Review of “Synergy between satellite observations of soil moisture and water storage anomalies for global runoff estimation”

We thank the Topical Editor (TE) and the two anonymous reviewers for their supportive review. In the following, the author replies (AC) to the TE and reviewers comments (red lines) are reported.
Italic text on the AC replies reports the changes made on the revised manuscript. Lines of the revised manuscript refer to the author's track-changes version of the manuscript.

TEC: Comments and replies

Dear Stefania Camici and co-authors,

based on the referee recommendations, I would like to see the following:

(1) A discussion of the data--model fit and considerations of how to improve this (re: Referee #2)

(2) A comparison and/or discussion of how STREAM compares to other similarly scoped and well-established models (I'll add PRMS and/or GSFLOW to the list, too) in order to identify why a modeler may want to use STREAM over these.

I am happy to discuss this further as you may find it helpful.

AC: We thank the TE for the helpfulness demonstrated with the authors. We modified the text to address the TE and reviewer comment’s as specified in the following replies. Specifically, in the revised manuscript:

1) we added references to underline that the performances of the STREAM v1.3 model over the west part of the Mississippi river basin and in particular over the Great Plains are similar to the ones obtained by other hydrological models. Lines 475-477 have been modified as follows:

In particular, over section 3 the STREAM v1.3 model overestimates the observed river discharge due the presence of large dams along the Missouri river, over the Great Plains region. This area is well known from other global hydrological models (e. g., ParFlow-CONUS and WRF-Hydro) to be characterized by low performances in terms of river discharge modelling (O'Neill et al., 2020, Tijerina et al., 2021).

O'Neill, M. M., Tijerina, D. T., Condon, L. E., and Maxwell, R. M., Assessment of the ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water balance components across the contiguous United States, Geosci. Model Dev., 14, 7223–7254, https://doi.org/10.5194/gmd-14-7223-2021, 2021.

Tijerina, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M. M., Sampson, K., ... and Maxwell, R., Continental Hydrologic Intercomparison Project, Phase 1: A Large-Scale Hydrologic Model Comparison Over the Continental United States, Water Resour. Res., 57(7), e2020WR028931, https://doi.org/10.1029/2020WR028931, 2021.

2) we better specified the strengths, the limitations, and the innovative aspects of the STREAM v1.3 model. For that Lines 531 to 589 have been modified as follows:

"Hereinafter, the strengths and the main limitations of the STREAM v1.3 model are discussed.

Among the strengths of the STREAM v1.3 model it is worth highlighting:

1. Simplicity. The STREAM v1.3 model structure: 1) limits the input data required (only precipitation, T_"air" , soil moisture and TWSA data are needed as input; LSM/GHMs require many additional inputs such as wind speed, shortwave and longwave radiation, pressure and relative humidity); 2) limits and simplifies the processes to be modelled for runoff/discharge simulation. Processes like evapotranspiration, infiltration or percolation, are not modelled therefore avoiding the need of using sophisticated and highly parameterized equations (e.g., Penman-Monteith for evapotranspiration, Allen et al.,1998, Richard equation for infiltration, Richard, 1931); 3) limits the number of parameters (only 8 parameters have to be calibrated) thus simplifying the calibration procedure and potentially reduce the model uncertainties related to the estimation of parameter values.

2. Versatility. The STREAM v1.3 model is a versatile model suitable for daily runoff and discharge estimation over sub-basins with different physiographic characteristics. The results obtained in this study clearly indicate the potential of this approach to be extended at the global scale. Moreover, the model can be easily adapted to ingest input data with spatial/temporal resolution different from the one tested in this study (0.25°/daily). For instance, satellite missions with higher space/time resolution, or near real time satellite products could be considered. As an example, the Next Generation Gravity Mission design studies all encompass double-pair scenarios, which would greatly improve upon the current spatial resolution of single-pair missions like GRACE and GRACE-FO (> 100'000 km²). The STREAM v1.3 model shows high flexibility also in the possibility to modify the subbasin delineation and to introduce additional observational river discharge data to be used for the model calibration.

