

Replies to reviews

“Same Streamflow, Different Water Stories: The Hidden Impacts of Streamflow-Only Calibration in Distributed Hydrological Modeling”

Nicolás A. Vásquez, Pablo A. Mendoza, Wouter Knoben, Martyn Clark, Tricia Stadnyk, and Naoki Mizukami

We thank the reviewer for the time and constructive comments on our paper. We provide responses to each individual point below. For clarity, the reviewer's comments are given in italics, and our responses are given in plain blue text.

Reviewer #2

This paper presents a valuable investigation of an important topic: how calibration strategies for distributed hydrological models affect not only streamflow performance, but also the physical realism of internal hydrological processes. The proposed regularization framework is interesting, and the results are generally well presented. Overall, I commend the authors for this useful contribution. I have a few comments that may help further improve the manuscript.

We thank the reviewer for the positive assessment of our work and for the constructive comments. We address each point in detail in our point-by-point responses to the specific comments.

General comments

1. The authors use Principal Component Analysis (PCA) to regularize the spatial distribution of model parameters. This is an interesting approach, but its physical basis could be explained more clearly. In particular, it would be useful to clarify what PC1 represents in each catchment. Does it mainly reflect soil texture, elevation, slope, or a combined physiographic gradient?

We thank the reviewer for this constructive suggestion, which echoes a related comment from Reviewer #1 regarding the physical interpretation of our PCA-based regularization. PC₁ was highly and significantly correlated with at least four grid cell attributes at Claro, Cautín and Futa, and two attributes in the remaining basins. In particular, correlations tend to be generally higher between PC₁ and elevation, clay content and sand content across basins. We have explicitly added this explanation in section 4.2.

The authors should also justify why only PC1 is used. Some hydrologically relevant spatial patterns may be represented by higher-order PCs, even if they explain less total variance. Reporting the PCA loadings and the variance explained by PC1 would make the method more transparent.

In this study, we only used the first component (PC₁) because (i) it explained between 46% (Colorado River basin) and 79% (Cautín River basin) of the total variance in the physiographic variables considered across the six catchments, and (ii) PC1 was highly and significantly correlated with at least two grid cell attributes in all basins. We have included this justification in section 4.2. Additionally, we will update Figure S1 (Supporting Information) to include a) the fraction of total variance explained by each component; b) two first principal components biplot.

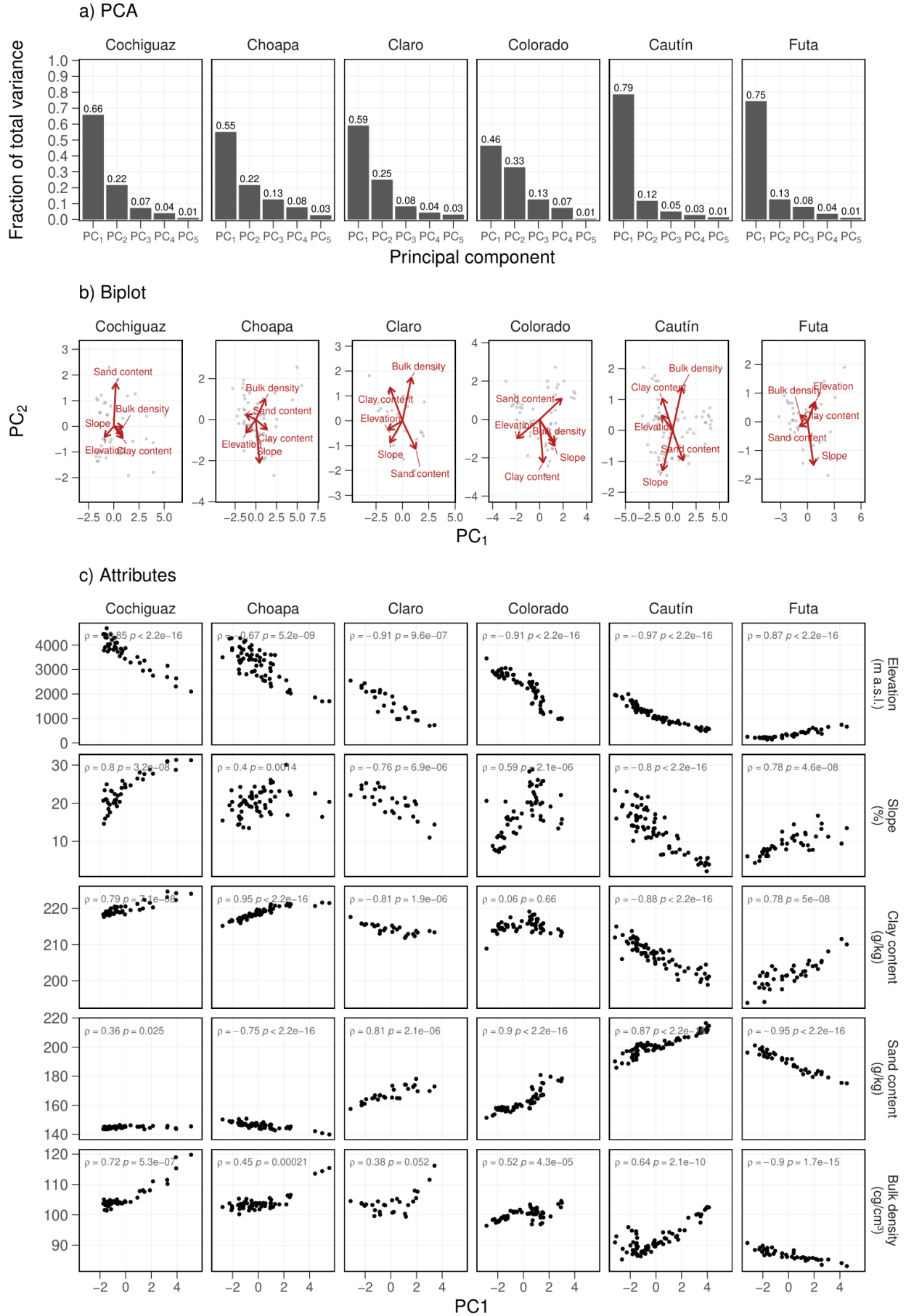


Figure S1. Principal component analysis to derive *a priori* parameter fields: a) fraction of total variance explained by each component; b) two first principal components biplot, and c) relationship

between PC_1 and the attributes, along with Spearman rank correlation coefficients and p-values. Each dot represents a grid cell ($0.05^\circ \times 0.05^\circ$ horizontal resolution).

