Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/hess-2022-168-RC1, 2022

Comment on hess-2022-168

Anonymous Referee #1

Referee comment on "Monitoring the extreme flood events in the Yangtze River basin based on GRACE and GRACE-FO satellite data" by Jingkai Xie et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2022-168-RC1, 2022

General comments

This manuscript is about the temporal downscaling of satellite based terrestrial water storage anomalies (TWSA) using data-based methods with in-situ meteorological data and assimilated soil moisture data as inputs. In addition, a new index is proposed incorporating downscaled TWSA estimates and daily precipitation to monitor flood events. The concepts are applied to four (sub-)basins of the Yangtze River Basin (YRB) in China.

Overall, the paper is well structured and written, and presents an interesting approach to temporally downscale GRACE/ GRACE-FO data to obtain daily TWSA estimates and its subsequent use in a daily normalized flood index. The authors generally use informative figures to illustrate their results. However, several issues need attention such as the concept and method to remove trends from the TWSA time series (see specific comment 7), the assumption that the data-based relationships between TWSA and hydrometeorological variables are time scale invariant (see specific comment 16) and the advantages of the proposed method and possible operationalisation (see specific comment 21). These and other specific and technical comments can be found below.

Specific comments

- L23: Only damage to structures and agriculture? No other types of flood damage and/ or casualties? See also line 38, where only a monetary value is mentioned without non-monetary damage and casualties/ victims.
- L31: Why does this study use the YRB as case study? How representative is the YRB for other basins in the world? To what extent can the methods and/ or results be generalized to other basins? This needs more attention in the manuscript, both in the introduction and discussion sections. The introduction of the YRB might be moved to

the end of the introduction section (after objective) to emphasize the role of the YRB being a case study to show the concepts and methods of this paper.

- L86-88: Which ranges are meant here? Ranges in space or in time (i.e. interannual variability)? Both temperature and precipitation ranges seem a bit narrow for a large river basin like YRB.
- L104-106: Why did the authors use the average of three types of GRACE and GRACE-FO solutions? Do these three types show equal performance in estimating TWSA? If not, wouldn't it be better to use weights depending on individual product performance when averaging the three products?
- L119-120: Is the Thiessen method the most suitable method for spatial averaging given for instance elevation differences in (sub-)basins?
- L143: Shouldn't the 3-hourly SMS data be summed instead of multiplied with 8? The SMS values for each 3-hourly time step are not necessarily the same, so multiplication with 8 will result in errors.
- L175-177: It is not completely clear which trends are considered here. As the authors indicate, long-term trends due to human activities should be removed in order to analyse the relation between 'natural' TWSA and hydroclimatic variables. The question is whether one can remove the human induced trend without removing part of a possible hydroclimatic (natural) induced trend. Did the authors analyse trends in precipitation and evapotranspiration (or temperature) to see whether climatic trends also could have been partly responsible for trends in TWSA? Should SMSA also be detrended similarly to TWSA? And did all relevant human activities cause a trend in TWSA or also abrupt changes in TWSA? Finally, it is doubtful whether the trend is always linear and hence linear methods can be used to remove the trends.
- L191: The number of hidden neurons of the MLP model has been set to five using a trial-and-error method. How has this been done and are the results reproducible? And did the authors obtain the same optimal number of hidden neurons for each of the four basins?
- L210-229: Are these formulas new or have these formulas already been described in the literature? In the latter case, this part can be removed from the manuscript.
- L234-240: Why is the regression equation linear? Shouldn't any non-linear transformations be considered or is it reasonable to assume linear relations between the three inputs and TWSA?
- L243-253: Several issues regarding the flood event selection are not clear. Why were five-day windows selected (step 1)? I can imagine that different windows need to be used for basins of different sizes. What is the background of the factor 1.11 (step 2)? How were the runoff events identified, i.e. how reproducible are the results (step 4)?
- L276: It would be good to also describe the spatial characteristics (e.g. different basins, averaging) of the analyses.
- L308-309: Why is TWSA consistent with temperature? What is the physical mechanism causing a high (positive) correlation between these variables?
- L362-363: Are the relationships between TWSA and hydroclimatic factors indeed complicated? Figure 4 suggests that these relations are rather straightforward as also confirmed by the authors when using linear correlations for the relations.
- L368-369: Is the limited availability of training data the main reason for the moderate performance of the LSTM model? In this study, it seems that sufficient data are available for training and validation. Another factor which might (partly) explain the moderate performance of this model compared to the MLP model is the possibly limited role of the memory function in the LSTM model, since relations between inputs and the output are quite direct without much memory effects.
- L374-378: An important assumption in this study is the independency of the input-output relations as determined by the data-based models on time scales, i.e. the relations found from monthly data are also assumed to be valid for daily data. The authors use as argument that 'the same scaling properties have been commonly assumed for baseline and future periods in temporal downscaling'. However, the situation/ conditions mentioned in this argument (translation from one period to

another) is different from the situation relevant in this study (translation from one (coarse) time scale to another (fine) time scale). Relations between (hydro)climatic variables and (other) hydrological variables are very different at different time scales. For instance, rainfall-runoff relations at hourly or daily time steps usually are highly non-linear, where relations at monthly or annual time scales are more or almost linear. Hence, it is doubtful whether the input-output relations at monthly time scales established by the three data driven models (in particular the MLP model) can be used for daily time steps as well. The authors should try to use some independent data sources (e.g. groundwater level measurements) at sub-monthly (preferably daily) scales to test the downscaled relations. Without such a validation/ testing it will be hard to assess the reliability of the results.