3. Computationally inexpensive. Due to its simplicity and the limited number of parameters to be calibrated, the computational effort for the STREAM v1.3 model is very limited."

However, some limitations have to be acknowledged for the current version of the STREAM v1.3 model:

1. Presence of reservoir, diversion, dams or flood plain. As the STREAM v1.3 model does not explicitly consider the presence of discontinuity elements along the river network (e. g, reservoir, dam or floodplain), discharge estimates obtained for sections located downstream of such elements might be inaccurate (see, e.g., river sections 1 and 2 in Figure 5).

2. Need of in situ data for model calibration and robustness of model parameters. As discussed in the results section, parameter values of the STREAM v1.3 model are set through an automatic calibration procedure aimed at minimizing the differences between simulated and observed river discharge. The main drawback of this parameterization *technique is that the models parameterized with this technique may exhibit (1) poor predictability of state variables and fluxes at locations and periods not considered in the calibration, and (2) sharp discontinuities along sub-basin boundaries in state flux, and parameter fields (e.g., Merz and Blöschl, 2004). To overcome these issues, several regionalization procedures, as for instance summarized in Cislaghi et al. (2020), could be conveniently applied to transfer model parameters from hydrologically similar catchments to a catchment of interest. In particular, the regionalization of model parameters could allow to: i) estimate discharge and runoff time series over ungauged basins overcoming the need of discharge data recorded from in–situ networks; ii) estimate the model parameter values through a physically consistent approach, linking them to the characteristics of the basins; iii) solve the problem of discontinuities in the model parameters, avoiding to obtain patchy unrealistic runoff maps. As this aspect requires additional investigations and it is beyond the paper purpose, it will not be tackled here.*

By looking at technical reviews of large-scale hydrological models (e.g., Sood and Smakhtin, 2015, Kauffeldt et al., 2016), it can be noted there are many established models, similar in objective and limitations to STREAM v1.3 model, already existing with support and user base (e.g., among others, Community Land Model, CLM, Oleson et al., 2013; European Hydrological Predictions for the Environment, E-HYPE, Lindström et al., 2010; H08, Hanasaki et al., 2008, PCR-GLOBWB, van Beek and Bierkens, 2008; Water – a Global Assessment and Prognosis WaterGAP, Alcamo et al., 2003; ParFlow–CLM, Maxwell et al., 2015; WRF-Hydro, Gochis et al., 2018). Some of them, e.g., ParFlow-CLM or WRF-Hydro have been specifically configured across the continental United States and showed good capability to reproduce observed streamflow data over the Mississippi river basin with performances decreased throughout the Great Plains (O'Neill et al., 2020, Tijerina et al., 2021) which is consistent with the results we obtained with STREAM v1.3 model. However, with respect to classical hydrological and land surface models, STREAM v1.3 is based on a new concept for estimating runoff and river discharge which relies on: (a) the almost exclusive use of satellite observations, and, (b) a simplification of the processes being modelled.

This approach brings several advantages: 1) satellite data implicitly consider the human impact on the water cycle observing some processes, such as irrigation application or groundwater withdrawals, that are affected by large uncertainty in classical hydrological models, 2) the satellite technology grows quickly and hence it is expected that the spatial/temporal resolution and accuracy of satellite products will be improved in the near future (e.g., 1 km resolution from new satellite soil moisture products and the next generation gravity mission); the STREAM v1.3 model is able to fully exploit such improvements; 3) STREAM v1.3 model simulates only the most important processes affecting the generation of runoff, and considers only the most important variables as input (precipitation, surface soil moisture and groundwater storage). In other words, the model does not need to simulate processes, such as evapotranspiration and infiltration and therefore it is an independent modelling approach for simulating runoff and river discharge that can be also exploited for benchmarking and improving classical land surface and hydrological models.

References have been added to the revised manuscript:

Maxwell, R. M., Condon, L. E., and Kollet, S. J.: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model

ParFlow v3, Geosci. Model Dev., 8, 923–937, https://doi.org/10.5194/gmd-8- 923-2015, 2015.