Regarding the number of PCs considered for spatial regularization, we conducted additional calibration experiments by maximizing $KGE(Q)$ in order to compare the following cases: (i) spatially uniform parameter sets; (ii) distributing only $b_{infiltr}$ as a function of PC_1 (i.e., $b_{infiltr} = a + b \cdot PC_1^c$); and (iii) distributing only $b_{infiltr}$ as a function of PC_1 and PC_2 (i.e., $b_{infiltr} = a + b \cdot PC_1^c + d \cdot PC_2^e$). A comparison of simulated annual cycles of streamflow (Figure R1) and other variables (Figure R3), as well as daily flow duration curves (Figure R2), will be included in the Supporting Information section. One can note that very similar $KGE(Q)$ values and streamflow annual cycles are obtained during the calibration period (Figure R1), and that the two regularization strategies applied to $b_{infiltr}$ improve $KGE(Q)$ during the evaluation period in Cochiguaz and Choapa. Additionally, the simulation of low flow volumes improves in Colorado and Futa when both PC_1 and PC_2 are considered for spatial regularization (Figure R2). However, the examination of simulated annual cycles of other variables than streamflow (Figure R3) reinforces that good streamflow simulations do not necessarily translate into annual cycles closer to observational references of other variables (see ET' results for Claro, Colorado, Cautín and Futa in Figure R3). We will include this point – i.e., that the main message of the manuscript does not change by including PC_2 in the parameter regularization scheme – in the discussion section of the revised manuscript.

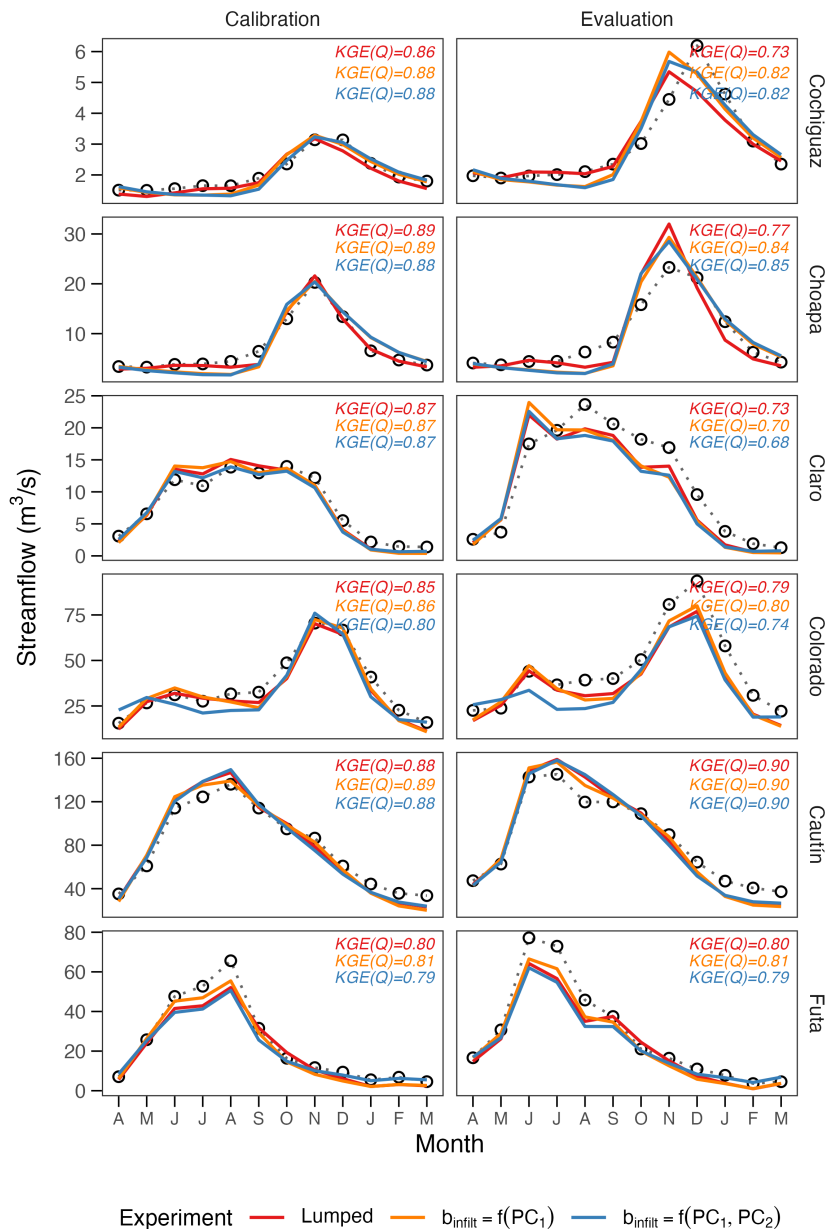


Figure R1. Simulated streamflow annual cycles for parameter sets obtained by maximizing KGE(Q) for the following cases: (i) spatially uniform parameter sets; (ii) distributing only b_{infiltr} as a function of PC_1 (i.e., $b_{\text{infiltr}} = a + b \cdot PC_1^c$); and (iii) distributing only b_{infiltr} as a function of PC_1 and PC_2 (i.e., $b_{\text{infiltr}} = a + b \cdot PC_1^c + d \cdot PC_2^e$). The calibration and evaluation periods are Apr/2005-Mar/2018 and Apr/1995-Mar/2005, respectively. The circles represent the observed mean monthly values.

