To Abhishek Abhi, 24 May 2022

I read through this interesting manuscript by Xie et al. with great interest. Here, the authors have attempted to resolve the coarse temporal resolution of GRACE(-FO) data by downscaling to the daily time series by employing machine learning methods. Further, a new flood index, namely, the normalized daily flood potential index (NDPFI), is proposed to better characterize the extreme flood events in the Yangtze river basin (YRB).

I have some suggestions that authors may find beneficial while revising their manuscript.

Response: Thanks for your valuable suggestions. We have carefully checked and revised our manuscript in accordance to your comments. The detailed responses to your comments are provided as follows.

Major suggestions

(1) Selection of flood events: I found the authors have selected only extreme flood event(s) to analyze and demonstrate the capability of NDFPI (Lines 395, 413). I think that this may not be the best case. Selecting other flood events that are not well captured in other indices may be a better choice to discern the outperformance of and additional insights by the proposed index over conventional indices, e.g., SPI, SPEI.

Response: Thanks for your valuable suggestions. The reason why we selected only extreme flood event(s) to analyse and demonstrate the capability of NDFPI is that the YRB is particularly prone to catastrophic floods due to its highly uneven rainfall pattern and high annual rainfall (more than 1100 mm) (Zhang et al., 2021; Xiong et al., 2021). In particular, it has been found that both the frequency and severity of extreme flood events generally showed an upward trend in the YRB in the recent decades owing to substantial changes in climate, infrastructure and land use (Huang et al., 2015; Liu et al., 2019; Yang et al., 2021; Zhang et al., 2008). Furthermore, previous studies have compared the difference between the similar indices using monthly GRACE/GRACE-FO satellite data and other traditional evaluation indices, such as SPI and SPEI. For example, Yan et al. (2021) and Yin et al. (2021) clearly indicated that the similar indices using monthly GRACE/GRACE-FO data can better reflect the evolution of flood events than traditional evaluation indices, such as SPI and SPEI, because the GRACE/GRACE-FO observations can measure the vertically integrated water storage over regions. Therefore, we mainly focus on downscaling the TWSA estimates derived from GRACE/GRACE-FO satellite data into daily values and demonstrating its application to monitor the extreme flood events at sub-monthly time scales for the YRB in this study. Nevertheless, we also agree with you that more efforts can be made to analyse the potential of NDFPI to monitor other flood events that are not well captured in other indices when more observations about PET are available in future study.

References

Xiong, J., Yin, J., Guo, S., Gu, L., Xiong, F., Li, N., 2021. Integrated flood potential index for flood monitoring in the GRACE era. J. Hydrol. 603, 127115.

Yan, X., Zhang, B., Yao, Y., Yang, Y., Li, J., Ran, Q., 2021. GRACE and land surface models reveal severe drought in eastern China in 2019. J. Hydrol. 601, 126640.

Yin, G., Park, J., 2021. The use of triple collocation approach to merge satellite- and model-based terrestrial water storage for flood potential analysis. J. Hydrol. 603, 127197.

Zhang, X., Zhang, G., Long, X., Zhang, Q., Liu, D., Wu, H., Li, S., 2021. Identifying the drivers of water yield ecosystem service: A case study in the Yangtze River Basin, China. J. Hydrol. 132, 108304.

(2) Terminology of wet and dry seasons: If I am correct, the wet season in the basin spans from June to September, and this is also the same period when the flood of 2020 is observed (e.g., Figure 10). The newly proposed index is also shown better to detect the extreme events during the wet season (e.g., as stated in lines 408, 427). However, in a few places (e.g., lines 433, 551, 817), it is mentioned that the flood events in 2020 as the ‘summer 2020’. Is there something that I missed or confused between summer or wet season?

Response: As described in Line 92, the wet season in the YRB spans from April to October, which is consistent with previous studies (Huang et al., 2015; Yang et al., 2010). To avoid ambiguity, we have changed ‘summer 2020’ with ‘Year 2020’ based on your suggestions.

References

Huang, Y., Salama, M.S., Krol, M.S., Su, Z., Hoekstra, A.Y., Zeng, Y., Zhou, Y., 2015. Estimation of human-induced changes in terrestrial water storage through integration of GRACE satellite detection and hydrological modeling: A case study of the Yangtze River basin. Water Resour. Res. 51 (10), 8494-8516.

Yang, S., Liu, Z., Dai, S., Gao, Z., Zhang, J., Wang, H., Luo, X., Wu, C., Zhang, Z., 2010. Temporal variations in water resources in the yangtze river (Changjiang) over the industrial period based on reconstruction of missing monthly discharges. Water Resour. Res. 46, W10516.