Gochis, D. J., Barlage, M., Dugger, A., FitzGerald, K., Karsten, L., McAllister, M., et al. (2018). The WRF-Hydro modeling system technical description, (Version 5.0). NCAR Technical Note. Retrieved from https://ral.ucar.edu/sites/default/files/public/WRFHydroV5TechnicalDescription.pdf

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., & Thielen, J.: Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level, Environ. Model. Softw., 75, 68-76, https://doi.org/10.1016/j.envsoft.2015.09.009, 2016.

Sood, A., and Smakhtin, V.: Global hydrological models: a review, Hydrol. Sci. J., 60(4), 549-565, https://doi.org/10.1080/02626667.2014.950580, 2015.

Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., … Yang, Z. -L.: Technical description of version 4.5 of the Community Land Model (CLM) (No. NCAR/TN-503+STR). http://dx.doi.org/10.5065/D6RR1W7M, 2013.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., & Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41(3-4), 295-319, https://doi.org/10.2166/nh.2010.007, 2010.

Hanasaki, N., Kanae, S., Oki, T., Masuda, K., Motoya, K., Shirakawa, N., ... , and Tanaka, K. :An integrated model for the assessment of global water resources–Part 1: Model description and input meteorological forcing, Hydrol. Earth Syst. Sci., 12(4), 1007-1025, https://doi.org/10.5194/hess-12-1007-2008, 2008.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., & Siebert, S.: Development and testing of the WaterGAP 2 global model of water use and availability, Hydrol. Sci. J., 48(3), 317-337, https://doi.org/10.1623/hysj.48.3.317.45290, 2003.

Van Beek, L. P. H., and Bierkens, M. F. P.: The global hydrological model PCR-GLOBWB: conceptualization, parameterization and verification.Utrecht University, Utrecht, The Netherlands, 1, 25-26, 2009.

Referee Report #1: Comments and replies

Line 162: replace “norther” with “northern”

AC: Accordingly, the text has been changed.

Line 235: replace “taken in to account” with “taken into account”

AC: Accordingly, the text has been changed.

Referee Report #2: Comments and replies

In this study, Camici and coauthors present a simplified conceptual discharge model that uses precipitation, soil moisture, and temperature to model quick runoff and GRACE-derived storage changes to model slow runoff. Although the rational behind the work is solid, though not particularly novel (https://www.sciencedirect.com/science/article/pii/S0012821X08006766, https://link.springer.com/chapter/10.1007/978-3-030-02197-9_1), the results are only good in the basin it's calibrated over, with little potential for transfer. The coauthors attempt to validate their model's utility by expressing it's easy of use, computational efficiency, and limited input data requirements, but it is far from the only model to check these boxes. Without comparison to some more commonly used models, say VIC, SWAT, Sacramento, or HEC-HMS, it's hard to convince people that they should use the presented STREAM model. I strongly encourage the coauthors to compare their results with other simplified conceptual discharge models to validate their model's utility.

AC: We modified the text to address the reviewer comment. Lines 531 to 589 have been modified as follows:

"Hereinafter, the strengths and the main limitations of the STREAM v1.3 model are discussed.

Among the strengths of the STREAM v1.3 model it is worth highlighting:

1. Simplicity. The STREAM v1.3 model structure: 1) limits the input data required (only precipitation, T_"air" , soil moisture and TWSA data are needed as input; LSM/GHMs require many additional inputs such as wind speed, shortwave and longwave radiation, pressure and relative humidity); 2) limits and simplifies the processes to be modelled for runoff/discharge simulation. Processes like evapotranspiration, infiltration or percolation, are not modelled therefore avoiding the need of using sophisticated and highly parameterized equations (e.g., Penman-Monteith for evapotranspiration, Allen et al.,1998, Richard equation for infiltration, Richard, 1931); 3) limits the number of parameters (only 8 parameters have to be calibrated) thus simplifying the calibration procedure and potentially reduce the model uncertainties related to the estimation of parameter values.