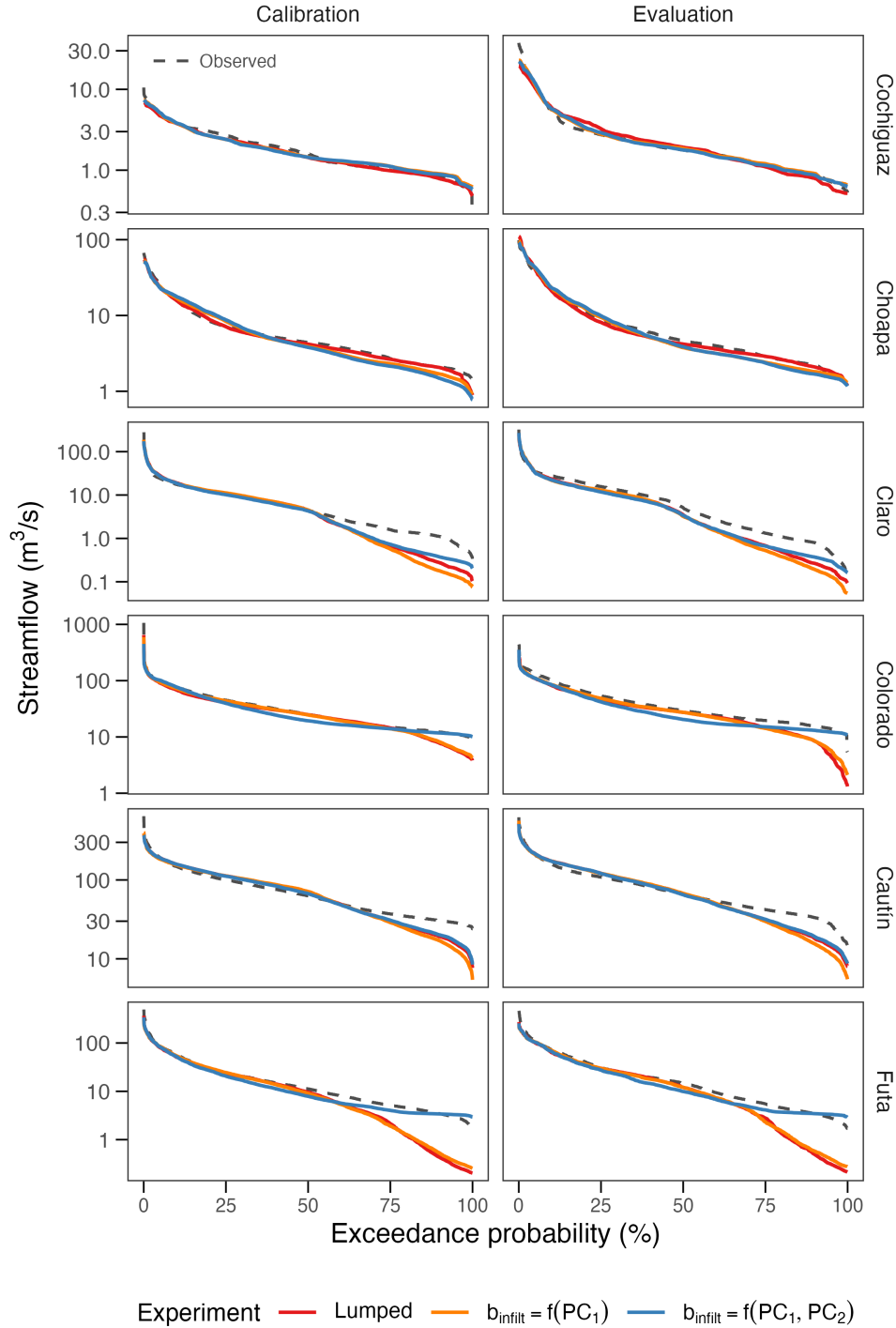


Figure R2. Simulated daily flow duration curves for parameter sets obtained by maximizing KGE(Q) for the following cases: (i) spatially uniform parameter sets; (ii) distributing only b_{infiltr} as a function of PC_1 (i.e., $b_{\text{infiltr}} = a + b \cdot PC_1^c$); and (iii) distributing only b_{infiltr} as a function of PC_1 and PC_2 (i.e., $b_{\text{infiltr}} = a + b \cdot PC_1^c + d \cdot PC_2^e$). The calibration and evaluation periods are Apr/2005-Mar/2018 and Apr/1995-Mar/2005, respectively.