(3) Handling the intermittent data gaps due to battery management (Line 111-112): It is not clear how the intermittent data gaps occurring about every six months in the GRACE and GRACE-FO TWSA time series were filled. Most probably, they were filled by linear interpolation or by the average of the bounding one-two months values. In either case, the filled values are likely (a) underestimating the actual (positive/negative peak) TWSA if the data gap happens to be in the peak of the wet or dry season (there

are a lot of such times, please see footnote of Table 3 in Abhishek et al., 2022), or (b) overestimation or underestimation in case of the high short-term fluctuations in the TWSA. Furthermore, this overestimation or underestimation can be critical given the topical issue of daily monitoring dealt with herein and, subsequently, might lead to inappropriate inferences in the YRB, which is highly vulnerable to floods. In my opinion, these data gaps can either be filled by running the three machine learning models (already used in the manuscript) for monthly TWSA, or alternatively, this can be stated as a likely source of uncertainty.

Response: Thanks for your valuable suggestions. Due to the problem of “battery management”, TWSA estimates in some months are not available for the GRACE and GRACE-FO satellites. As you suggested in this comment, these data gaps can be filled by running the three machine learning models (already used in the manuscript) for monthly TWSA. In fact, our previous study (Xie et al., 2019) also indicated that monthly TWSA estimates could be well reconstructed by the MLP model that was also applied in this study. Therefore, we further compared the monthly TWSA estimates reconstructed by the three machine learning models and that acquired from GRACE and GRACE-FO satellites (shown in Figure R1). According to Figure R1, we can find that these data gaps have been well filled by the three machine learning models. Furthermore, as shown in Line 515 to line 519, we have also added some sentences to clearly describe the uncertainty induced by these data gaps based on your suggestions.

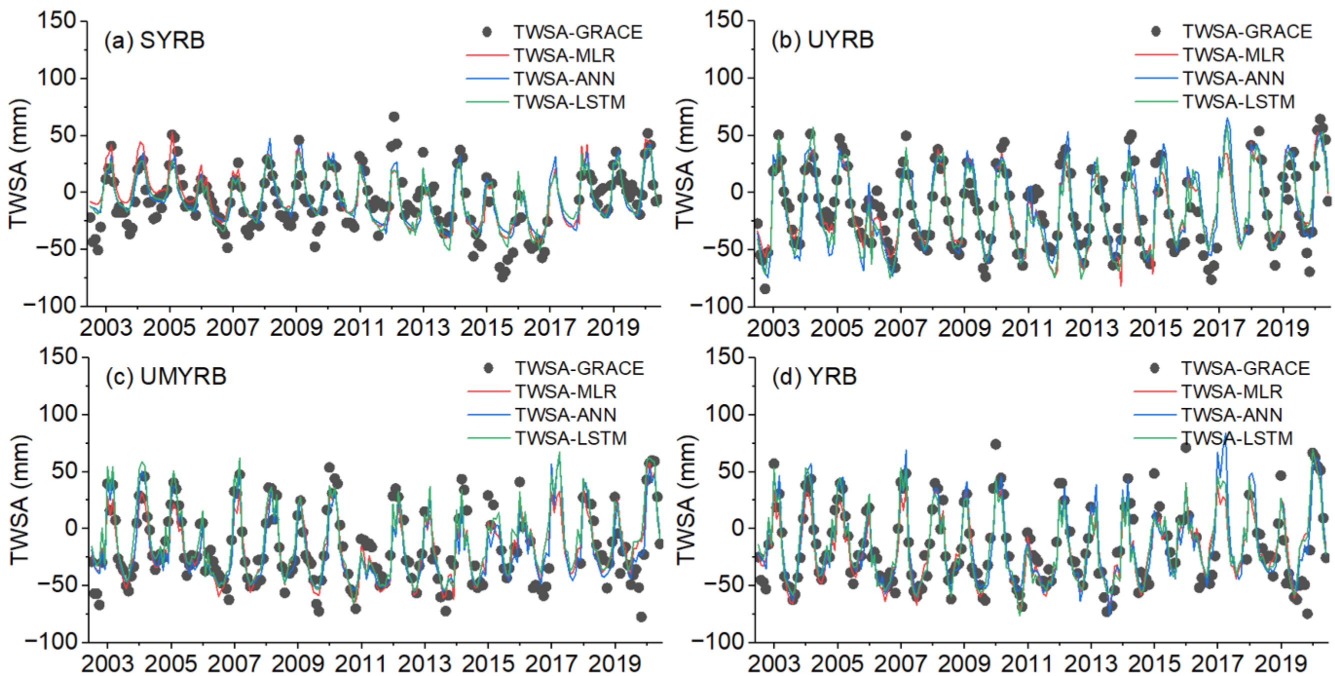

Figure R1. Comparison between monthly TWSA estimates reconstructed by the three machine learning models and that acquired from GRACE and GRACE-FO satellites for the YRB.