- L379-381: The relations determined at a monthly scales are used with daily inputs as well. I would not call this downscaling as the same relations are used for different time steps.
- L385-387: The results in Figure 6 show that the daily TWSA estimates do not capture the minimum monthly TWSA observations in all basins for several years. That is opposite to what the authors mention in this sentence. It would be good to also compare daily and monthly TWSA estimates in addition to the comparison between daily TWSA estimates and monthly TWSA observations. This might partly explain this behaviour.
- L401-402: It is doubtful whether soil moisture is the dominant component of TWSA. Most water is stored as (saturated) groundwater or surface water and the soil moisture storage is limited compared to these two storage components. Also changes in groundwater and surface water storage can generally be much larger than changes in soil moisture storage.
- L444-446: Do the lag times vary with the size of the basins? I can imagine that larger basins (e.g. the entire YRB) will have larger lag times than smaller basins.
- L478-485: The advantages of the proposed method and possible operationalisation need some nuance. Besides the considerations mentioned under specific comment 16, it should be emphasized that operationalisation would mean application of the MLP model at a daily scale using daily temperature, precipitation and soil moisture storage as inputs. In fact the TWSA observations are not directly used for early flood detection, but only indirectly for establishing relations at a monthly scale which are used at a daily scale. As such, the early flood detection is still mainly based on in situ data and will be hardly applicable to poorly or ungauged areas. Hence, the authors are encouraged to investigate whether observed TWSA estimates at a monthly time scale can be downscaled to a daily time scale without using in situ data or at least without using these data in an operational context.
- L517-518: Only water consumption data or also for instance data on reservoir operation?

Technical corrections

- L8: ‘monitor flood’ instead of ‘monitor the floods’.
- L20: ‘satellite data’ instead of ‘satellites data’.
- L25: ‘floods’ instead of ‘flood’.
- L33: ‘cycle’ instead of ‘cycles’.
- L35: ‘in recent decades’ instead of ‘in the recent decades’.
- L48: ‘based on a copula function’ instead of ‘based on the copula function’.
- L57: ‘few attention’ instead of ‘few attentions’.
- L59: ‘using temporally downscaled GRACE data’ instead of ‘using the temporally

downscaled GRACE data'.

- L62: 'to monitor extreme flood events' instead of 'to monitor the extreme flood events'. Similar corrections should be made in the remainder of the manuscript. Preferably, a native English speaking person should check and correct the document.
- L80: What do the authors mean with 'three-step ladder distribution'? Why three steps and not e.g. two, four or five steps?
- L100: 'can be found in Fig 1' instead of 'can refer to Fig. 1'.
- L118: 150 stations for both precipitation and temperature?
- L142: 'to daily and monthly estimates' instead 'o daily and monthly estimates'.
- L153: 'which consists of four steps' instead of 'which is made of four steps'.
- L158: The role of scenarios is not clear yet.
- L159: What is meant with 'scaling properties'? This is not clear here.
- L161: 'Step 2' instead of 'Step 3'?
- L200: 'existing' instead of 'existed'.
- L240-242: Shouldn't this sentence be moved to section 4.6?
- L252-253: Reformulate this sentence.
- L270-271: 'the typical difference between the storage change and precipitation'; what do the authors mean with this?
- L280: Remove 'generally'.
- L290-295: The subscripts in the formulas are not clear. Subscripts are used for both the time step and observed values, simulated values do not have an additional subscript. For average values an unclear subscript is used as well without having a clear meaning or role. The authors are advised to use subscripts for observed and simulated values (i.e. 'o' and 's') and put the time step between brackets.
- L302-303: This sentence includes quite some repetition.
- L306: Is the maximum TWSA value 130.9 mm (this line) or 130.1 mm (line 300)?
- L312: 'resulting' instead of 'resulted'.
- L349: 'difference' instead of 'decrease' (two times).
- L361: 'of' instead of 'higher than' and 'of' instead of 'lower than'?
- L368: 'ANN models' instead of 'ANNs models'.
- L371: 'satellite data' instead of 'satellites data'.
- L454: 'Discussion' instead of 'Discussions'.
- L460: Is hysteresis the right term here?
- L461: What do the authors mean with the term 'predisposition'?
- L504: What is 'deficit' here?
- L511: 'activities' instead of 'activates'.
- L513: What kind of surface conditions are meant here?
- L533-534: It is not clear what is meant with 'spatially correlated features'.
- L755: What is 'Now'? 2022?
- L760: Is it meaningful to show the overall performance mixing training and validation periods?
- L768: The Three Gorges Reservoir is difficult to identify on the map.