Topical Editor

Dear Dr. Camici and co-authors,

Thank you for your patience in my work to provide a response. I have been having a bit of difficulty in deciding what to do, considering the long time that your paper has been going back and forth through the editorial process alongside the fact that the new and independent referee feels, as do I, that the presentation of the work remains difficult to follow.

I was hoping to suggest more minor revisions, but based on the points of the new referee, added to my own concerns, I am suggesting "major revisions". Indeed, the new referee indicates similar questions and concerns to those that I have been having throughout the entire review process. If you respond to these fully and with appropriate changes to your paper, I would hope to be able to move forward more quickly with it. These include substantial improvements in the writing and consistency of your paper, without which I feel that others will find it difficult to understand (thereby significantly decreasing its impact), and appropriate contextualization of your work.

I am sorry that this news is not better: I realize that we have been at this for a while, and have been considering the best way to address these comments. Please feel free to be in contact if any additional clarification be needed,

R: We tried to do our best to make the manuscript and the revision more complete as possible. We hope we will be able to dispel all doubts of the reviewers and Topical Editor and to see the manuscript finally published on the GMD journal.

Specifically, we carried out the following additional analysis and modifications in the revision or in the revised manuscript:

1) A comparison between the modelled STREAM and PRMS river discharge data over the gauging stations selected in the Mississippi river basin. The analysis confirmed a good capability of the STREAM model to provide accurate river discharge estimates even better than those obtained through a comprehensive hydrological model such as the PRMS model (see reply to reviewer 1, below). However, as this analysis is beyond the scope of the manuscript, the authors prefer do not include it in the revised manuscript;
2) A sensitivity analysis of STREAM parameters. According to the suggestion of the Reviewer 2, a sensitivity analysis has been added to the revised manuscript in the paragraphs 5.4 and 6.3. This analysis allowed us to identify the most sensitive parameters of STREAM model and it poses the basis for the regionalization of model parameters;
3) According to the suggestions of reviewer 2, the STREAM model description and the related figure (Figure 2 in the revised manuscript) were accurately modified to better illustrate the model structure and the simulated processes;
4) The innovative aspect of the STREAM model has been in-deep highlighted. Through a comparison with conceptual hydrological models available in literature (see Lines 649-656) we

stressed the parsimonious structure of the model and the key role of satellite observations that allowed us to simplify the model structure (see Lines 625-636).

Reviewer #1

Dear Dr. Camici and co-authors,

Thank you for your revisions. However, you have chosen to only respond to select portions of the comments, which gives me no choice but to require further revisions. Furthermore, I have encountered some small errors and some unclear and/or misleading language in your revisions, which I have tried to correct, but I would ask that you take some care (or perhaps just place a bit of context in front of your enthusiasm at what you have built) in order to accurately describe the model's capabilities and limitations.

As a result, I have undertaken a read of your paper by myself. I am not fully familiar with all of the details involved in developing a hydrological model. For this reason, and because of the prior reviews (major revisions and reject, though I think that we agree that the latter's suggestion lies in contrast to their middling evaluation and lack of detail), I am going to send it out to one more referee. If the referee's decision is favorable, I will ask you to respond to both their reviews and mine.

Unanswered portions of my two requests:

(1) "considerations of how to improve this [the data--model fit]"

R: To answer this question in the previous revision of the manuscript we highlighted the possibility to regionalize the model parameters (see Lines 537-543; 684-698), avoiding to add any analysis related to the regionalization. The main reason behind this choice lies in the fact that the authors in the present manuscript would like to introduce the STREAM model and stress how satellite data can provide precious information in modeling river discharge leaving any additional analysis to future manuscripts. However, as this aspect seems crucial for the reviewer and the Topical Editor, we carried out a preliminary analysis to regionalize the model parameters.

For the application of the STREAM model in the Mississippi basin, we identified 53 sub-catchments each one with own characteristics in terms of land-cover, topography, climate, etc. Through the calibration procedure, for the same sub-catchments we identified 5 groups of parameters (each one for each calibration site). By using this parameter sets we tried a first attempt of parameters regionalization. In particular, based on the results of the sensitivity analysis (see paragraph 6.3 in the revised manuscript), we fixed some model parameters (the less sensitive, i.e., gamma, T, Cm, C and D) to constant values (equal to the mean values obtained for the 53 sub-catchments) while we looked for plausible relationships with the basin characteristics for the others (i.e., alpha, beta and m). Specifically, we linked alpha to the aridity index, beta and m to the bulk density.

The results of this first attempt of model parameter regionalization are shown in the following. Figure1R shows the performances, in terms of KGE, obtained through the calibrated parameter set (on the left upper map) and the variation on model performances obtained by considering the regionalized model parameters (on the right upper map). The bottom plot of Figure 1R shows the modelled river discharge obtained by considering the calibrated (STREAM cal.) and the regionalized (STREAM reg.) parameter sets. Figure 2R illustrates the performances of the model over the sections used for the model calibration.

As expected, both the figures highlight a worsening of model results when the regionalized set of parameters are used for running the model. However, the deterioration is lower than 30% and it can be retained acceptable.

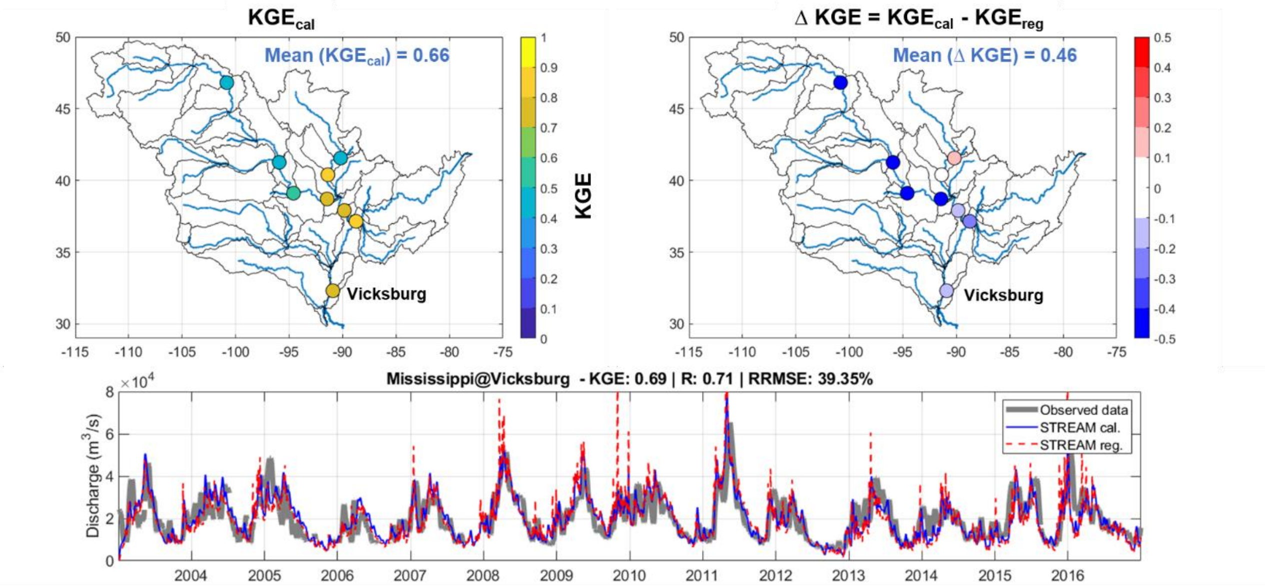

Figure 1R. Performances obtained through the STREAM calibrated/ regionalized (left/right map) parameter sets.