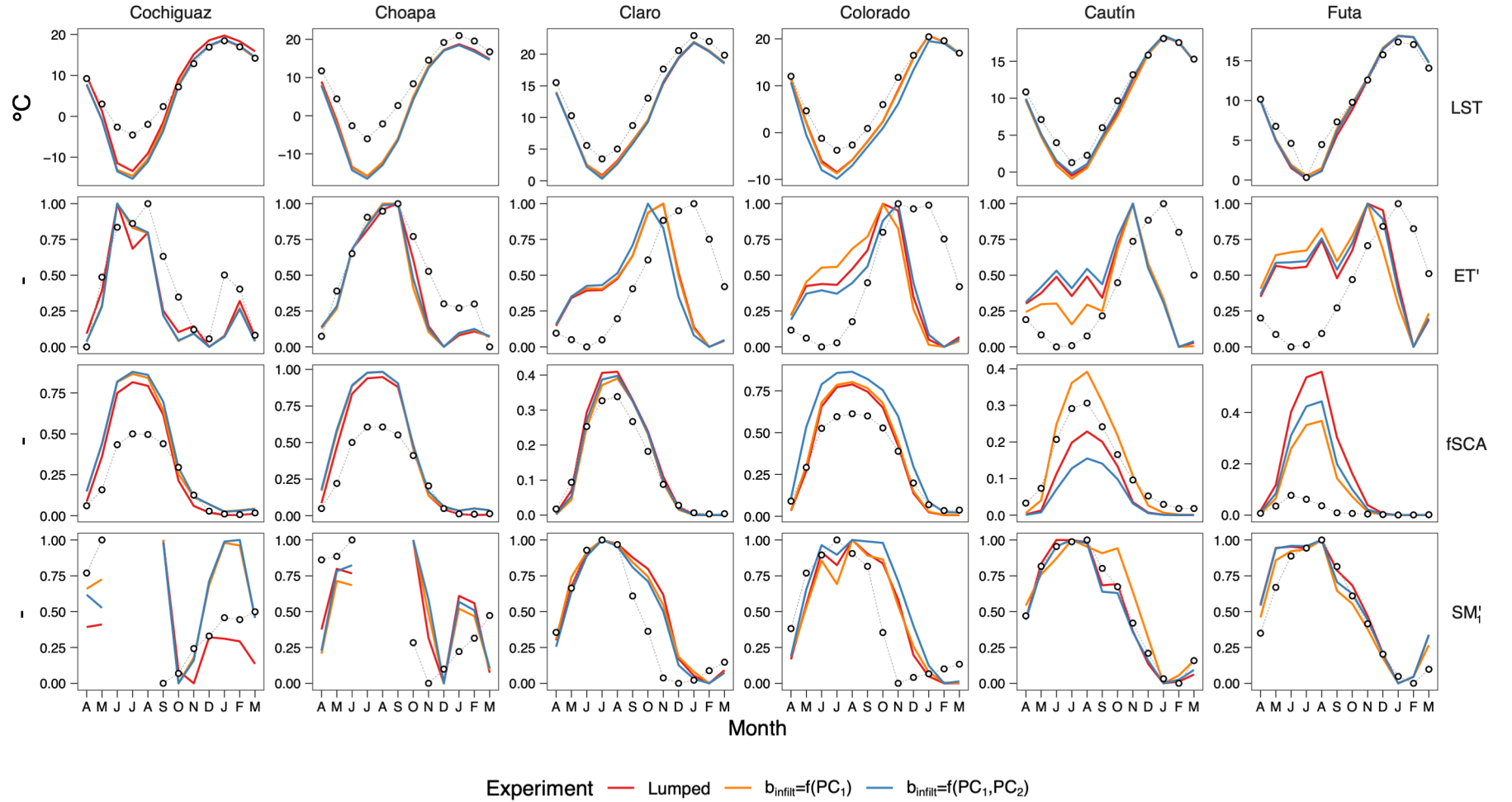


Figure R3. Simulated annual cycles of Land Surface Temperature (LST), dimensionless Evapotranspiration (ET'), fractional Snow Covered Area (fSCA) and Soil Moisture for the first (upper) soil layer (SM₁') for parameter sets obtained by maximizing KGE(Q) for the following cases: (i) spatially uniform parameter sets; (ii) distributing only b_{infil} as a function of PC₁ (i.e., $b_{infil}=a+b*PC_1^c$); and (iii) distributing only b_{infil} as a function of PC₁ and PC₂ (i.e., $b_{infil}=a+b*PC_1^c+d*PC_2^e$). The circles represent the mean monthly observational reference values, and results are displayed for the calibration period.

In addition, the authors may consider discussing the choice of input attributes used for PCA. Besides the selected soil and topographic variables, other descriptors such as land cover, vegetation type, rooting depth, drainage area, flow accumulation, or topographic wetness index may also be hydrologically relevant.

This is an excellent point. In this study, we only considered bulk density, clay and sand content, elevation, and slope in the PCA-based parameter regularization approach. Nevertheless, previous work in the Andes suggests that other hydrologically relevant attributes may be incorporated in this framework. For example, Molina et al. (2007) found that the observed runoff variability in the southern Ecuadorian Andes can be explained by a combination of surface vegetation cover and land use. Crespo et al. (2011) concluded that the temporal distribution of runoff above 2500 m a.s.l. in the southern cordillera of the Ecuadorian Andes is mainly controlled by the lateral saturated hydraulic conductivity, the soil water retention and the antecedent soil moisture content. Roa-García et al. (2011) showed that the transformation of headwater catchments from forests and wetlands to grasslands is most likely contributing to the reduction in their water regulation capacity in the Andes of Colombia. More recently, Gordillo & Pineda (2021) concluded that soil-bedrock interface and lithology exert a first-order control on hydrologic partitioning of volcanic mountain sub-catchments of the Mira River basin, Ecuador.

In response to this comment, we will expand the discussion of the input attributes used for PCA by incorporating the above information.

2. The calibration results may depend on the random initial conditions, random seed, or stochastic search path of the calibration algorithm. This is particularly relevant because the paper focuses on equifinality and differences among calibration configurations.

I do not suggest repeating the full experimental design for all catchments, as this may be computationally expensive. However, a limited robustness test could be useful. For example, the authors could repeat a subset of calibration experiments for one representative catchment, or one catchment from each hydrological regime, using several random seeds. Reporting the spread in final objective-function values, calibrated parameters, and key diagnostic variables such as ET, SM, LST, and fSCA would help determine whether the main conclusions are robust to calibration stochasticity.

We thank the reviewer for raising this point. We conducted exploratory analyses (section S.6 in the Supporting Information) to examine, for a given Q-based calibration metric, the interplay between equifinal parameter sets and the regularization scheme for simulating annual cycles of Q, ET', SM', LST, and fSCA (Figure S34). The results show that ET' seasonality can vary substantially among near-optimal parameter sets depending on the combination of regularization scheme and catchment (e.g., spatially constant parameters at Cautín; regularized Depth₂ at Colorado). These results suggest that modelers prioritizing internal process fidelity should verify the annual cycle of ET (and other relevant variables) using a small ensemble of behavioral parameter sets rather than the single best-fit solution. We have incorporated these points at the end of section 6.2 (Discussion).

