Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., author comment AC1
https://doi.org/10.5194/hess-2022-204-AC1, 2022

Reply on RC1

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-AC1, 2022

In this document, we provide detailed answers to the comments raised by Reviewer 1 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:

I enjoyed reading the manuscript. This study combines several remote sensing products to improve the hydrologic model's physics together with streamflow performance. I only have several concerns regarding the presentation of the work and the framework.

Thanks for working with us to improve this paper. We appreciate the constructive comments and we have provided detailed responses below.

1. L159: USBR dataset needs a reference (url/doi)

The original U.S. Bureau of Reclamation (USBR) webpage that provides the flow data records will be included in the updated version: https://www.usbr.gov/lc/region/g4000/NaturalFlow/provisional.html

The naturalized flow data we obtained from USBR will also be uploaded onto the same Zenodo online archive where we stored the model parameters (it will be mentioned in the "Open Research" section in the updated paper).

2. L239: why monthly and not daily streamflow performance was targeted in calibration? daily water balance is key for hydrologic models. Monthly fit is easier and reducing the value of baseline simulation.

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 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.

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.

3. L248-Fig3: in this section (3.3) I read what has been done but I couldn't find answers for the question "how". Framework needs elaboration. Baseline simulation is clear but other steps are not clear.

We acknowledge that we have not provided an in-depth description of the "how" in each step mentioned section 3.3. The reason is that we considered section 3.3 as an overview of the calibration methods, while we preferred reporting the details of the Forcing-adj, Veg-adj, and Snow-adj steps in sections 4.3.1, 4.3.2, and 4.3.3 of the Results, respectively. To address the Reviewer's comment, we will:

- *change the name of Section 3.3 to "Model improvements with remote sensing products: overview of stepwise calibration strategy";*
- *mention in section 3.3 that the details of each step are better described in sections 4.3.1, 4.3.2, and 4.3.3 of the Results.*

4. Most importantly, model calibration is an exercise of fine tuning of the model parameters. Before calibration a robust sensitivity analysis (SA) must be applied for such sophisticated models with many parameters to reduce the search space. Did authors apply SA in their study?

We did not apply a systematic robust SA, but we adopted a hybrid approach based on the physics and sensitivity of single parameters. Specifically, we first carefully considered the physical equations implemented in the model to simulate the processes related to the observed variable (i.e., land surface temperature, LST, snow cover, SC). These equations are reported in the manuscript Appendix. We then identified the set of key parameters that are (1) involved in the equations, (2) spatially variable, and (3) not derived from any type of direct or indirect observation. We then computed the spatial correlation between these parameters and the pattern of the errors between simulated and remotely sensed LST. The outcomes of these analyses are reported in Figure 7 of the manuscript. For the parameters with the largest correlation, we performed a sensitivity analysis to verify that that parameter importantly affects the simulation of LST, although we did not mention it in the paper. We ultimately focused on the subset of parameters that exhibited the largest sensitivity.

5. The authors followed a stepwise approach but sensitive analysis (sobol's, LHS O-A-T, Morris etc) may reveal parameter interactions which can be important to consider during calibration. The authors should discuss the implications of parameter interaction in their framework.

We agree with Reviewer 1 that a sensitivity analysis targeting multiple parameters at the same time can ultimately lead to improved performance. However, this type of analysis would be very computationally expensive for a model like VIC and the size of the Colorado River Basin and would most likely require an entirely separate study. Here, our focus is instead to highlight the importance of accounting for spatially variable observations from remote sensors in the calibration process. As discussed in the answer to the previous comment #4, we focused the calibration on a set of parameters identified based on model physics and the impact of these parameters on the spatial variability of the errors between simulated and remotely sensed LST and SC. The calibration was then performed by focusing on one parameter in each step.

That said, we will highlight in the paper that conducting a sensitivity analysis targeting *multiple parameters is one of the subsequent research tasks to further advance the use of remotely sensed products for model improvement.*

6. It would be good to simultaneously use LST and snow RS data on uncertainty reduction via model calibration.

We agree with the Reviewer that this would be an interesting idea. However, it would require the use of an automatic calibration routine with an optimization function that accounts for both daily LST and monthly SC, which would require significant computational power and would be out of the scope of this study, as for the case of the multiparameter sensitivity analyses (comment #5). To our knowledge, studies that investigated the utility of remotely sensed products in large basins like the Colorado River are still very few; therefore, it is still necessary to separately gain insights into the utility of each remotely sensed dataset, along with the associated parameters and equations. Once this knowledge based on single variables is built, calibration strategies that target multiple variables could be better designed in future work.