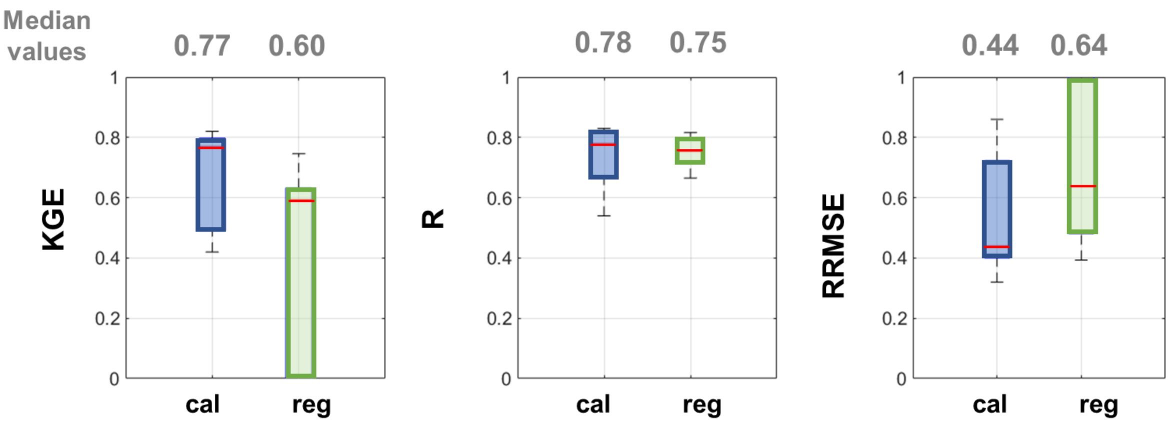

Figure 2R. Performances scores over the sections used for the model calibration (see Table 2 in the manuscript). Cal and reg stand for to calibrated and regionalized parameter set.

Although these results are very interesting, we remark that this is just a preliminary analysis for the regionalization of the STREAM model parameters and for the reasons expressed above, the authors would like do not include this part in the manuscript. We hope the reviewer could understand our point of view.

(2) Consider PRMS and/or GSFLOW; especially the former is scoped much more similarly to STREAM than some of your targets for comparison, which in turn affects the validity of some of your statements below.

R: The authors feel that this request is beyond the scope of the paper, that is to present an innovative and simple model able to estimate runoff and river discharge by exploiting, as much as possible, satellite data. The authors are aware that in literature there are a plethora of hydrological models (e.g., among others PRMS or GSFLOW) more complete in the process's representation with respect to STREAM model. Based on that, the main intent of the authors is not to introduce a new hydrological model but demonstrate the high informative content of the satellite observations in reproducing some processes (e.g., evapotranspiration or human impact) without the need to explicitly to model/parametrize them into an hydrological model. Being this is mind, is clear that any comparison with a classical hydrological model (e.g., PRMS or GSFLOW) is meaningless.

However, as this point seems crucial for the reviewer and the Topical Editor, the authors carried out a comparison of the STREAM results against the ones obtained through the PRMS model. Before to carried out the comparison, a remark is needed. As correctly stated by the reviewer, STREAM and PRMS model are quite similar from a technical point of view as both discretize the basin into subbasins and streams (HRU and segments for PRMS) but they are quite different for the processes they represent. PRMS model (https://pubs.usgs.gov/tm/06/b09/tm6b9.pdf) is designed to simulate the full hydrologic cycle as determined by the energy and water budgets of the plant canopy, snow-pack, and soil zone on the basis of distributed climate information (temperature, precipitation, and solar radiation). In this respect, STREAM model is far away from PRMS as the modelling component, and consequently the number of parameters, is reduced at minimum whereas satellite observations are exploited as much as possible to represent unmodelled processes.

The result of the model's comparison is shown in the following, but it will not be included in the manuscript because, as stated before, according the atuthirs it is beyond the purpose of the study.

The river discharge data simulated through the PRMS model have been downloaded from the "Application of the National Hydrologic Model Infrastructure with the Precipitation-Runoff Modeling System (NHM-PRMS), by HRU Calibrated Version" (https://www.sciencebase.gov/catalog/item/5a4ea3bee4b0d05ee8c6647b). Specifically, the variable "nsegment_seg_outflow" (i.e., the streamflow leaving a segment as described in Table 1 on https://www.sciencebase.gov/catalog/item/5a4ea3bee4b0d05ee8c6647b) contained in the homonym csv dataset has been extracted for the PRSM model. To distribute output values spatially, the nseg_ID from the national-extent Geospatial Fabric has been used (GIS Features of the Geospatial Fabric for National Hydrologic Modeling, https://www.sciencebase.gov/catalog/item/5410bd38e4b07ab1cd98c45f).

Then, for each station along the Mississippi river the modelled STREAM and PRMS river discharge time series have been compared. Figures 3R-5R illustrate the comparison between modelled and observed river discharge data whereas the performance scores (Kling-Gupta efficiency index, KGE, correlation coefficient, R, and relative root mean square error, RRMSE) are summarized in the Table 1R.

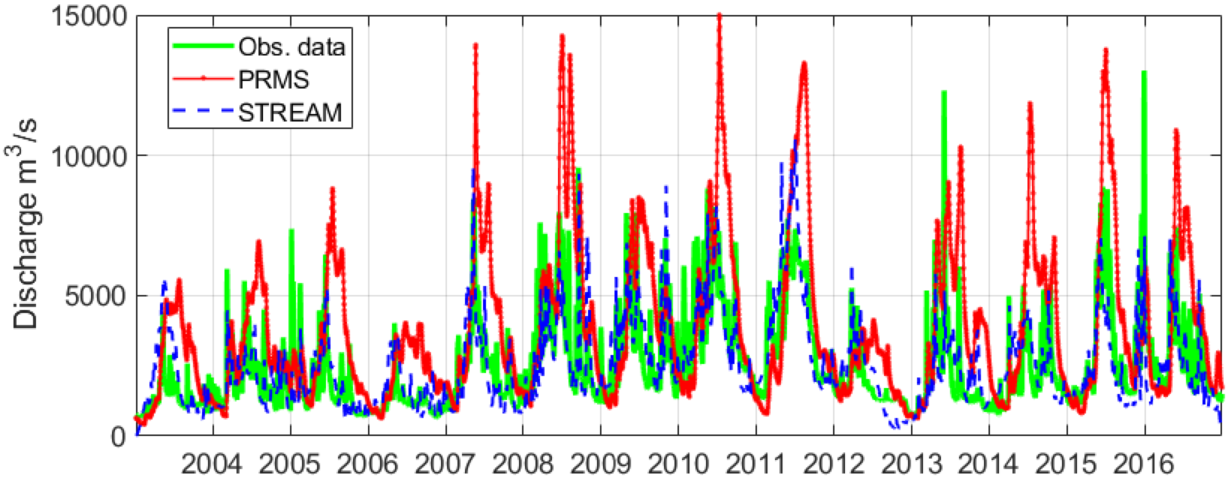

Figure 3R. Hermann station in the Missouri river (gauging station n. 4 in the manuscript). Comparison between observed and modelled PRMS and STREAM river discharge data.