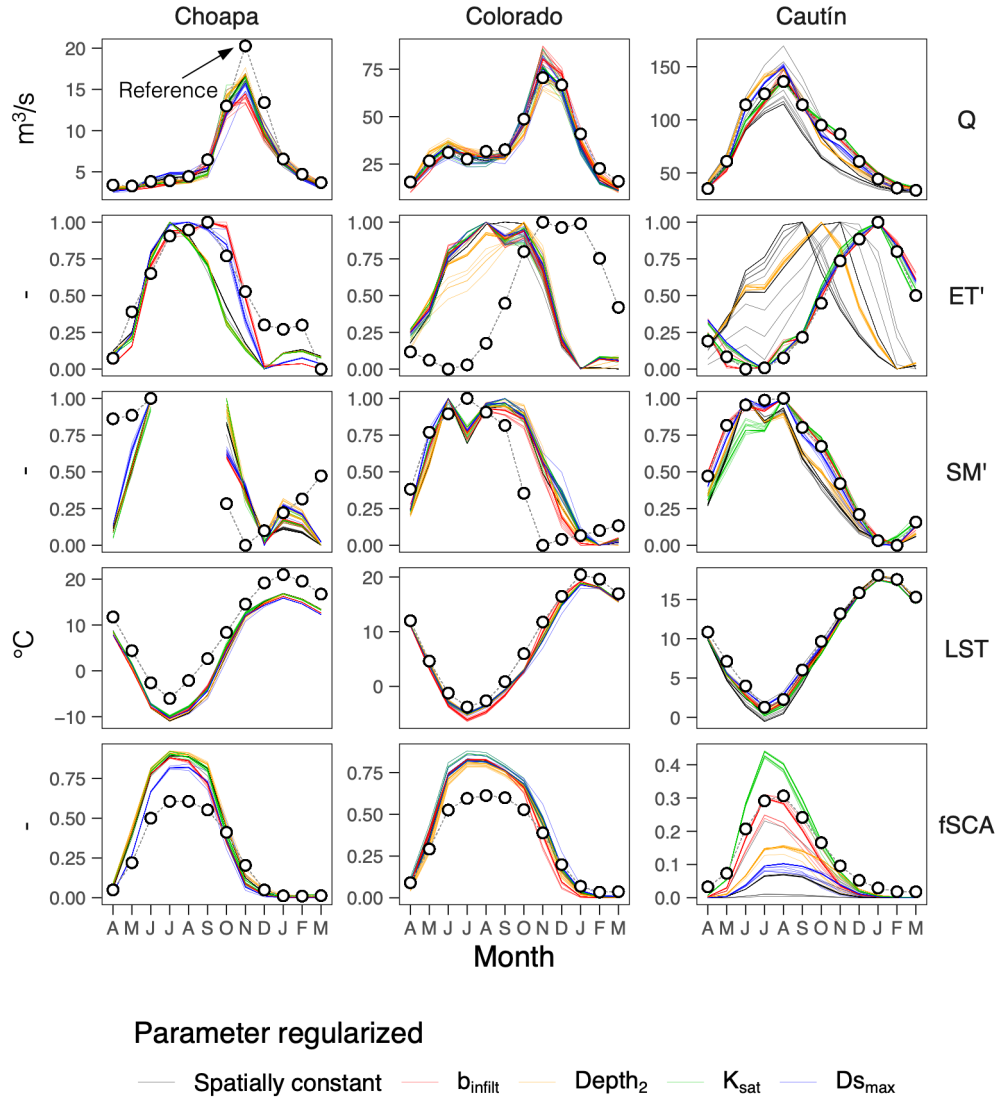


Figure R4. Catchment-scale annual cycles of streamflow (Q), normalized evapotranspiration (ET') normalized soil moisture (SM₁), land surface temperature (LST) and fractional snow covered area (fSCA) for the Choapa, Colorado, and Cautín River basins during the calibration period (2005–2018). Each line corresponds to a behavioral parameter set selected using two criteria: (1) the best 1% of parameter sets ranked by objective function value, or (2) all parameter sets with an objective function value exceeding $\max[OF(Q)] - 0.02$, where $OF = 0.5 \cdot KGE(Q) + 0.5 \cdot KGE(1/Q)$. Streamflow observations and remotely sensed variables are referred to as “reference” and represented with white dots. The normalization of ET and SM₁ is computed as $X' = (X - X_{min}) / (X_{max} - X_{min})$. Note that winter corresponds to JJA, while summer to DJF.

Additionally, we examined the sensitivity of our results and conclusions to the initial parameter set and seed number used in DDS optimizations for the following cases: (i) spatially uniform parameter sets; (ii) distributing only b_{infil} as a function of PC_1 (i.e., $b_{infil} = a + b \cdot PC_1^c$); and (iii) distributing only b_{infil} as a function of PC_1 and PC_2 (i.e., $b_{infil} = a + b \cdot PC_1^c + d \cdot PC_2^e$). We performed these calibration experiments only at the Futa River basin (i.e., the smallest one), to reduce the computational burden. The results show that, despite better $KGE(Q)$ values being achievable with different random seeds and including both PC_1 and PC_2 in the regularization scheme (Figure R5), none of the new parameter sets yield reasonable annual cycles of variables other than streamflow (Figure R6). We have incorporated these points in section 6.2 (Discussion).

