

Hydrol. Earth Syst. Sci. Discuss., author comment AC4
<https://doi.org/10.5194/hess-2022-204-AC4>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC2

Mu Xiao et al.

Author comment on "On the value of satellite remote sensing to reduce uncertainties of regional simulations of the Colorado River" by Mu Xiao et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-204-AC4>, 2022

In this document, we provide detailed answers to the comments raised by Reviewer 2 on our manuscript "On the Value of Satellite Remote Sensing to Reduce Uncertainties of Regional Simulations of the Colorado River" by Mu Xiao et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2022-204-RC1>, 2022.

Our answers are reported in italic after the Reviewers' comments that we have numbered.

General comment:

The authors present a well-written and well-motivated study on the value of remote sensing data to improve hydrological simulations. I enjoyed reading the manuscript and only have a few comments that will require some attention by the authors.

Thanks for reviewing the paper and helping us to improve it. Below, we provide point-to-point responses to the Reviewer's comments.

1. The authors chose to evaluate the model against monthly runoff. In my point of view this cannot be justified since the remote sensing data are used at daily scale to evaluate the model. In order to achieve a balanced evaluation of the model runoff should also be evaluated at daily scale.

This is a very good point that was mentioned by both Reviewers. The river is heavily regulated and the highest resolution of the naturalized flow records from the US Bureau of Reclamation (USBR) that is currently available is monthly. This is the main reason why we could not extend our calibration and validation against discharge at a daily scale. We also note that other modeling studies of the Colorado River Basin are based on the same monthly dataset.

Given the lack of daily streamflow data, we might argue that adding daily remotely sensed products to the model testing phase is even more critical to capture daily dynamics.

2. More details are required with respect to the "adjustment of the VIC parameters" as presented in section 3.2. Did the authors conduct a manual calibration or were the parameter values estimated via automatic calibration? Please specify.

It is a manual calibration. For the baseline simulation of section 3.2, we manually changed

the soil parameters to improve streamflow performance in the "traditional" way. In the revised version of the manuscript, we will specify in Sections 3.2 and 3.3 that the calibration performed in the baseline and the other steps are manual.

3. One of my main concerns relates to how the fit between observed and simulated spatial patterns was assessed. The authors chose to do a grid-wise evaluation of the simulation. I think this makes sense for the forcing adjustment where the temporal dynamics of simulated LST are strongly linked to the forcing data (air Temp). However, when it comes to model parameters, I would suggest to evaluate the model against spatial pattern that are aggregated over time, for example a long-term average annual (or summer) LST map. Evaluating a model at daily scale will always be very much affected by the quality of the forcing data and the model parameters have a limited affect here. Nevertheless, the imprint of the model parameters emerges when aggregating the simulation results over time and quantifying the spatial pattern match (e.g. with help of the SPAEF metric (<https://doi.org/10.5194/gmd-11-1873-2018>)) instead of the grid-to-grid comparison. Along these lines, the spatial patterns of RMSE and Bias presented in Figure 9 do not show a clear improvement of the model developments. Maybe a spatial pattern oriented evaluated of the long term average LST patterns is more insightful.

Thanks for pointing this out. Reviewer 1 had also a similar comment related to the need to use metrics that quantify the match of spatial patterns. However, there is a detail of our approach that, perhaps, was did not properly explain. As mentioned in lines 255-258: "The first two steps [of the calibration] were guided by metrics quantifying the agreement between simulated and remotely sensed LST, including the correlation coefficient (CC), root mean squared error (RMSE), and Bias (mean LSTV - mean LSTM) between: (1) time series of daily LSTV and LSTM at each grid cell, and (2) daily spatial maps".

In particular, the maps and metrics shown in Figures 5, 6, 7, and 9 that drove the calibration effort are based on RMSE, CC, and Bias between simulated and observed time series at each pixel. Therefore, for these cases, it is not possible to compute spatial metrics like SPAEF and SSIM (as suggested by Reviewer 1). The only case where we computed metrics between maps is Figure 8, which we used as additional measures of calibration accuracy.

Regarding the option to use long-term averages of LST, we decided not to do so because we wanted to assess the model's ability to capture dynamics at higher temporal resolutions, especially considering that we do not have streamflow data at a daily scale, as mentioned in the answer to comment #1.

To address this comment, we plan to:

- *Clarify that metrics based on pattern-oriented metrics cannot be applied to the maps resulting from pixel-based time series errors.*
- *Add in the revised manuscript figures showing the changes in the long-term average of LST and SC (Figures R1 and R2 in the attached document), along with a table (Table R1 in the attached document) reporting SPAEF and SSIM that target the ability to match the spatial patterns, as also requested by Reviewer 1. As shown in table R1, the values of these metrics are in line with the overall trend of RMSE and Bias.*

4. The authors only present maps of the three selected metrics. For interested readers, observed and simulated maps of the actual variables will provide insightful information. The authors could select single days or long-term averages of LST (night and day) and snow cover to illustrate the catchment characteristics and how the model represents those.

Thanks for the suggestion. As mentioned in the answer to the previous comment, in the

revised version we will include maps of long-term averages of LST and SC (Figures R1 and R2 in the attached document).

5. It would be interesting to see how the three steps of model development affect the water balance of the model. The authors only present the simulated runoff of the various simulations, but aggregated numbers of evapotranspiration, runoff, groundwater recharge, etc. would provide relevant alternative information.

This is another good suggestion. In the Revised manuscript, we will add the annual components of the water balance, including precipitation, evapotranspiration, and changes in water storage, so that the readers can have better information about how the water balance changes during our stepwise experiments.

Please also note the supplement to this comment:

<https://hess.copernicus.org/preprints/hess-2022-204/hess-2022-204-AC4-supplement.pdf>