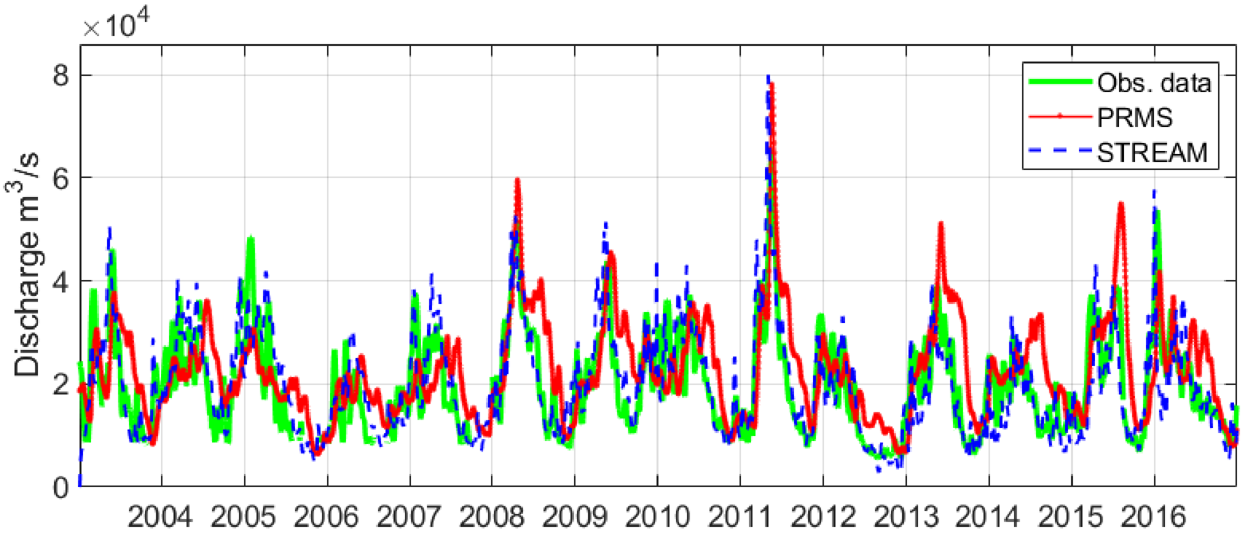

Figure 4R. Vicksburg station (gauging station n. 10 in the manuscript). Comparison between observed and modelled PRMS and STREAM river discharge data.

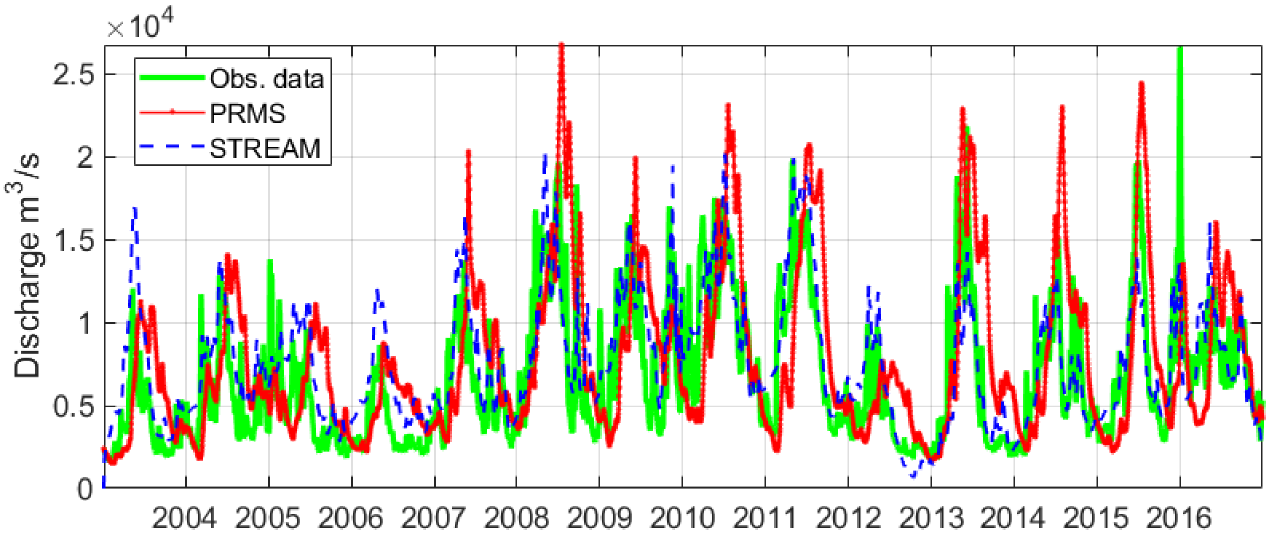

Figure 5R. Chester station (gauging station n. 8 in the manuscript). Comparison between observed and modelled PRMS and STREAM river discharge data. Please, note that the data of this gauging station has not be used for the STREAM model calibration.

Table 1R. Performance scores between observed and modelled river discharge data over the gauging stations selected in the Mississippi river.

Station ID	KGE		R		RRMSE (%)	
	STREAM	PRMS	STREAM	PRMS	STREAM	PRMS
1	0.39	-0.84	0.44	0.60	88.10	177.68
2	0.46	-0.42	0.56	0.67	77.89	137.89
3	0.53	-0.50	0.72	0.68	62.44	132.29
4	0.75	0.04	0.77	0.60	43.42	104.99
5	-1.23	-0.38	0.25	0.64	348.72	219.79
6	0.77	0.60	0.79	0.60	45.23	59.83
7	0.52	0.66	0.61	0.76	71.87	56.35
8	0.71	0.49	0.83	0.56	38.89	66.69
9	0.64	0.57	0.82	0.76	79.21	92.19
10	0.78	0.59	0.78	0.62	32.45	44.23
11	0.82	0.60	0.82	0.66	44.63	61.11
Mean over the basin	0.46	0.13	0.67	0.65	84.80	104.82

As it can be noted, both the models are quite accurate in reproducing the river discharge data observed in some gauging stations (e.g., station 6, 7, 9, 10 an 11) but, on overall the STREAM model overperforms the PRMS model. In particular, the STREAM model shows better performances also over the stations 1, 2 and 3 located downstream large dams and over the stations not used for calibration (e.g., station 8).

