Editor's comment to the Author:

We have now received the reports of the two reviewers with comments on the revised manuscript. Though one of the reviewers is very positive, the other still has some concerns. After my own assessment of the manuscript and the reviewers' comments, I agree with Reviewer #1 and believe that the manuscript needs further revisions to clarify some methodological aspects, and their implications for the analysis and results.

It is important to evaluate/discuss the possible implications of leakage errors (section 2.2). Please also include further analysis for the selection of the 2004-2009 period and its representation of normal water conditions over the domain. Addressing these comments will further improve the value of the manuscript contributions.

Thank you for the assessment of our manuscript and provision of the comments. The manuscript cannot be improved so much without the help from the reviewers and yourself.

We have revised the manuscript to describe how 'normal' the period 2004-2009 compared to the entire period using the annual precipitation data in Line 147-153: mean annual P was 1444.0±138.0 mm over the period of 2004-2009, comparable to 1461.3±150.7 mm over the entire study period, which means that the period of 2004-2009 is representative of the normal condition over the study period.

Regarding the leakage errors along the basin boundary, we have added more information in the section 2.2 about GRACE data in Line 123-133.

We hope that with the newly added clarification, the manuscript can meet the standard of the journal in terms of the rigor in describing the data, methods and results. Sincere thanks to you all!

Reviewer #1

RC1: The authors argue that they do not intend to estimate leakage errors following Wiese et al., 2016, because they primarily used TWSA to infer water availability changes instead of the absolute values.

This logic is flawed. Method from Wiese et al 2016 has been widely used to quantify leakage error related to TWS variability and trend. To my knowledge most studies adopt their method to address leakage error associated with trends. For this study, the region to the north of the study domain is known to have experienced sustained increase in TWS linked to reservoir filling, which makes it necessary to evaluate if any of those signals have leaked into the analyzed domain.

AC1: First of all, we want to confirm that we agree that the measurement and leakage errors are important when using GRACE data for quantitative analysis. Many studies have discussed the methods for error estimation such as Long et al., (2017) and Wiese et al., (2016), although in many hydrologic applications the error estimations are not described such as in Muskett et al., (2009) and Yang et al., (2014).

We read carefully again the Wiese et al., 2016 paper and double checked the product information of GRACE data (Release 6). Wiese et al., 2016 paper states clearly that their procedures to reduce leakage errors across the land/ocean boundary are "unique to the JPL RL05M mascon solution and are not directly applicable to other GRACE mascon solutions. The reason for this is that JPL RL05M is currently the only available mascon solution that parameterizes the gravity field in terms of equal-area 3° spherical cap mascons. Other available mascon solutions parameterize the gravity field in terms of a finite spherical harmonic expansion of 1° mascon elements". The CSR RL06 mascon solution uses a grid of 1-degree and the hexagonal tiles that span across the coastline are split into two tiles along the coastline to minimize the leakage between land and ocean signals (Product Highlights at http://www2.csr.utexas.edu/grace/RL06_mascons.html). The CSR RL06 data include all necessary corrections (http://www2.csr.utexas.edu/grace/RL06_mascons.html). The JPL RL06 uses mascon grids of 3-degree, and scaling factors derived from CLM land surface model at 0.5-degree resolution are provided which are multiplied to calculate the TWSA. We then checked the JPL RL06 product information in the netcdf file, and found that apparently this RL06 product has already incorporated the coastline leakage error provided by Wiese et al. ,2016 method (see below).

```
months_missing              = '2002-06;2002-07;2003-06;2011-01;2011-06;2012-05;2012-10;2013-03;2013-08;2013-09;2014-02;2014-07;2014-12;2015-06;2015-10;2015-11;2016-04;2016-09;2016-10;2017-02'
postprocess_1               = ' OCEAN_ATMOSPHERE_DEALIAS_MODEL (GAD), MONTHLY_AVE, ADDED BACK TO OCEAN PIXELS ONLY'
postprocess_2               = 'Water density used to convert to equivalent water height: 1000 kg/m^3'
postprocess_3               = 'Coastline Resolution Improvement (CRI) filter has been applied to separate land/ocean mass within mascons that span coastlines'
GIA_removed                 = 'ICE6G-D: Peltier, W. R., D. F. Argus, and R. Drummond (2018) Comment on the paper by Purcell et al. 2016 entitled An assessment of ICE-6G_C (VM5a) glacial isostatic adjustment model
geocenter_correction        = 'Swenson, Chambers, and Wahr (2008), J. Geophys. Res., 113, 8410.'
C_20_substitution           = 'Cheng, M., Ries, and Tapley (2011), J. Geophys. Res., 116, B01409.'
user_note_1                 = 'The accelerometer on the GRACE-B spacecraft was turned off after August 2016.  After this date, the accelerometer on GRACE-A was used to derive the non-gravitational accelerations a
journal_reference           = 'Watkins, M. M., D. N. Wiese, D.-N. Yuan, C. Boening, and F. W. Landerer (2015) Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap
CRI_filter_journal_reference = 'Wiese, D. N., F. W. Landerer, and M. M. Watkins (2016) Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution, Water Resour. Res., 52, doi:10.1002/2016WR0193
```

In this case, we have added a few more sentences to include the above information for a better description of how errors in the GRACE TWSA are processed. They are given in line 123-133.