2. Versatility. The STREAM v1.3 model is a versatile model suitable for daily runoff and discharge estimation over sub-basins with different physiographic characteristics. The results obtained in this study clearly indicate the potential of this approach to be extended at the global scale. Moreover, the model can be easily adapted to ingest input data with spatial/temporal resolution different from the one tested in this study (0.25°/daily). For instance, satellite missions with higher space/time resolution, or near real time satellite products could be considered. As an example, the Next Generation Gravity Mission design studies all encompass double-pair scenarios, which would greatly improve upon the current spatial resolution of single-pair missions like GRACE and GRACE-FO (> 100'000 km²). The STREAM v1.3 model shows high flexibility also in the possibility to modify the subbasin delineation and to introduce additional observational river discharge data to be used for the model calibration.

3. Computationally inexpensive. Due to its simplicity and the limited number of parameters to be calibrated, the computational effort for the STREAM v1.3 model is very limited."

However, some limitations have to be acknowledged for the current version of the STREAM v1.3 model:

1. Presence of reservoir, diversion, dams or flood plain. As the STREAM v1.3 model does not explicitly consider the presence of discontinuity elements along the river network (e. g, reservoir, dam or floodplain), discharge estimates obtained for sections located downstream of such elements might be inaccurate (see, e.g., river sections 1 and 2 in Figure 5).

2. Need of in situ data for model calibration and robustness of model parameters. As discussed in the results section, parameter values of the STREAM v1.3 model are set through an automatic calibration procedure aimed at minimizing the differences between simulated and observed river discharge. The main drawback of this parameterization technique is that the models parameterized with this technique may exhibit (1) poor predictability of state variables and fluxes at locations and periods not considered in the calibration, and (2) sharp discontinuities along sub-basin boundaries in state flux, and parameter fields (e.g., Merz and Blöschl, 2004). To overcome these issues, several regionalization procedures, as for instance summarized in Cislaghi et al. (2020), could be conveniently applied to transfer model parameters from hydrologically similar catchments to a catchment of interest. In particular, the regionalization of model parameters could allow to: i) estimate discharge and runoff time series over ungauged basins overcoming the need of discharge data recorded from in–situ networks; ii) estimate the model parameter values through a physically consistent approach, linking them to the characteristics of the basins; iii) solve the problem of discontinuities in the model parameters, avoiding to obtain patchy unrealistic runoff maps. As this aspect requires additional investigations and it is beyond the paper purpose, it will not be tackled here.

By looking at technical reviews of large-scale hydrological models (e.g., Sood and Smakhtin, 2015, Kauffeldt et al., 2016), it can be noted there are many established models, similar in objective and limitations to STREAM v1.3 model, already existing with support and user base (e.g., among others, Community Land Model, CLM, Oleson et al., 2013; European Hydrological Predictions for the Environment, E-HYPE, Lindström et al., 2010; H08, Hanasaki et al., 2008, PCR-GLOBWB, van Beek and Bierkens, 2008; Water – a Global Assessment and Prognosis WaterGAP, Alcamo et al., 2003; ParFlow–CLM, Maxwell et al., 2015; WRF-Hydro, Gochis et al., 2018). Some of them, e.g., ParFlow-CLM or WRF-Hydro have been specifically configured across the continental United States and showed good capability to reproduce observed streamflow data over the Mississippi river basin with performances decreased throughout the Great Plains (O'Neill et al., 2020, Tijerina et al., 2021) which is consistent with the results we obtained with STREAM v1.3 model. However, with respect to classical hydrological and land surface models, STREAM v1.3 is based on a new concept for estimating runoff and river discharge which relies on: (a) the almost exclusive use of satellite observations, and, (b) a simplification of the processes being modelled.