On overall, by comparing PRMS and STREAM model we can conclude that the satellite observations offer a high information content avoiding to model complex processes like the evapotranspiration.

With this analysis, the authors hope to have dispelled all doubts of the reviewers and Topical Editor.

General comments:

(A) You note multiple times that you bypass the need to compute evapotranspiration becuase you have soil-moisture data available. This seems to be a large advantage of STREAM. Is it? Is it innovative compared to what others have done? And what is the technical justification for bypassing ET? To me, it would seem that ET can reduce soil moisture, but can also reduce the total amount of water that makes it to the point at which it can infiltrate. How, therefore, do you account for the pathway of the water droplet and these possible losses, and might they be significant?

R: In standard hydrological models, precipitation and evapotranspiration data are used as input to simulate the soil water storage that is the key variable for partitioning precipitation in infiltration and runoff. Therefore, evapotranspiration observations are needed even though their measurement or simulation is particularly difficult (Parr et al., 2015). In STREAM model we exploit satellite observations of surface soil moisture and soil water storage from GRACE to avoid the need of using evapotranspiration data. Therefore, yes it is a large advantage of STREAM that will has been highlighted in the revised paper.

This aspect has been highlighted in the manuscript multiple times (see Lines 148-153):

"Unlike classical LSMs, STREAM exploits the knowledge of the system states (i.e., soil moisture and TWSA) to derive river discharge and runoff, and thus it 1) skips the modelling of the evapotranspiration fluxes which are known to be a non-negligible source of uncertainty (Long et al. 2014), 2) limits the uncertainty associated with the over-parameterization of soil and land parameters and 3) implicitly takes into account processes, mainly human-driven (e.g., irrigation, change in the land use), that might have a large impact on the hydrological cycle and hence on runoff."

see Lines 625-636:

"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 model is able to fully exploit such improvements; 3) STREAM model models 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 parametrize processes, such as evapotranspiration and percolation 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."

or see Lines 644-646:

"*Processes like evapotranspiration 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)*;

Parr, D., Wang, G., & Bjerklie, D. (2015). Integrating remote sensing data on evapotranspiration and leaf area index with hydrological modeling: Impacts on model performance and future predictions. Journal of Hydrometeorology, 16(5), 2086-2100.

(B) Is the "terrestrial water storage anomalies" common knowledge among hydrologists? I would suggest defining this (anomalies from what?) and writing the source and a bit more background information.

R: We modified the manuscript specifying this aspect (see Lines 130-134):

"GRACE and its successor mission GRACE-FO provide monthly snapshots of the Earth's gravity field. The temporal variation is therefore relative to the temporally mean gravity field and, hence, the time variations of water storage are fundamentally relative to the mean storage. This relative water storage variation is termed Total Water Storage Anomaly (TSWA)."

Additional specific comments, by line:

226-227: You write, "Although the mascon size is smaller than the inherent spatial resolution of GRACE, the model exhibits a relatively high spatial resolution." Here you need to separate the true resolution of the data from the gridded resolution of the product and make it clear whether it is just that the product has an inappropriately high resolution (which I do not think that you mean) or whether there are ways to go above the spatial resolution of GRACE (perhaps expressed in terms of spherical harmonic degree and order?)

R: We modified this sentence to better explain this aspect (see Lines 229-232):

"Although the mascon size is smaller than the inherent spatial resolution of GRACE of about 2.5°×2.5° or 64'000 km2 (Vishwakarma et al., 2018), the model exhibits a relatively high spatial resolution. This is attributed to a statistically optimal Wiener filtering, which uses signal and noise full covariance matrices."

287-288: You have not demonstrated the efficiency of your snow component, and you then compare efficiency to accuracy, which are two different things.

R: The reviewer is right. The sentence has been moved from this part and changed as follows (see Lines 680-683):

"Snow modelling. A potential limitation of the current version of the STREAM model is related to the rain/snow differentiation, based on the degree-day coefficient. A different scheme based e.g., on the wet bulb temperature like in IMERG (Wang et al., 2019; Arabzadeh and Behrangi, 2021), could be investigated in future developments."

289-291: I see the help of a wet-bulb temperature, but here it appears that you are combining information on grid resolution with information on input data type, which are not addressing the same fundamental topic/limitation

R: This part has been moved in the paragraph 7, "Discussion" and it has been modified as follows:

"Snow modelling. A potential limitation of the current version of the STREAM model is related to the rain/snow differentiation, based on the degree-day coefficient. A different scheme based e.g., on the wet bulb temperature like in IMERG (Wang et al., 2019; Arabzadeh and Behrangi, 2021), could be investigated in future developments."

548-557: This is one place where you could compare with other similar models available today. To my knowledge, STREAM is indeed simpler.

R: We thank the reviewer for the suggestion. The paragraph that in the previous version of the manuscript was:

"1. Simplicity. The STREAM v1.3 model structure: 1) limits the input data required (only precipitation, air temperature, 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."

as been modified as follows (see Lines 640-656):

"Simplicity. The STREAM model structure: 1) limits the input data required. Only precipitation, air temperature, soil moisture and TWSA data are needed as input whereas 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 and river discharge simulation. Processes like evapotranspiration 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); 3) limits the number of parameters (only 8 parameters have to be calibrated) thus simplifying the calibration procedure and potentially reduces the model uncertainties related to the estimation of parameter values.

In particular, the STREAM model is even simpler than the classical semi-distributed conceptual hydrological models available in literature. As an example, for the comparison we could refer to the Hydrologiska Byråns Vattenbalansavdelning model (HBV, Bergström 1995) or to the Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS, Feldman, 2000). HBV model counts 14 parameters to be calibrated and needs precipitation, air temperature and potential evapotranspiration as input data. Similar input data are required for HEC-HMS which counts 23 parameters. Both the models, uses conceptual equations to estimate the soil losses and to model the soil water storage."

560: "Clearly indicate" and other such effusive language should be removed; change to "indicate" and refer to a section and/or figure in the paper that demostrates your claim.

R: The sentence that in the previous version was:

"*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*.

has been modified as follows (see Lines 657-661):

"The STREAM model is a versatile model suitable for daily runoff and river discharge estimation over sub-basins characterized by different physiographic/climatic characteristics (see e.g., the outcomes obtained for the gages 9 and 11 located in the driest and wetter part of the Mississippi basin). This aspect is paramount as it gives an insight about the potential of the model to be extended at the global scale."

563-565: What resolution with the Next Generation Gravity Mission hydrological data product have? It isn't clear that this will be higher from what you have said without any references

This has been specified in the sentence (see Lines 603-609):

“. Further developments in this direction are expected with the ESA’s Next Generation Gravity Mission (NGGM), a candidate Mission of Opportunity for ESA–NASA cooperation in the frame of the Mass Change and Geosciences International Constellation (MAGIC) that will enable long-term monitoring of the temporal variations of Earth’s gravity field at relatively high temporal (down to 3 days) and increased spatial resolutions (up to 100 km). This implies also that time series of GRACE and GRACE-FO can be extended towards a climate series (Massotti et al., 2021).”