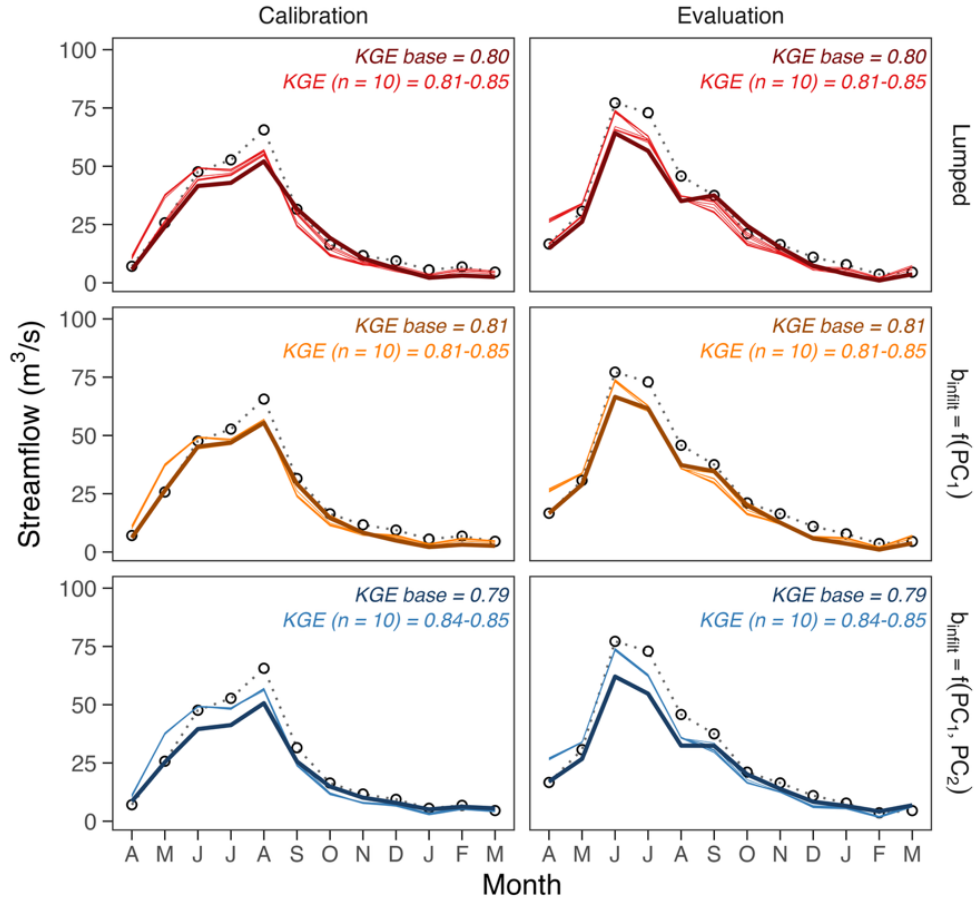


Figure R5: Streamflow annual cycles for additional calibrations conducted for the Futa River Basins to verify the impact of different initial parameter sets and seeds. In all the experiments, KGE(Q) was maximized using DDS (budget = 1,000 simulations), and results are presented for the following cases: (i) spatially uniform parameter sets; (ii) distributing only b_{infil} as a function of PC_1 (i.e., $b_{infil} = a + b \cdot PC_1^c$); and (iii) distributing only b_{infil} as a function of PC_1 and PC_2 (i.e., $b_{infil} = a + b \cdot PC_1^c + d \cdot PC_2^e$). The thicker lines represent the original results to facilitate comparisons, and the circles represent the mean monthly observational reference values.

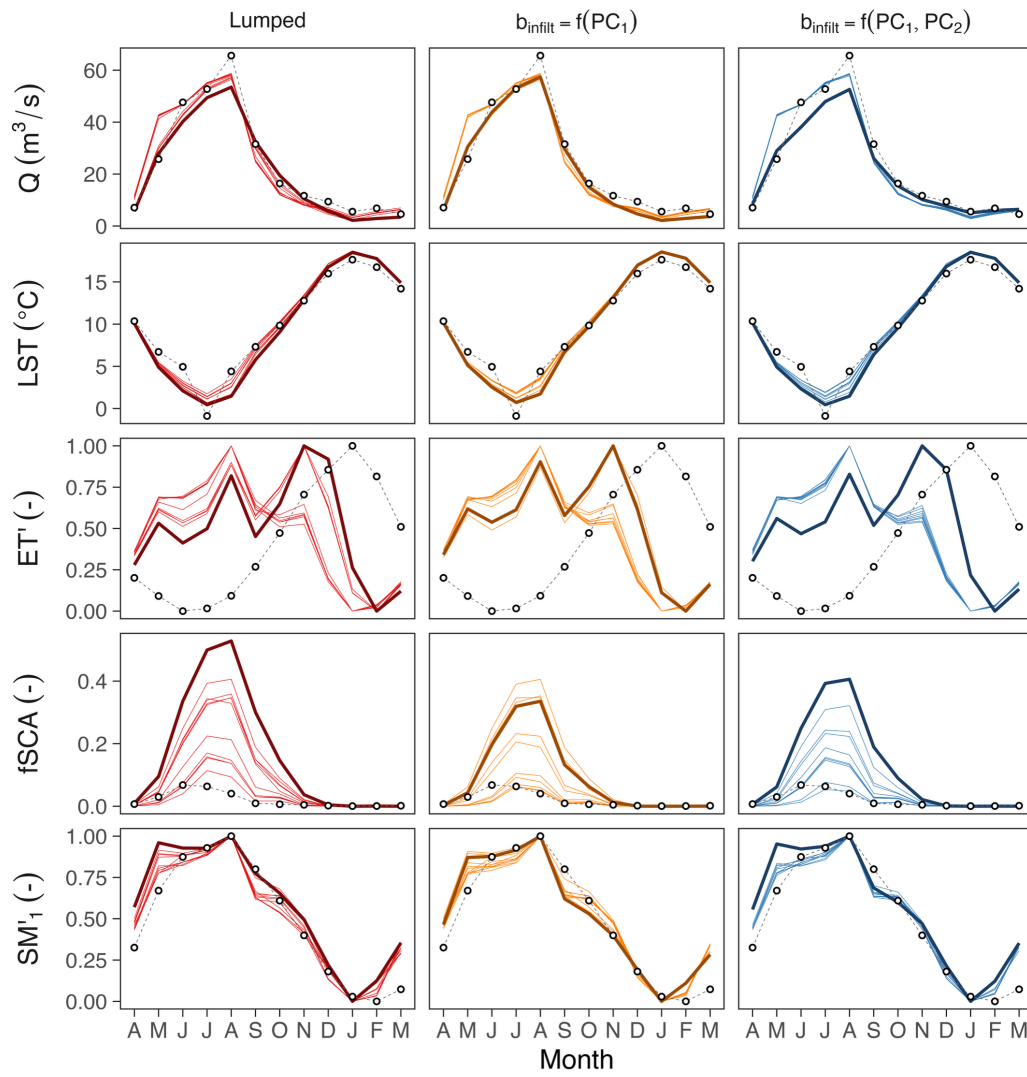


Figure R6: Simulated annual cycles of streamflow, land surface temperature (LST), dimensionless Evapotranspiration (ET'), fractional snow covered area (fSCA), and dimensionless Soil Moisture for the first (upper) soil layer (SM'_1) for parameter sets obtained by maximizing KGE(Q) with DDS (budget = 1,000 simulations) and different initial parameter sets and seeds. The thicker lines represent the original results to facilitate comparisons, and the circles represent the mean monthly observational reference values.