"Meanwhile, TWSA estimates in some months are not available for the GRACE and GRACE-FO satellites due to the problem of "battery management". Although all these missing months can be effectively filled by different machine-learning based models, it may overestimate or underestimate the actual TWS especially for some extreme values in the peak of the wet or dry season (Abhishek et al., 2022)."

References

Abhishek., Kinouchi, T., Abolafia-Rosenzweig, R., Ito, M., 2022. Water Budget Closure in the Upper Chao Phraya River Basin, Thailand Using Multisource Data. Remote Sens. 14(1), 173.

Xie, J., Xu, Y.P., Gao, C., Xuan, W., Bai, Z., 2019. Total basin discharge from GRACE and Water balance method for the Yarlung Tsangpo River basin, Southwestern China. J. Geophys. Res. Atmos. 124, 7617-7632.

Minor suggestions

Line 32: 'severe extreme'. Choosing one of these two words may be better.

Response: As described in Line 58, we have chosen one of these two words based on your suggestions.

"The YRB has been regarded as one of the most sensitive and vulnerable regions that suffered from **severe** floods due to its highly uneven rainfall pattern (Zhang et al., 2021)."

Line 52: 'evaluation' may not be necessary.

Response: As described in Line 43, we have deleted this word based on your suggestions.

"Previous studies have clearly indicated that the proposed indices using GRACE/GRACE-FO data can better reflect the evolution of flood events than traditional indices, such as standardized precipitation index (SPI) and standardized precipitation evapotranspiration index (SPEI), because the GRACE/GRACE-FO observations can measure the vertically integrated water storage over regions (Yan et al., 2021; Yin et al., 2021)."

Line 57: either 'a limited attention has been paid' or 'very few studies have paid attention to'.

Response: As described in Line 49, we have rewritten this sentence based on your suggestions.

"To date, **very few studies have paid attention to** monitor flood events at sub-monthly time scales using GRACE data."

Line 59: Since downscaling of the TWSA data is the aim (Line 60), the phrase 'using the temporally downscaled GRACE data' may be removed.

Response: As described in Line 50, we have removed the phrase 'using the temporally downscaled GRACE data' from this sentence based on your suggestions.

"Given the rapid occurrence and evolution of some extreme events within a short period, there is a great need to monitor the flood events at a finer temporal resolution (e.g. day), which has important implications for better understanding the mechanisms of extreme flood events development in the YRB."

Line 81: Reference of Fig. 1 may be provided here.

Response: We have added the reference of Fig. 1 in this sentence as described in Line 82.

"The terrain of the YRB generally decreases from west to east with altitudes ranging from -142 m to 7143 m above the sea level **(shown in Fig. 1).**"

Line 84: 'temperature' to 'temperate'

Response: As described in Line 87, we have changed 'temperature' to 'temperate' based on your suggestions.

"The YRB is located in typically subtropical and **temperate** climate zones, which is dominated by three types of monsoons, namely the Siberian northwest monsoon winds in winter, the Indian southeasterly monsoon winds and the East Asian monsoon in summer (Kong et al., 2020)."

Lines 106, 119, 131: 'basins' to 'subbasins'

Response: As described in Line 109, 125 and 137, we have changed 'basins' to 'subbasins' based on your suggestions.

Line 131: How the 'extreme flood events' were extracted is not clear? Providing a reference to Section 4.4 may be better.

Response: As described in Line 137 to Line 139, we have added a reference in this sentence, which can help us better understand how the extreme flood events are extracted in this study. In addition, we describe in detail the key steps about how the extreme flood events were extracted in Section 4.4.

“Meanwhile, extreme flood events in the YRB and its individual subbasins during the study period can be extracted from daily time series of streamflow observed from the above hydrological stations **(Tarasova et al., 2018)**. More details about how the extreme flood events are extracted will be described in the following Section 4.4.

“**4.4 Flood event selection**

A nonparametric algorithm suggested by Tarasova et al. (2018) is adopted to identify runoff events in this study, which has been widely applied in many different basins over the world because of its advantages in identifying flood events (Fischer et al., 2021; Giani et al., 2022; Lu et al., 2020; Winter et al., 2022). A brief procedure of this algorithm is described as follows: (1) picking out local minima within nonoverlapping five-day windows with respect to the entire streamflow time series; (2) examining the extracted series of minima with the goal of finding turning points, all of which are usually defined as the points that are at least 1.11 times smaller than their neighboring minima; (3) reconstructing the base flow hydrograph according to the linear interpolation between the turning points, which are previously obtained in Step (2); (4) screening the streamflow time series to identify runoff events after the separation of base flow. Traditionally, a typical runoff event can be characterized by three main components, namely peak, beginning, and end points. A peak refers to the maximum of streamflow for a specific period. The beginning point refers to the closest point in time when total runoff is equal to base flow before the peak. Similarly, the end point denotes the closest point in time when total runoff is equal to base flow after the peak.”

