., 2017) where readers can obtain periods of available satellite observations.

Flechtner, F., Sneeuw, N. and Schuh, W.-D., Eds.: Observation of the System Earth from Space - CHAMP, GRACE, GOCE and future missions: GEOTECHNOLOGIEN Science Report No. 20, Springer-Verlag, Berlin Heidelberg., 2014.

Karthikeyan, L., Pan, M., Wanders, N., Kumar, D. N. and Wood, E. F.: Four decades of microwave satellite soil moisture observations: Part 2. Product validation and inter-satellite comparisons, Advances in Water Resources, 109, 236–252, https://doi.org/10.1016/j.advwatres.2017.09.010, 2017.

l. 30, "[…] spatial resolutions are relatively coarse […] models that primarily focus on global or continental scale";

The statement will be modified to:

.. spatial resolutions are coarse due to the limitation of sensors and models that focus on ...

l. 190 "[…] distributed unevenly across the continent."

The statement will be modified to:

... are distributed across the continent...

C18: Inline hyperlinks, except for the sources (l. 467 onward), make the text difficult to read, l. 60, l. 62, l. 114, l. 156, l. 161, l. 189.

R18: We thank Reviewer #3 for the comment. We will move the inline hyperlinks to Data availability sections.

C19: Table 1, l. 110, appears to be misplaced, l. 69. The reference in l. 106 could point back.

R19: We thank Reviewer #3 for the comment. Table 1 is placed after a paragraph when it is first mentioned, which is after Sect. 2.2 (at line 110).

C20: l. 130, "still"?

R20: "still" will be removed from the sentence. The sentence will read:

The soil map is also derived from the Harmonized World Soil Database but at 0.05° grid spacing…

C21: l. 132, Figure 1 does not show the high resolution of the data. Maybe include a detail?

R21: We thank Reviewer #3 for the comment. Please note that only rederived parameters are shown in Fig. 1. These parameters are used in the CABLE 0.05° configuration. This is stated in lines 131 – 132:

All rederived parameters are shown in Fig. 1.

C22: l. 149, 320 years → 32 years

R22: Please note that the model is spinning up for 320 years to obtain the initial states. This is stated in lines 149 – 150:

initial states are obtained using a 320-year spinup, i.e., performing ten repeated runs between 1981 and 2012.

C23: l. 185, "Murray-Darling Basin", see Figure 1.

R23: We thank Reviewer #3 for the suggestion. The statement will be modified to

...Murray-Darling Basin (see Fig. 1).

C24: l. 250, § 2.3 should be § 2.4

R24: We thank Reviewer #3 for the correction. Change will be made.