7. My biggest concern is about the spatial structure of the selected hydrologic model (VIC) which is a semi-distributed model. In such model parameters get the same value in the same subbasins which inevitably leads to uniform parameter fields and resultant uniform flux maps. One way to avoid this, is using fully distributed models with parameter regionalization tool based on pedo-transfer functions using soil and vegetation properties.

We appreciate the Reviewer's comment and point out three issues that, hopefully, address this concern.

First, VIC is a macroscale hydrologic model with a gridded domain where most of the parameters and all outputs do vary spatially. For example, all parameters identified in Fig. 7 vary in space. See also the maps in Figs. 1d and 1e that report the vegetation fraction, f_v*, and soil depth that are used in the baseline simulation.*

Second, the parameters of the baseline simulations are mainly based on the products derived by Bohn and Vivoni (2019), who utilized high-resolution (from 500 m to 1 km) remote sensing products to generate several model parameters in the same grid at 1/16°(~6-km) resolution used in our manuscript. This is mentioned in Section 3.2 of the manuscript.

Finally, we have more than 15,000 pixels for the Colorado River Basin which allow the spatial variability to be appropriately captured.

8. The authors used bias-sensitive error metrics (rmse, bias) and CC as bias insensitive metric. CC must be used with cautious it can be affected by outliers in the sample. High CC values are not always informing. Instead spatial metrics (SSIM, FSS etc) could be preferred.

We would like to clarify one important issue. 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 SSIM, FSS, etc. The only case where we computed metrics between maps is Figure 8, which we used as additional measures of calibration accuracy.

Regarding the role of CC, we fully agree with the Reviewer. When we carried out our analyses, we noted that the CC alone cannot be a good indicator for model evaluation because it is not robust, as pointed out by the Reviewer, and because it is always high (>0.8) even in the baseline (Lines 475-476). Because of this, our model adjustments (Forcing-adj, Veg-Adj, and Snow-adj) are mainly based on either RMSE or Bias of the time series.

To address the concerns of both Reviewers, in the revised manuscript, we computed the Structural Similarity Index Measure (SSIM) and the Spatial Efficiency metric (SPAEF) of long-term LST climatology (see Tables R1 in the attached document). The values of these metrics are in line with the overall trend of RMSE and Bias; the table will be then added to the Supporting Information of the revised manuscript version.

9. Fig6: the readers can be curious why median night time bias for baseline is usually less than other 3 cases.

As shown in Fig. 6, the median bias for nighttime LST in the baseline case is negative. This result was ascribed to the negative bias of the air temperature forcings from the Livneh dataset, which we removed using the PRISM long-term normal products in the "Forcing-adj" calibration step. The resulting median bias for nighttime LST becomes close to 0 or slightly positive in the "Forcing-adj" step, thus improving the simulation (see also Figures 8 and 9). The bias does not change significantly in the other steps because they do not involve changes in parameters that affect nighttime LST. These details are described in Section 4.3.1, where we concluded that "…the Forcing-adj simulations improved Bias, which was reduced in most subbasins" (see Lines 370-371).

10. Fig7 should be better explained. How correlation between parameters is assessed?

The correlation coefficients in Fig. 7 are the Pearson Correlation Coefficient (CC) between two spatial fields in each subbasin. The first spatial field is either T_{air} or any of the parameters shown in the rows of the heatmaps (e.g., Elevation, Porosity, …, LAI, and f_V), while the second spatial field is either RMSE (heatmaps on left) or the Bias (heatmaps on right) between time series of LST_V and LST_M at each domain pixel. The columns in each heatmap report the correlation coefficient in each subbasin. For example, the first pixel on the top left is showing the correlation coefficient between T_{air} and Daytime RMSE in the Green subbasin.

In the revised version of the paper, we will update the text to better present this information.

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2022-204/hess-2022-204-AC1-supplement.pdf