565-566: Cite source for GRACE data-product resolution

Done, see Lines 229-231:

“Although the mascon size is smaller than the inherent spatial resolution of GRACE of about 2.5°×2.5° or 64’000 km2 (Vishwakarma et al., 2018), the model exhibits a relatively high spatial resolution.”

566: What does "high flexibility" mean? Use specific wording about how STREAM works and what the processes are.

R: The sentence that in the previous version of the manuscript was:

“*The STREAM v1.3 model shows high flexibility also in the possibility to modify the sub-basin delineation and to introduce additional observational river discharge data to be used for the model calibration.*”

has been modified as (see Lines 667-670):

“Additionally, the STREAM model shows highly flexibility as: 1) it can accommodate application domains comprising single or multiple basins of any size; and 2) the sub-catchment delineation procedure can be easily adapted to introduce intermediate outlets along the river in correspondence of gages with available observed river discharge data, useful for model calibration.”

570: Define "very limited". Do you have quantitative information on CPU-hours per unit model time, e.g., from your examples?

R: The definition “very limited” was referred to the time needed to execute a model run that for the presented case study was less than 2 seconds on a machine with 16 GB RAM and 4 Core. This aspect has been specified in the revised manuscript.

The sentence in the previous version of the manuscript:

“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.”

has been modified as follows (see Lines 671-674):

“Due to its simplicity and the limited number of parameters to be calibrated, the computational effort for the STREAM model is very limited (model runs requiring seconds to minutes). For instance, a run of the STREAM model over the presented case study takes less than 2 seconds on a machine with 16 GB RAM and 4 Core.”

Reviewer #2

In addition to a good editorial review, the paper would benefit from an inclusion of the sensitivity analysis and a parameter regionalization study. The model and workflow could easily be adapted to other basins in the US where data are readily available to look into whether model parameters could be regionally attributed to some kind of land-cover characterization. A good editorial review could give the paper more room to include these analyses. The availability of the data are an issue, with both soil moisture and GRACE data only available through 2016. A more clear picture as to the potential resolution of that would be informative.

I'd like to see a really good overview figure which shows an example of each one of the input data sets in it's gridded form, drawn to scale, and mapped to the basins, as well as a time slice of the model showing precipitation, temperature, runoff, and TWSA over the basin. This would be helpful when describing the data used, and also the STREAM method.

I've added many editorial comments to the text, these should be considered suggestions and were simply my way of trying to highlight where I was having difficulty reading the text. There are other more substantive comments as well.

R: We thank the reviewer for the supportive review that allowed us to improve the manuscript. According the reviewer' suggestions, we carried out the following additional analysis and modifications to the manuscript:

1) a sensitivity analysis that allowed us to highlight the main parameters controlling the STREAM model output. The analysis has been introduced in paragraph 5.3. The paragraph 6.3 collects the related results;
2) a first attempt of model parameters regionalization. The results of this analysis are reported below but the authors, as better specified in the following, would like to not include them in the revised manuscript;
3) we specified the problem about the input data availability;
4) we accurately revised the model description and we added a figure to illustrate the input data used over the basin;
5) we carefully modified the text following the reviewer comments. For this part we do not reply line by line. We greatly appreciated the effort of the reviewer in improving the manuscript.

<u>1. Sensitivity analysis.</u>

In the reviewed version of the manuscript we added a sensitivity analysis based on the Variance-Based sensitivity method (VBSA, (see paragraph 5.4). We modified the text as follows (see Lines 475-492):

"To investigate how the variation of the STREAM parameters influences the variation of the STREAM model outputs, a global sensitivity analysis has been carried out. Specifically, the Variance-Based sensitivity analysis (VBSA, Sobol 1993) implemented into the Sensitivity Analysis For Everybody toolbox (SAFE, Pianosi et al., 2015, https://www.safetoolbox.info/) has been applied. VBSA relies on the variance decomposition and consists of assessing the contributions to the variance of the model output from variations in the parameters. In this study, we use as sensitivity index the first-order (main effect) index, which measures the variance contribution from variations in an individual input

factor alone (i.e., excluding interactions with other factors) and the total sensitivity indices, which measure the total contribution of a single input factor or a group of inputs including interactions with all other inputs. The following steps were carried out to execute the VBSA. Firstly, the locality-sensitive hashing (LSH) technique was used to generate 15000 samples from the model parameter space (see Table 1A). Previous hydrological studies (e.g., Tang et al., 2007) recommend the LHS sampling method for its sampling efficiency. Secondly, 15000 STREAM model runs were executed and the corresponding KGE values (11x15000 values, one for each gauging station for each run) were retained. Thirdly, the parameters and the 15000 KGE samples were used in the SAFE toolbox to compute the sensitivity indices.
For major details on the workflow needed to implement the VBSA the reader is referred to Noacco et al. (2020).”

The results of this analysis are illustrated in paragraph 6.3:

“The results of the VBSA, are illustrated in Figure 7a in terms of main effect indices and in Figure 7b in terms of total effect. Specifically, the figure refers to Vicksburg station but similar results have been obtained for all the 11 gauging stations in the Mississippi basin. By looking at Figure 7, we observe that the model parameters most influencing the model response are β and m, i.e., the two parameters controlling the slow-flow runoff response of the lower soil storage. In particular, the total effect sensitivity index of these two parameters is higher than the main effect sensitivity index. This means that these two parameters have an effect on the model output not only through their individual variations but also through interactions with other parameters. Instead, the other five parameters (α, T, γ, C, D and Cm) have low main and total effect indices, and consequently, these parameters have a small effect, both direct and through interactions, on model response. Among these, only the α parameter shows a slightly high main and total effect sensitivity indices.

This outcome is very important as it allows to clearly distinguish model parameters which values should be carefully determined when calibrating the model (β and m and partially α) from the least sensitive (T, γ, C, D and Cm) which values could be set values within the model parameters’ range of variability and then excluded during the calibration phase.”