This approach brings several advantages: 1) satellite data implicitly consider the human impact on the water cycle observing some processes, such as irrigation application or groundwater withdrawals, that are affected by large uncertainty in classical hydrological models, 2) the satellite technology grows quickly and hence it is expected that the spatial/temporal resolution and accuracy of satellite products will be improved in the near future (e.g., 1 km resolution from new satellite soil moisture products and the next generation *gravity mission); the STREAM v1.3 model is able to fully exploit such improvements; 3) STREAM v1.3 model simulates only the most important processes affecting the generation of runoff, and considers only the most important variables as input (precipitation, surface soil moisture and groundwater storage). In other words, the model does not need to simulate processes, such as evapotranspiration and infiltration and therefore it is an independent modelling approach for simulating runoff and river discharge that can be also exploited for benchmarking and improving classical land surface and hydrological models."*

R2: Line 278: The rain/snow differentiation model should be expanded on within the study. Rain/snow differentiation based on temperature and elevation is passably good, but at a large grid size like 25 x 25 km, the topographic complexity of higher elevations is lost. A differentiation scheme like that used in IMERG may be preferred, but isn't necessary. Still, this should be acknowledged, however briefly.

AC: We thank the reviewer to outline this aspect. In the manuscript we mentioned that (see Lines 285-290):

"In particular, according to Cislaghi et al. (2020), SWE is modelled by using as input T_{air} and a degree-day coefficient, C_m , to be estimated by calibration. We have to acknowledge that, even though this rain/snow differentiation method works quite efficiently at a large grid size like the one used in the study (25 x 25 km), the topographic complexity of higher elevations can be lost. A different differentiation scheme based e.g., on the wet bulb temperature like in IMERG (Wang et al., 2019; Arabzadeh and Behrangi, 2021), would be preferable but is out of the purpose study."

References have been added to the revised manuscript.

Wang, Y. H., Broxton, P., Fang, Y., Behrangi, A., Barlage, M., Zeng, X., and Niu, G. Y.: A wet-bulb temperature-based rain-snow partitioning scheme improves snowpack prediction over the drier western United States, Geophys. Res. Lett., 46(23), 13825-13835, https://doi.org/10.1029/2019GL085722, 2019.

Arabzadeh, A., and Behrangi, A.: Investigating Various Products of IMERG for Precipitation Retrieval Over Surfaces With and Without Snow and Ice Cover, Remote Sens., 13(14), 2726; https://doi.org/10.3390/rs13142726, 2021.

R2: Lines 345-348: Using a calibration tool would be preferable to manually adjusting to maximize Kling-Gupta. Perhaps one was used, but it is not specified. Also, does paragraph 5.1 relate to calibration, or is it paragraph 5.4?

AC: For the maximization of the Kling-Gupta efficiency Index we used a standard gradient-based automatic optimization method. This has been specified in the manuscript (see Lines 351-352)

"For model calibration, a standard gradient-based automatic optimization method (Bober 2013) was used."

The reference has been added tot the revised manuscript.

Bober, W. Introduction to Numerical and Analytical Methods with MATLAB for Engineers and Scientists; CRC Press, Inc.: Boca Raton, FL, USA, https://doi.org/10.1201/b16030, 2013.

R2: Section 5.1: "1. Input data collection" is unnecessary to include.

AC: Accordingly this part has been removed from the revised manuscript. In the new manuscript it can be read (see Lines 356-361):

1. Sub-basin delineation. STREAM v1.3 model is run in the semi-distributed version over the Mississippi River basin. The TopoToolbox (https://topotoolbox.wordpress.com/), a tool developed in Matlab by Schwanghart et al. (2010), and the SHuttle Elevation Derivatives at multiple Scales (HydroSHED, https://www.hydrosheds.org/) DEM of the basin at the 3′′ *resolution (nearly 90 m at the equator) have been used to derive flow directions, to extract the stream network and to delineate the drainage basins over the Mississippi River basin. In particular, by considering only rivers with order greater than 3 (according to the Horton-Strahler rules, Horton, 1945; Strahler, 1952), the Mississippi watershed has been divided into 53 sub-basins as illustrated in Figure 3.*

R2: Line 414-415: It is not clear to me what "to get to the right answers for the right reasons" means in this context and its tedious to hunt it down in the cited paper.

AC: This aspect has been expanded in the revised manuscript. The rationale behind the well-known concept "to get to the right answers for the right reasons" is that the hydrological models are today highly performing and able to reproduce a lot of hydrological variables. For that, the model performances should not only be evaluated against observed streamflow or associated signature measures, but complementary datasets representing internal hydrologic states and fluxes, such as soil moisture and evapotranspiration could be used to evaluate the capability of the model to simulate spatially distributed land surface fluxes controlled by local soil moisture availability and land surface hydrology.