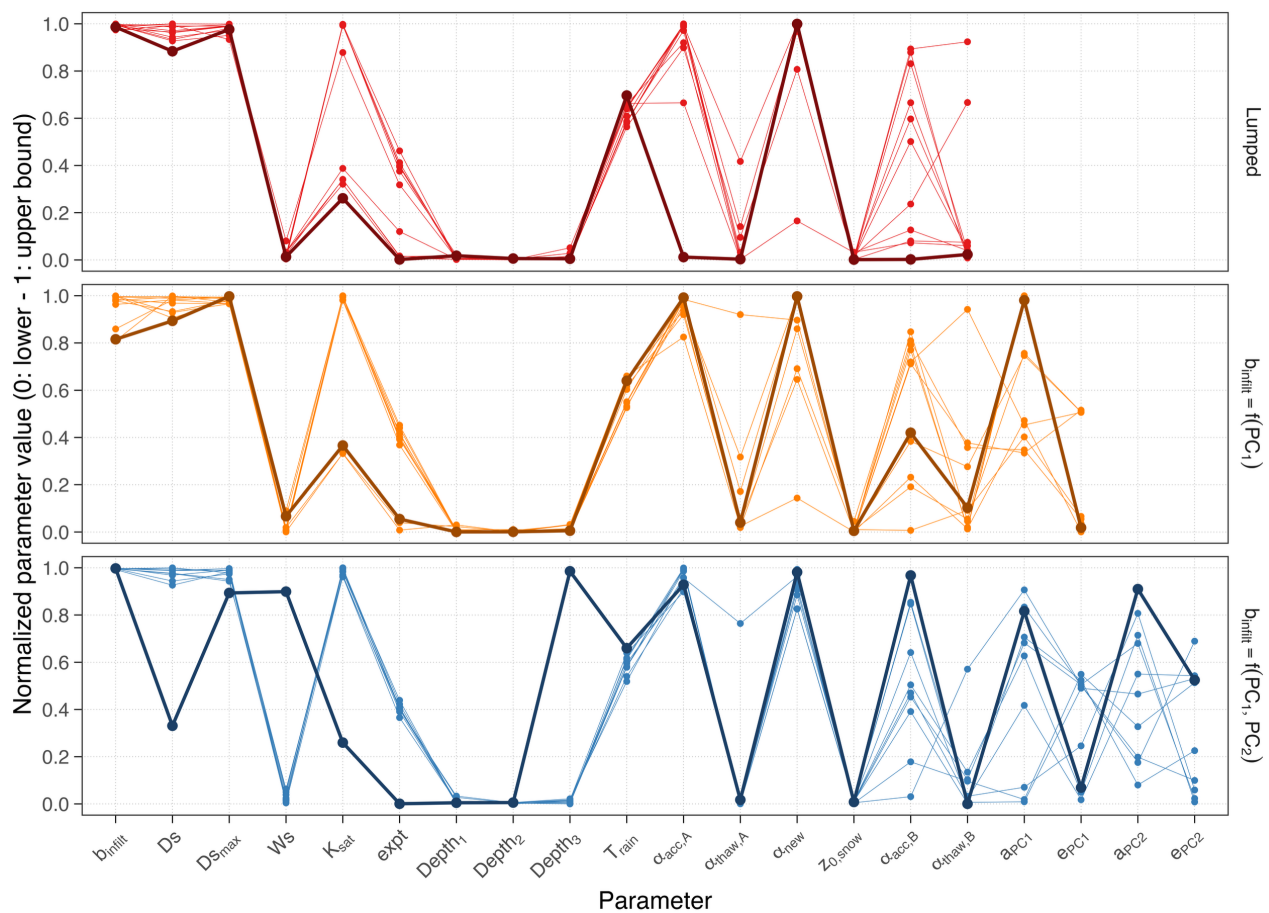


Figure R7: Normalized parameter values obtained by maximizing KGE(Q) with DDS (budget = 1,000 simulations) and different initial parameter sets and seeds. The thicker lines represent the original results to facilitate comparisons.

3. The manuscript evaluates model realism using gridded or remotely sensed products for ET, LST, fSCA, and SM. These products are valuable, but they also contain uncertainties related to retrieval algorithms, spatial resolution, temporal aggregation, and scale mismatch. Therefore, the interpretation of model performance against these products should be more cautious.

For example, when simulations differ from a gridded ET or SM product, this should not necessarily be interpreted as definitive model error. The authors may revise some statements to indicate that the simulations deviate from the selected reference product, rather than implying that the model is necessarily physically incorrect. A brief discussion of observational uncertainty and scale mismatch would improve the balance of the interpretation.

We agree with the reviewer that the gridded/remotely sensed products used to evaluate model realism (ET, LST, fSCA, and SM) carry their own uncertainties related to retrieval algorithms, horizontal resolution, temporal aggregation, and scale mismatch with our simulations. In the revised manuscript, we will revise the wording throughout sections to clarify that the different products are observational references. For example, we have reworded one of the last sentences in the introduction as follows: “To quantify the ability to replicate spatial patterns from reference SRS ET, LST, SM₁ and fSCA products, we used biased and unbiased performance metrics.”