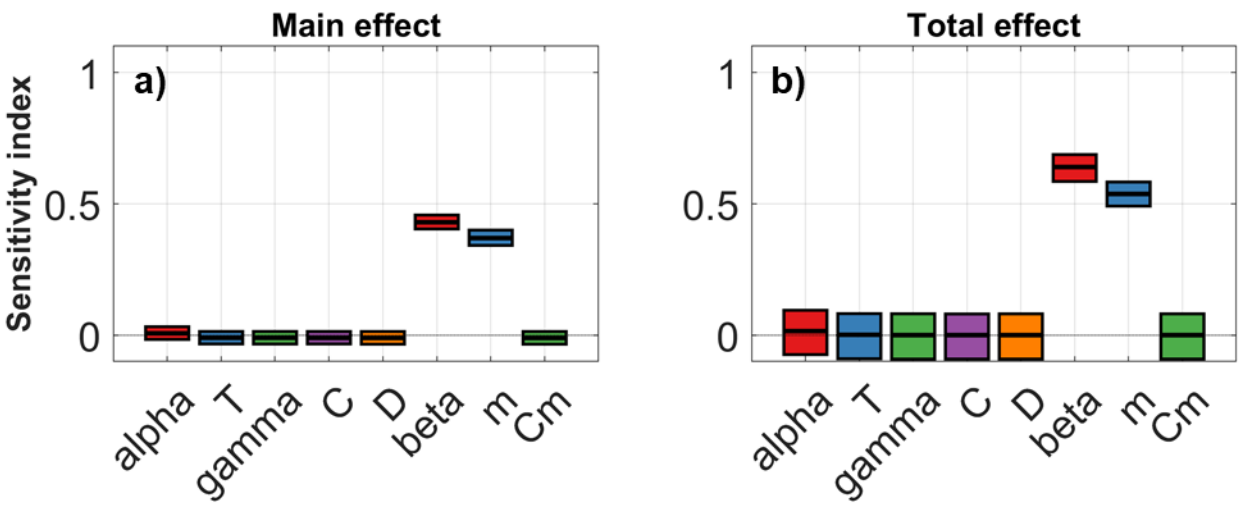

Figure 7. Main effect a) and total effect b) sensitivity indices calculated using the VBSA method for Vicksburg gauging station. The boxes represent the 95% bootstrap confidence intervals and the central black lines indicate the bootstrap mean.

2. Model parameter regionalization

The reviewer is right that the model and the workflow could easily be adapted to other basins in the US where data are readily available. However, the intent of the manuscript was to introduce a first version of the STREAM model, to highlight the limits and advantages of the model with respect to a classical hydrological model and, above all, to demonstrate the usefulness of satellite data in supporting the river discharge modelling. For that, we believe that the application of the model to other basins and carry out a parameter regionalization study is material for a future paper.

Nevertheless, as this aspect seems crucial for the reviewers and the Topical Editor, we addressed a fist attempt of model parameters regionalization. Indeed, for the application of the STREAM model in the Mississippi basin, we identified 53 sub-basins each one with own characteristics in terms of land-cover, topography, climate, etc. Through the model calibration, for the same sub-basins we identified 5 group of parameters (each one for each calibration site) and we tried to identify specific relationship between them and the sub-basin characteristics. In particular, based on the results of the sensitivity analysis, we fixed some model parameters (the less sensitive, i.e., gamma, T, Cm, C and D) to constant values (equal to the mean values obtained for the 56 sub-basins) whereas we search plausible relationships with the basin characteristics for the others (i.e., alpha, beta and m). Specifically, we linked alpha to the aridity index, beta and m to the bulk density.

The results of this first attempt of model parameter regionalization are shown in the following. Figure 1R shows the performances, in terms of KGE, obtained through the calibrated parameter set (on the left upper map) and the variation on model performances obtained by considering the regionalized model parameters (on the right upper map). The bottom plot of Figure 1R shows the simulated river discharge obtained by considering the calibrated (STREAM cal.) and the regionalized (STREAM reg.) parameter sets. Figure 2R illustrates the performances of the model over the sections used for the model calibration.

As expected, both the figures illustrate a worsening of model results when the regionalized set of parameters are used for running the STREAM model. However, the deterioration is lower than 30% and it can be retained acceptable.

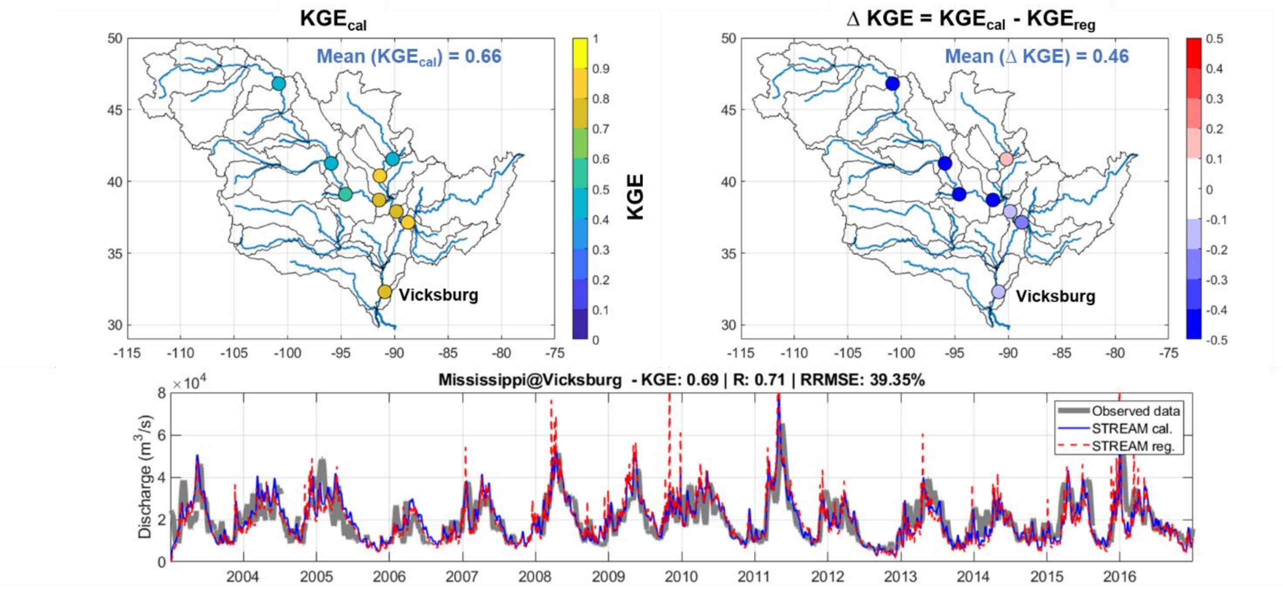

Figure 1R. Performances obtained through the STREAM calibrated/ regionalized (left/right map) parameter sets.