Long, D., Pan, Y., Zhou, J., Chen, Y., Hou, X., Hong, Y., et al. (2017). Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models. *Remote Sensing of Environment*, *192*, 198–216. https://doi.org/10.1016/j.rse.2017.02.011

Muskett, R. R., & Romanovsky, V. E. (2009). Groundwater storage changes in arctic permafrost watersheds from GRACE and insitu measurements. *Environmental Research Letters*, *4*(4). https://doi.org/10.1088/1748-9326/4/4/045009

Yang, Y., Long, D., Guan, H., Scanlon, B. R., Simmons, C. T., Jiang, L., & Xu, X. (2014). GRACE satellite observed hydrological controls on interannual and seasonal variability in surface greenness over mainland Australia. *Journal of Geophysical Research: Biogeosciences*, *119*(12), 2245–2260. https://doi.org/10.1002/2014JG002670

RC2: The authors justify their use of 2004-2009 baseline using the following argument: (a) the 2004-2009 baseline is "suggested by JPL," (b) calculating anomalies using the entire period as the new baseline is "more of the anomaly of anomaly," (c) the entire period is not significantly longer than the 2004-2009 period, and (d) the newly calculated mean annual anomaly is close to zero.

I want to first stress that for the purpose of identifying anomalous dry or wet conditions, *the suitability of a certain choice of baseline should be evaluated based on how normal the water conditions are during that baseline period. In practice longer baseline is often chosen to average out anomalies across different time scales.*

The authors did not evaluate how normal the water conditions were during the 2004-2009 period but instead provided four problematic arguments. (a) The 2004-2009 baseline is indeed used by JPL since earlier releases of the GRACE data but claiming that the 2004-2009 baseline is suggested by JPL would be of a different nature. In fact, JPL has provided guidelines regarding how to adopt a different baseline on the same link provided by the authors in the rebuttal letter. (b) Removing the mean over the entire period do not provide results that correspond to an anomaly of anomaly; they are simply anomalies referenced to a new baseline period. (c) The duration of the entire period is more than twice as that of the 2004-2009 period. (d) This sounds peculiar. Shifting the baseline will increase the magnitude of some of the anomalies while reducing the rest.

Finally, if the authors insist using the shorter period as the baseline, they should clarify how well that shorter period represents the normal water condition in the analyze domain.

AC2: We would like to thank you for such detailed comments again. With all respect, we keep the 2004-2009 as the baseline for all variables for the reason below.

The anomaly is the value deviating from a reference. If we had the true absolute water storage value for each month, then we can use any period as baseline reference to calculate the anomaly; but GRACE products are already storage anomaly (without absolute storage values). Therefore, subtracting the GRACE monthly anomaly data by the all-time mean, which is what we meant 'the anomaly of anomaly', will lead to near-zero results.

Anyway, the calculations with GRACE data show that the monthly and annual data using 2004-2009 baseline and entire period baseline have slight difference with the former having generally larger values than the latter (Fig. 1 a-b). So, from this view, it has no problem with the entire period as baseline.