We have rephrased the beginning of section 3.2 as follows:

“We use remotely sensed fractional snow-covered area, actual evapotranspiration, land surface temperature, and soil moisture as observational references for hydrological model evaluation (Figure 2).”

We have reworded the second paragraph of section 4.3 as:

“(We conduct) two model calibrations that consider spatially constant parameters and incorporate remotely sensed ET (Q+ET) and soil moisture in the first layer (Q+SM₁) as observational references within the optimization process.”

In section 4.3.2, we have re-worded some sentences as follows:

“We use the spatial pattern efficiency metric (Eq. 5) proposed by (Dembélé, Hrachowitz, et al., 2020) to quantify deviations between ET, SM, LST, and fSCA simulations, and the selected remotely sensed observational references”.

“To assess the model's capability to replicate remotely sensed annual cycles of ET and SM, we normalize ET and SM annual values as $X'=(X-X_{\min})/(X_{\max}-X_{\min})$.”

Regarding the uncertainties associated with SRS products, we have added the following text in the Discussion section:

“In this study, we used a single SRS product for ET, SM₁, LST, and fSCA. However, SRS products carry substantial uncertainties stemming from inputs, retrieval methods, horizontal resolution, temporal aggregation and algorithm-specific assumptions (McCabe et al., 2017). Moreover, different SRS products may yield considerable differences in the temporal dynamics and the spatial patterns used as observational references, affecting hydrological model calibration results (Dembélé, Ceperley, et al., 2020). Here, additional uncertainties were introduced when upscaling (i.e., ET, LST, and fSCA) or downscaling (i.e., SM₁) remotely sensed variables to match the 0.05° x 0.05° horizontal grid of VIC model simulations. Therefore, the model performance results presented here should be interpreted with caution, and they do not represent deviations from the "truth".”

We thank the reviewer for encouraging this more balanced interpretation, which we believe will strengthen the credibility of our conclusions.

Minor specific comments

1. Line 55: Please define “super-parameter” when it first appears.

We thank the reviewer for catching this. We will add a definition of the term “super-parameter” at its first appearance (Line 55), clarifying that it refers to global coefficients that link local model parameters with observable attributes.

2. Line 72: Please provide the full names of LST, ET, fSCA, and SM at first use.

We thank the reviewer for this observation. Please note that these variables are already defined in the abstract (L4-5) and the introduction (L32-34).

3. Introduction: The motivation for parameter regularization could be stated more explicitly. The authors could briefly explain the difficulties of calibrating many spatially distributed parameters, including overparameterization, equifinality, limited parameter identifiability, and computational cost. This would better motivate the need for regularization.

We have added the following text in the introduction:

“On the other hand, the implementation of distributed hydrological models requires the specification of a much larger number of parameter values in comparison to the lumped counterpart. Hence, the parameter optimization process becomes an ill-posed problem, involving a high-dimensional parameter space search that increases the computational cost considerably. This problem is typically addressed through regularization (P. Pokhrel & Gupta, 2010; Samaniego et al., 2010; de Lavenne et al., 2019).”

Further, we will clarify in the discussion section that spatial parameter regularization strategies are not intended to improve parameter identifiability and/or mitigate equifinality; instead, they seek to reduce the dimensionality of the parameter estimation problem in distributed hydrological modeling (Pokhrel et al., 2008; Prafulla Pokhrel & Gupta, 2010; de Lavenne et al., 2019).

We hope our responses satisfactorily address the reviewer's concerns, and we thank them again for their valuable and constructive feedback, which we believe will substantially improve the quality of the manuscript.

References

- Crespo, P. J., Feyen, J., Buytaert, W., Bücken, A., Breuer, L., Frede, H. G., & Ramírez, M. (2011). Identifying controls of the rainfall-runoff response of small catchments in the tropical Andes (Ecuador). *Journal of Hydrology*, 407(1–4), 164–174. <https://doi.org/10.1016/j.jhydrol.2011.07.021>
- Dembélé, M., Hrachowitz, M., Savenije, H. H. G., Mariéthoz, G., & Schaefli, B. (2020). Improving the Predictive Skill of a Distributed Hydrological Model by Calibration on Spatial Patterns With Multiple Satellite Data Sets. *Water Resources Research*, 56(1), 1–26. <https://doi.org/10.1029/2019WR026085>
- Dembélé, M., Ceperley, N., Zwart, S. J., Salvatore, E., Mariéthoz, G., & Schaefli, B. (2020). Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Advances in Water Resources*, 143. <https://doi.org/10.1016/j.advwatres.2020.103667>
- Gordillo, J., & Pineda, L. E. (2021). Unravelling runoff processes in Andean basins in northern Ecuador through hydrological signatures. *Hydrological Processes*, 35(9). <https://doi.org/10.1002/hyp.14354>
- de Lavenne, A., Andréassian, V., Thirel, G., Ramos, M. -H., & Perrin, C. (2019). A Regularization Approach to Improve the Sequential Calibration of a Semidistributed Hydrological Model. *Water Resources Research*, 55(11), 8821–8839. <https://doi.org/10.1029/2018WR024266>
- McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., et al. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences*, 21(7), 3879–3914. <https://doi.org/10.5194/hess-21-3879-2017>
- Molina, A., Govers, G., Vanacker, V., Poesen, J., Zeelmaekers, E., & Cisneros, F. (2007). Runoff generation in a degraded Andean ecosystem: Interaction of vegetation cover and land use. *Catena*, 71(2), 357–370. <https://doi.org/10.1016/j.catena.2007.04.002>
- Pokhrel, P., & Gupta, H. V. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Research*, 46(1), W01505. <https://doi.org/10.1029/2009WR008066>
- Roa-García, M. C., Brown, S., Schreier, H., & Lavkulich, L. M. (2011). The role of land use and soils in regulating water flow in small headwater catchments of the Andes. *Water Resources Research*, 47(5). <https://doi.org/10.1029/2010WR009582>
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), W05523. <https://doi.org/10.1029/2008WR007327>