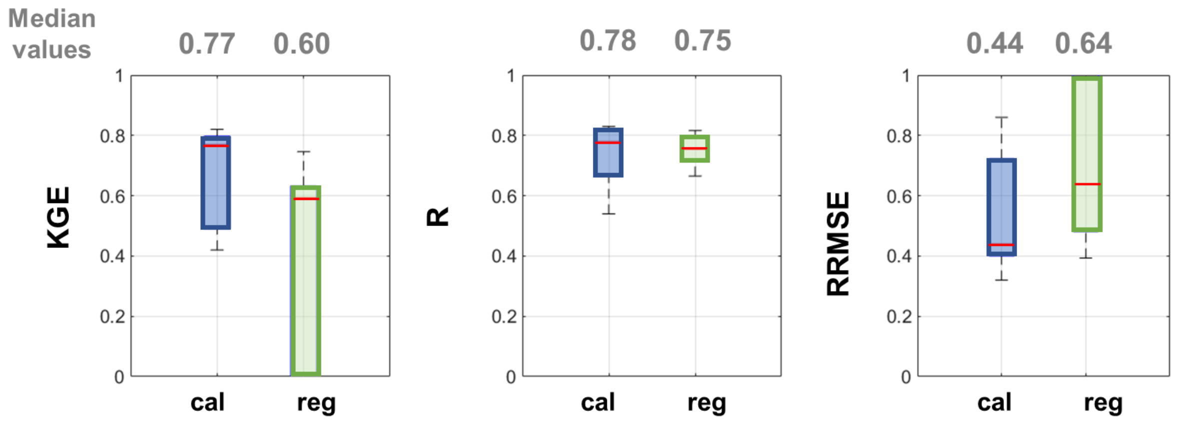

Figure 2R. Performances scores over the sections used for the model calibration (see Table 2 in the manuscript). Cal and reg stand for to calibrated and regionalized parameter set.

Although these results are very interesting, we remark that this is just a preliminary analysis for the regionalization of the model parameters and for the reasons expressed above, the authors would like do not include this part in the manuscript. We hope the reviewer could understand our point of view.

3. Input data availability

The soil moisture and GRACE data availability is not an issue. Soil moisture data are available from 1978 to present (2022) and can be freely downloaded at:

on https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture?tab=form.

In this respect there was an error in the previous version of the manuscript, as it was written:

"In this study, the daily combined ESA CCI soil moisture product v4.2 is used and it is available at global scale with a grid spacing of 0.25°, for the period 1978-2016."

In the revised version of the manuscript, we have corrected the sentence as follows (see Lines 219-221) :

"In this study, the daily combined ESA CCI soil moisture product v4.2 is used. It is available at global scale with a grid spacing of 0.25°, for the period 1978 to present.

For GRACE data, there is a data gap between the end of GRACE mission (October 2017) and the launch of GRACE-FO (May 2018). Depending on the solutions (spherical harmonics, SH or mass concentration, mascons) considered for the GRACE data, the period of missing data spans, at most, from August 2016 (for mascons, like in the present study) to June 2018. In literature there are numerous studied intended to fill this data gap (e.g., Landerer et al., 2020) that could be used to extend the analysis period for the STREAM model.

We addressed this aspect in the revised version of the manuscript (see Lines 242-248).

"The GRACE data used here are available from January 2003 to July 2016, which suffices to demonstrate the STREAM capabilities. With its successor mission GRACE Follow-On (GRACE-FO), launched early 2018, the time series of time-variable gravity has reached a nearly uninterrupted time span of about 20 years, thus allowing a continued and operational use of STREAM. The existing interruptions, short ones due to mission operations or technical failures, but also the one-year gap between GRACE and GRACE-FO can be dealt with in various ways, e.g. by data driven gap filling (Yi and Sneeuw, 2021)."

and Lines 596-609:

"This outcome along with the one obtained in the paragraph 6.3, demonstrating the high sensitivity of the model parameters related to slow-flow river discharge component, confirm the paramount role of TWSA in estimating river discharge. In this respect, the availability of GRACE data up to July 2016 could represent an issue for the model application beyond that date. However, the GRACE-FO along with the numerous literature studies devoted to fill the GRACE data gap between GRACE and GRACE-FO (see e.g., Landerer et al., 2020 or Yi and Sneeuw, 2021), can provide the needed data to extend the STREAM model application up to present. Further developments in this direction are expected with the ESA's Next Generation Gravity Mission (NGGM), a candidate Mission of Opportunity for ESA–NASA cooperation in the frame of the Mass Change and Geosciences International Constellation (MAGIC) that will enable long-term monitoring of the temporal variations of Earth's gravity field at relatively high temporal (down to 3 days) and increased spatial resolutions (up to 100 km). This implies also that time series of GRACE and GRACE-FO can be extended towards a climate series (Massotti et al., 2021)."

Moreover, we specified that to run the STREAM model, near real time data could be used. For the purpose, GPM Early run precipitation data and the EUMETSAT H16 soil moisture data could be used. In addition, the feasibility of providing products derived from GRACE in near-real time has been demonstrated in the Horizon 2020 project European Gravity Service for Improved Emergency Management (EGSIEM, Jäggi et al., 2019).

We add a sentence in the text as (Lines 662-666):

"For instance, satellite missions with higher space/time resolution (e.g., GPM Final Run, ASCAT and NGGM-MAGIC) or near-real time products (e.g., GPM Early Run, EUMETSAT H16, GRACE European Gravity Service for Improved Emergency Management, EGSIEM GRACE data Jäggi et al., 2019) could be considered."

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., ... & Yuan, D. N. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. Geophysical Research Letters, 47(12), e2020GL088306.

4. Revision of model description and addition of a figure to illustrate the input data used over the basin

We in-depth revised the model description and modified Figure 1 (in the following) adding the four layers of gridded input data needed to run the model (Figure 1b) and the times series for a point inside the basin (Figure 1c).