Line 142: ‘o’ to ‘to’

Response: According to the comments from Reviewer #1, this word has been removed from this manuscript. More details about this change also can refer to the response to the Comment #6 from Reviewer #1

Line 146, 530: using the single model output for soil moisture may include the implicit biases. How about using the ensemble mean from multiple model outputs, maybe even within the GLDAS series, subject to the availability and consistency with the study period.

Response: Thanks for your valuable suggestions. As documented in many studies (Proulx et al., 2013; Wei et al., 2022; Xiong et al., 2021), four different land surface models, i.e., Mosaic, Noah, Community Land Model (CLM), and Variable Infiltration Capacity (VIC) Model are provided by Global Land Data Assimilation System (GLDAS) datasets (https://disc.gsfc.nasa.gov/gldas). They have been proven to have good accuracy and can be used for the comparison of GRACE data. Previous studies (Liu et al., 2022; Long et al., 2014; Wang et al., 2021; Zhang et al., 2022) indicate that soil moisture in the Noah land surface model has a higher correlation with the GRACE-derived TWSA estimates compared to other land surface models including the Yangtze River Basin. Furthermore, our previous study (Xie et al., 2019) also suggests that the Global Land Data Assimilation System version 2.1 (GLDAS 2.1) Noah land surface model shows a superiority than the other models. Given the above reasons, we used the single model output for soil moisture in this study. Nevertheless, we also agree with you that the ensemble mean from multiple model outputs can be applied in our next study to further consider the uncertainties or implicit biases induced by model outputs.

References

Liu, M., Pei, H., Shen, Y., 2022. Evaluating dynamics of GRACE groundwater and its drought potential in Taihang Mountain Region, China. J. Hydrol. *in press*, https://doi.org/10.1016/j.jhydrol.2022.128156.

Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., Li, B., Chen, L., 2014. Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. Remote Sens. Environ. 155, 145-160.

Proulx, R.A., Knudson, M.D., Kirilenko, A., VanLooy, J.A., Zhang, X., 2013. Significance of surface water in the terrestrial water budget: A casestudy in the Prairie Coteau using GRACE, GLDAS, Landsat,and groundwater well data.

Wang, L., Peng, Z., Ma, X., Zheng, Y., Chen, C., 2021. Multiscale gravity measurements to characterize 2020 flood events and their spatio-temporal evolution in Yangtze river of China. J. Hydrol. 603, 127176.

Wei, M., Zhou, H., Luo, Z., Dai, M., 2022. Tracking inter-annual terrestrial water storage variations over Lake Baikal basin from GRACE and GRACE Follow-On missions. J. Hydrol. Reg. Stud. 40, 101004.

Xiong, J., Yin, J., Guo, S., Slater, L., 2021. Continuity of terrestrial water storage variability and trends across mainland China monitored by the GRACE and GRACE-Follow on satellites. J. Hydrol. 599, 126308.

Xie, J., Xu, Y.P., Gao, C., Xuan, W., Bai, Z., 2019. Total basin discharge from GRACE and Water balance method for the Yarlung Tsangpo River basin, Southwestern China. J. Geophys. Res. Atmos. 124, 7617-7632.

Zhang, X., Li, J., Dong, Q., Wang, Z., Zhang, H., Liu, X., 2022. Bridging the gap between GRACE and GRACE-FO using a hydrological model. Sci. Total Environ. 822, 153659. Water Resour. Res. 49, 5756-5764.

Line 157: ‘machine’ learning-based

Response: As described in Line 158, we have changed this sentence as you suggested.

“In Step 2, the relationship between TWSA estimates and all hydro-climatic factors at monthly time scales for the YRB can be built by using three different **machine** learning-based models, namely MLP model, LSTM model and MLR model respectively.”

Line 168: ‘served’ to ‘used’ or ‘employed’ or some other appropriate verb.

Response: As described in Line 171, we have changed this sentence as you suggested.

“Specifically, three types of models, namely, the artificial neural network (ANN), the recurrent neural network (RNN) and the multiple linear regression (MLR) are **used** as the statistical downscaling methods.”

Line 411: ‘rivers’ may not be the best word. Please see another suitable word, if possible.

Response: As described in Line 409, we have rewritten this sentence as you suggested.

“According to the Yangtze River Conservancy Commission of Ministry of Water Resources, the YRB has suffered from catastrophic flooding in 2020.”

Reference

Abhishek, Kinouchi, T., Abolafia-Rosenzweig, R., Ito, M., 2022. Water Budget Closure in the Upper Chao Phraya River Basin, Thailand Using Multisource Data. Remote Sens. 14.

Response: The results shown in the study of Abhishek et al. (2022) are very useful, which has been cited in our revised manuscript, as described in Line 517.