Lines 414-415 have been modified as (see Lines 411-416 of the revised manuscript):

"3. External validation aimed to test the capability of the model "to get the right answers for the right reasons" (Kirchner 2006). The rationale behind this concept is that the hydrological models are today highly performing and able to reproduce a lot of hydrological variables. For that, the model performances should not only be evaluated against observed streamflow, but complementary datasets representing internal hydrologic states and fluxes, e.g., soil moisture, evapotranspiration, runoff etc) should be considered."

R2: Line 500-501: I would encourage you to include a precipitation map as a figure to illustrate your point.

AC: We thank the reviewer for the suggestion. A figure, showing the mean annual precipitation data obtained by TMPA 3B42 V7 and GSWP3 datasets over the Mississippi river basin has been added to the supplementary material. As it can be noted, both the datasets identify a strong difference between the western (dry) and the eastern (wet) area of the basin.

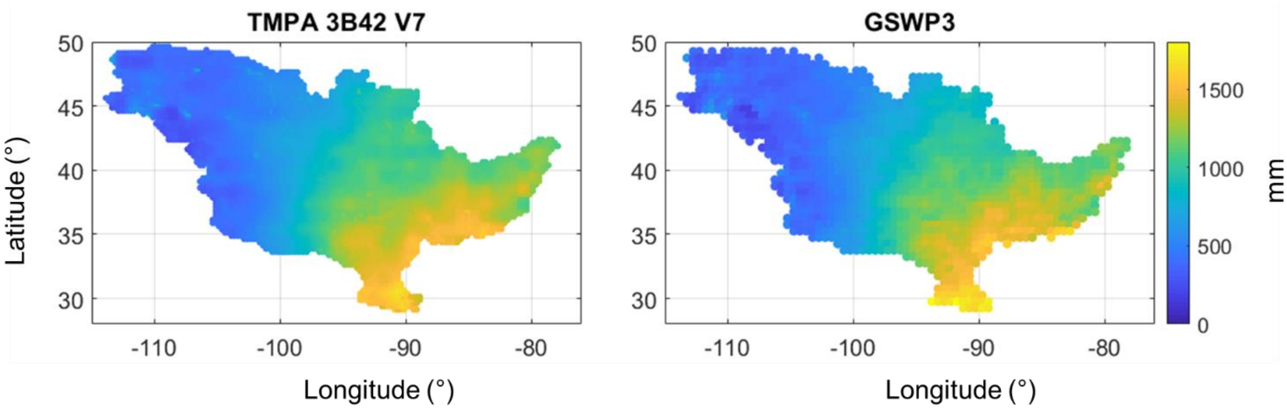

Figure. Mean annual precipitation data over the period 2003-2014 obtained by TMPA 3B42 V7 and GSWP3 datasets over the Mississippi river basin.

R2: Line 595: By the author's own admission (Lines 486-490), the model may not be suitable to reproduce discharge in basins not calibrated over. This should be changed to something less absolute. "Under some circumstances, the STREAM model can be used to estimate discharge in basins not calibrated over, especially those without upstream dams with comparable size and land cover." Or something similar.

AC: We thank the reviewer for this helpful suggestion. The sentence has been modified as:

"Conversely, the performances over river section 8, whose parameters have been set equal to the ones of river section 10, are quite high (KGE equal to 0.71, 0.80 and 0.77 for the entire, the calibration and the validation period, respectively; R equal to 0.83, 0.84 and 0.84 for the entire, calibration and validation periods, respectively). This outcome demonstrates that under some circumstances, the STREAM v1.3 model can be used to estimate river discharge in basins not calibrated over, especially those without upstream dams and with comparable size and land cover.

Although it is expected that the performances of STREAM v1.3 model, as any hydrological model calibrated against observed data, can decrease over the gauging sections not used for the calibration, the findings obtained above raises doubts about the robustness of model parameters and whether it is actually possible to transfer model parameters from one river section to another with different interbasin characteristics."