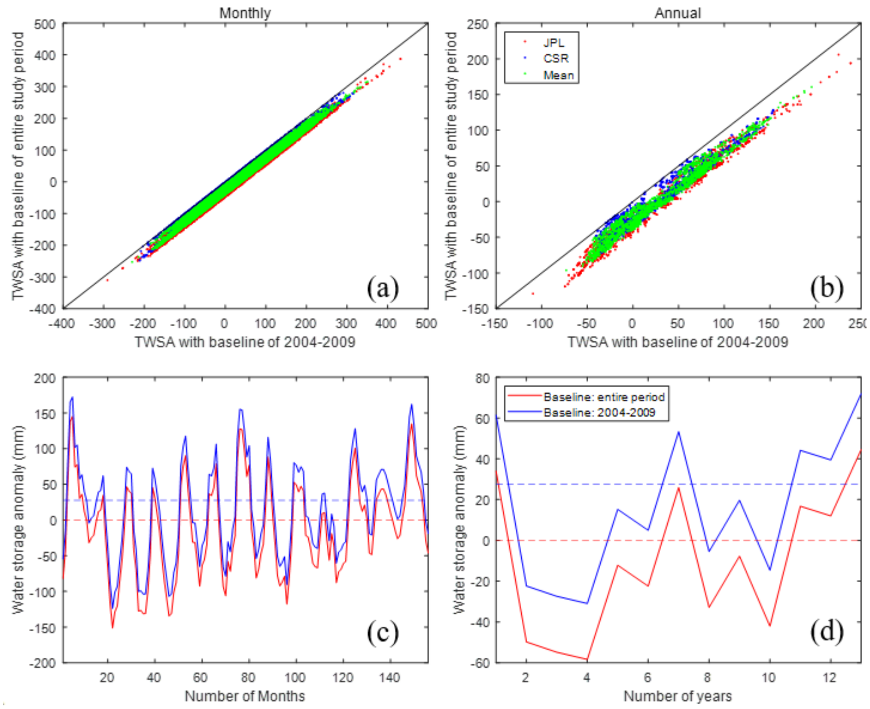

Fig. 1 (a-b) Comparison of TWSA calculated using two baseline periods at the monthly and annual scale. Data points include all GRACE data in the study area over 156 months and 13 years respectively. (c-d) Mean water storage anomaly over the entire study area with two baseline periods. Dashed lines are the all-time average over the area.

However, when we looked at the all-time mean TWSA over the entire study area, we found the mean value with baseline of entire period is nearly zero (Fig. 1 c-d). This is also true for the average value for each grid with baseline of the entire period (Fig. 2), that is, the all-year mean value for each grid is nearly zero, although the annual trends basically are the same with these calculated with baseline of 2004-2009.

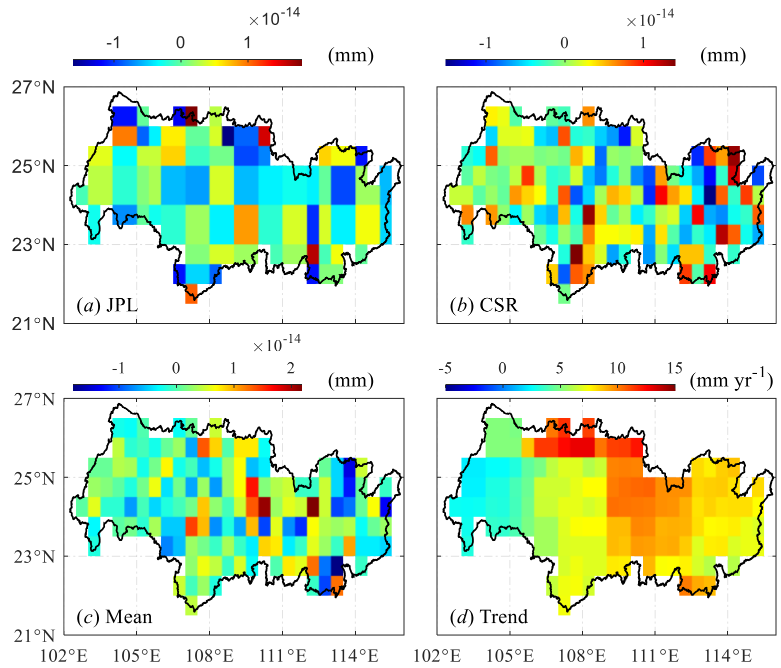

Fig. 2 Spatial distribution of men annual TWSA calculated with baseline of the entire study, and the trends.

Putting that aside, we calculate the annual mean precipitation of the basin and found it was 1461.3±150.7 mm over the entire period, comparable to 1444.0±138.0 mm over 2004-2009. The means that the period of 2004-2009 can be representative of the normal condition. We have added this explanation in the revision (line 147-153).

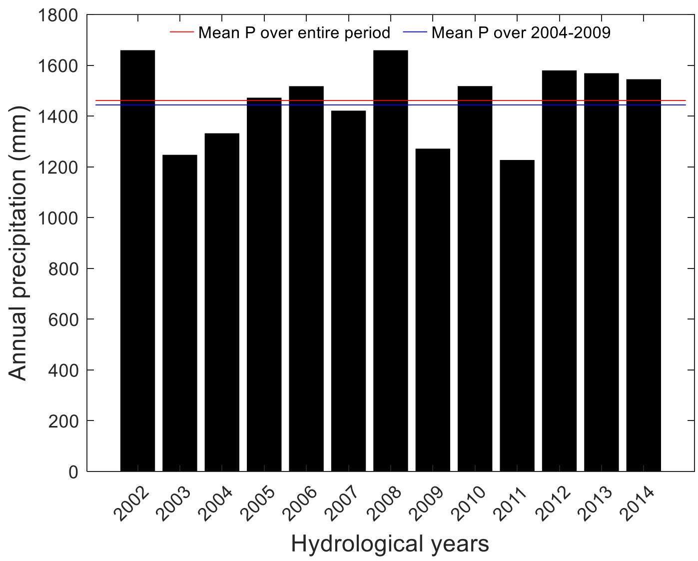

Fig. 3 Variations of mean annual precipitation over the basin. Red and blue lines in the figure mark the average value over the entire period and over 2004-2009, respectively.