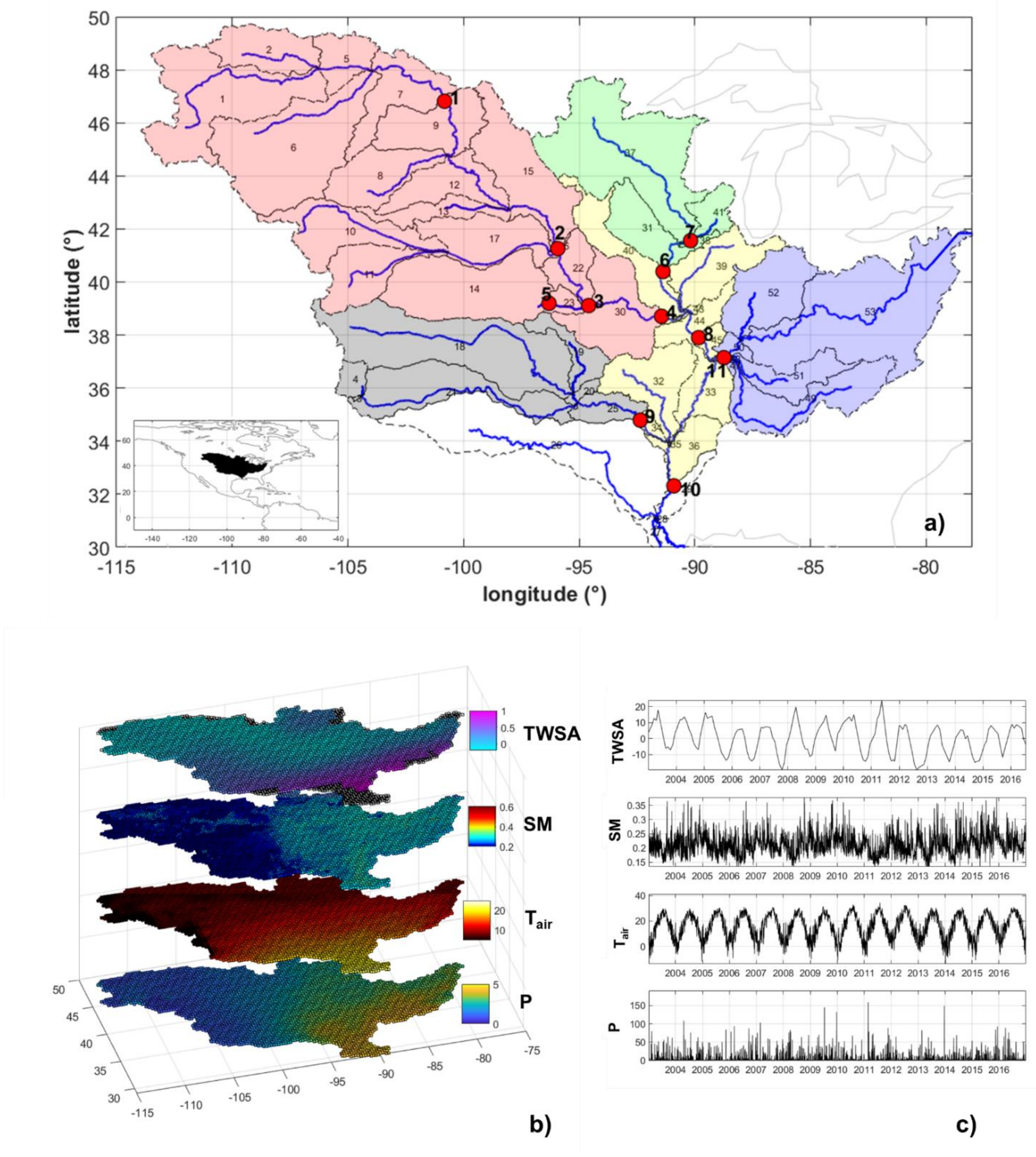

Figure 1. Mississippi river basin. Figure 1a) illustrates the sub-catchments delineation. The black dashed lines and the numbers in the map identify the 53 sub-catchments (tributary and directly draining areas) in the Mississippi basin, blue lines represent the mainstem of each sub-catchment. Red dots indicate the location of the river discharge gauging stations; different colours identify different inner cross-sections (and the related contributing sub- catchments) used for the model calibration. Figure 1b) shows the gridded mean daily values of the input data for the period 2003-2016. Figure 1c) illustrates the input time series over a point located inside the basin.

About the model description, we modified the manuscript (see Lines 259-378) and Figure 2 (see below) to better describe the processes modelled at the cell scale, at the sub-catchment scale or along the river network. Accordingly, we divided section 4.2 into three sub-sections. i.e., "4.2.1 Runoff generation at cell scale", "4.2.2 Sub-Catchment Streamflow Calculation" and "4.2.3 Runoff Routing through River Networks". Please refer to Lines 257- 370 for the modifications

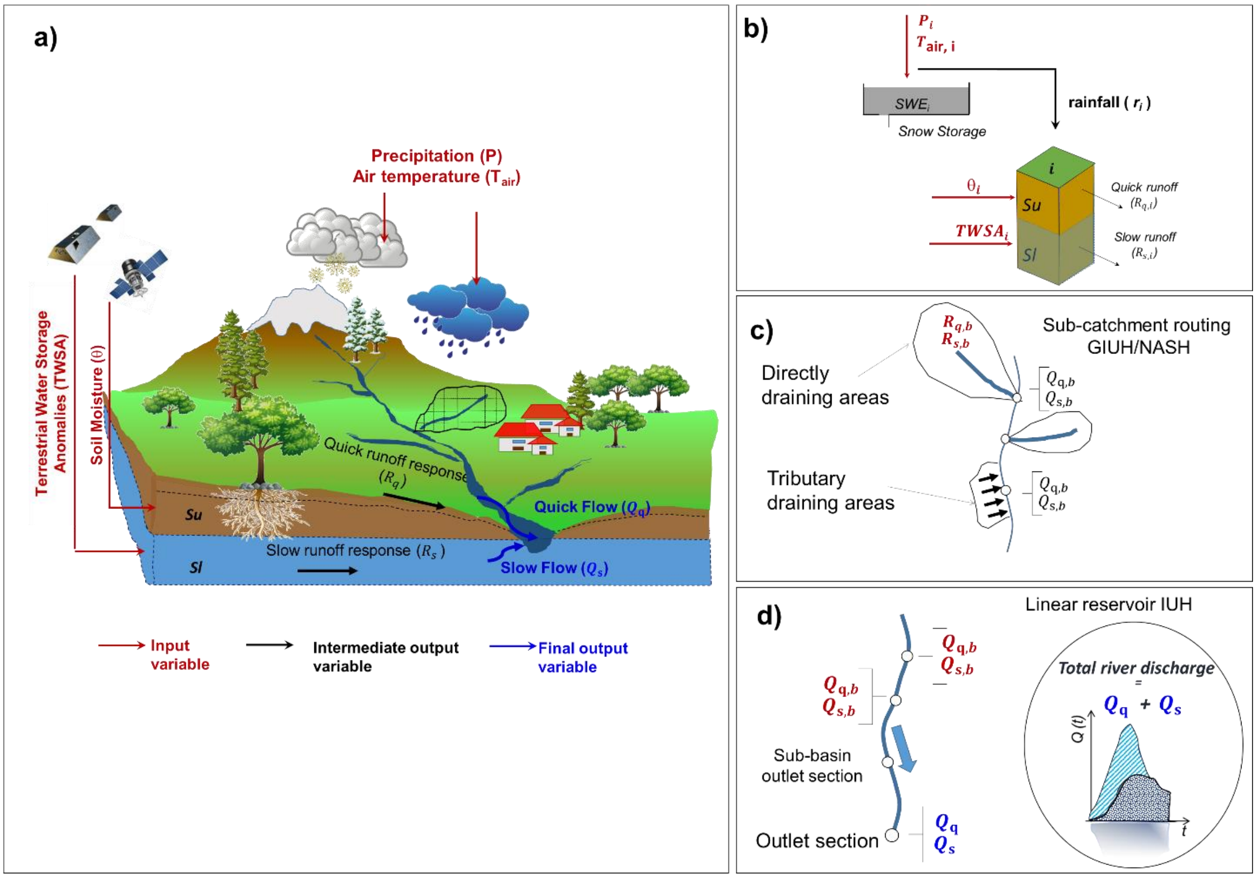

Figure 2. Configuration of the STREAM model adopted for runoff and river discharge estimation. Figure 2a) gives an overview of the needed input data and the variables can be obtained as model output. Figure 2b) illustrates the runoff generation at cell scale. Figure 2c) refers to the sub-catchment river discharge calculation and Figure 2d) illustrates the river discharge routing through river networks. Red arrows indicate input variables; black arrows indicate intermediate output variables; blue arrows indicate final output variables. Please refer to text for symbols.