I apologize for the time it has taken me to re-review this paper. I recommend that the paper be accepted for publication pending minor revision. While I still have some concerns about the work, the review process seems to have reached a natural stopping point. Improvements have been made in the 2 rounds of review and the authors are likely ready to move on to other things.

> *Our Response:*
> We thank the reviewer's suggestions. We address the following issues one-by-one. It is true that the authors are preparing another paper, and will extend the discussions such as comment #1 in the following.

Remaining concerns:

1) I am confused as to why Ks (the fluvial erodibility coefficient) does not influence soil thickness as stated in the reply to reviewer file. It seems as though it must if the soil thicknesses have been measured in portions of the landscape with significant fluvial erosion (if not then there is no point in having that component of the model present). Related: the results of the surface water routing model seem unrealistic (the maximum flow depth, even in channels, is 1 mm).

> *Our Response:*
> We thank the reviewer's comment. In our sensitivity analysis, the parameter Ks is less significant compared to hillslope diffusion coefficient. This is expected qualitatively but not quantitatively before we perform the analysis. We conduct the overland flow simulation using annual precipitation rate of 363 mm/yr as stated in the caption. Among the two hillslopes, there is no stream channels. There are stream pathways but these have only been submerged for a very short period, with even no surface water at all during dry years. We are preparing another manuscript to address in more details the relationship between surface flow water depth and precipitation.

2) The description of the Bayesian approach to calibration was very helpful but also raised some questions. For example, the range of prior values for h_0 is 0.15-0.5 m but the plots only include a subset of this range. Another example: the model RMSE seems to be essentially independent of E_thre (i.e., the pdf is not humped but largely uniform across the range of prior values). This is unfortunate and somewhat troubling since the main point of the proposed method is that it is a good idea to hand off from one model to another depending on the value of E_thre. If the results are independent of E_thre, it would seem logical to conclude that this study site is not a good test of the hybrid approach, no? I am also confused by the units of E_thre in the author response file (the values are negative and range from about -3 to -1 x 10^-6 – very different from the range in Table 1).

> *Our Response:*
> We thank the reviewer's reminder. The range listed in the table has been updated to be consistent with the range in the Bayesian approach. The results show relatively high nonlinearity in the sensitivity analysis (Figure 3). We believe that the 'pdf' plot that the reviewer mentions here is the histogram plot in Figure S7. In the histogram plots for $E_{thre}$, an optimal (low RMSE) value shows an obviously better value which gives smallest RESE than the rest of the values, even though the rest of the values show a relatively

uniform RESEs. In Figure S8, $E_{thre}$ shows a clearly better value at the north-facing hillslope, but it is more uniform in the south-facing hillslope. The possible reason would be that south-facing hillslope has less sampling points than the north-facing hillslope. More sampling points at the south-facing hillslope may improve the performance of $E_{thre}$. Regarding the sign, it is the author's mistake in that the negative sign should have been positive in the plots, which are all revised. We include the following sentences in the revised manuscript: "Moreover, among the posterior distribution of each parameter, $P_o$ and $E_{theta}$ show closer to uniform in the south-facing hillslope than the north-facing hillslope. The reason may be that the north-facing hillslope has more sampling points, which provides better estimation than the south-facing hillslope."

3) I would have liked to have seen more information on why the authors believe that the global parameter search yielded a best-fit solution and not some local optimum. I am most familiar with using Markov-Chain Monte Carlo for searching parameter spaces. Whether or not one finds a global or local optimum using MCMC depends sensitively on the details of how the set of parameter values is perturbed from one iteration to the next. Need more info to be convinced.

*Our Response:*

We agree that MCMC is one of the common methods as a Bayesian method, and it is known to be efficient. However, the first set of the reviewer's comments asked about the global optimum and parameter interaction, so that the exhaustive parameter combination over the range would be more beneficial to address these comments. This sampling-resampling method (Smith and Gelfand 1992) is also a valid and common method for Bayesian inference. Compared to an iterative method such as MCMC, this method is advantageous since it can parallelize simulations over available cores and it is robust against simulation failures. Moreover, we used grid search and Bayesian method calculate the posterior distribution of the parameters, aiming at finding the best set of parameters that fits the minimum RMSE (Figure S8). In Figure S7, we show the grid search of parameters. We plot the histograms of parameter values, which corresponds to RMSEs smaller than a threshold. By keeping reducing the threshold, we can see that the mostly selected parameters are consistent, which guarantees the global optimum instead of local optimum.

Reference:

A. F. M. Smith and A. E. Gelfand, Bayesian statistics without tears: a sampling-resampling perspective, Am. Statistician, 46 (1992), pp. 84–88.

Some typos remain (e.g., parament instead of parameter on line 57).

*Our Response:*

We thank the reviewer's reminder. It has been corrected.