Responses to the comments of RC1:

My main problem with this manuscript is that the authors have misrepresented CLM5 as having a constant soil depth of 8.03 m. This is not true. Variable soil thickness has been implemented into this newest version of the model (see the CLM Technical Note at https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/tech_note/index.html). With this implementation, the number of hydrologically active layers varies from grid cell to grid cell.

Response: This part of the model description in the previous manuscript was inaccurate, which was based on the previous versions of CLM. We have corrected the issue throughout the revised manuscript. We have treated the constant soil depth of 8.03 m as a default option as described in Section 2.2.2.1 in Lawrence et al. (2018) and examined how the soil depth changes affect the runoff simulations through the sensitivity tests.

Lawrence, D. et al.: Technical Description of version 5.0 of the Community Land Model (CLM5), National Center for Atmospheric Research, Boulder, Colorado, 2018.

Therefore, what the authors are investigating here is not the addition of variable soil thickness but the impact of vertical resolution with variable soil thickness. They need to make this crystal clear in Section 5.1. Also, are the bottom of the "soil" columns all at the same depth in each of the SLN sensitivity tests? If so, the authors should state that. If not, the authors should note what the bottom depth in each test simulation is.

Response: We have revised Section 5.1 based on the suggestion from this reviewer. Yes, the bottom of the soil columns is at different depths across the watershed in each of the SLN sensitivity tests except for the default run with the fixed soil depth of 8.03 m (see Section 2.2.2.1 in Lawrence et al. 2018). The spatial distribution of the soil depth is shown in Figure S1, and was also added to the manuscript as Figure 1b.

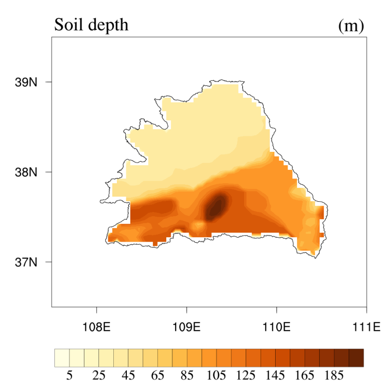

Figure S1. The geographic distribution of the soil depth for the WRB.

The goal of this work is to improve the runoff simulated in this region of interest, the Wuding River Basin. It is clear that the increase in vertical resolution improves the simulations, but the runoff is still biased. I am pleased to see the improvements made by adding the realistic river network and the changes made to improve the evapotranspiration.

Response: Thanks for the positive comments.

Of minor note, at Line 134, the authors introduce the acronym WRB without earlier definition in the body of the text. They do define this in the abstract, but they need to define it in the body of the manuscript.

Response: Thanks! This has been corrected in the revised manuscript.

Finally, shouldn't the Nash-Sutcliffe efficiency be between 0 and 1 in magnitude?

Response: The Nash-Sutcliffe efficiencies can range from -infinity to 1.0. Please see the following reference:

Nash, J. E. and Sutcliffe, J. V., River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10 (3), 282-290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Responses to the comments of RC2:

Jin et al. present a sensitivity analysis of improving runoff simulation using CLM5 for loess plateau watershed. They evaluate the parameter parameters and significantly improve the simulation performance. I feel this paper is well written with scientific insights. However, I have some serious concerns and suggestions and hope the authors can clarify/answer.

L24: What is the depth of the 150 soil layers?

Response: The geographic distribution of the soil depth in the Wuding River Basin is shown in Figure S1 for our responses to the comments of RC1, and this figure was also added to the manuscript as Figure 1b. The soil depth data were from observations, and were used for the 150 soil layer model simulations.

L27: What's the "higher-resolution"? I suggest adding them explicitly in the abstract.

Response: We removed this vague phrase and revised the statement as follows in the abstract:

“In addition, when compared with the default version with 20 soil layers, CLM5 with 150 soil layers slightly improved runoff simulations.”

L71: It’s surprising that the authors didn’t mention the seasonal and/or extreme precipitation impacts on the surface and subsurface runoff, as well as other physical processes such as erosion in the LP geology.

Response: We added a statement here to discuss the issue as follows:

“Especially, extreme rainfall events that mostly occur during the summer monsoon season (Tian et al. 2020) produce strong soil erosion and a large amount of fast infiltration-excess surface runoff to the river channels in hillslope areas, sometimes causing severe flooding.”

However, in this study, we focused mostly on the effects of the model structures on the runoff simulations, which is the main objective of this study. The effects of the forcing data on the runoff simulations and soil erosive processes are beyond the scope of this study. Therefore, we did not conduct an analysis related to extreme precipitation and soil erosion. However, based on our final results, we can see that the model relatively well captured the magnitude and phase of the observed runoff peaks (Figure 8) during which extreme precipitation events could often occur. This implyed that CLM may be able to handle the impact of extreme precipitation on the runoff in our study region when such events occur. In addition, we should indicate that CLM has no capacity to simulate soil erosion for this region.

L92: Another benefit of a hydrological model can be that models output the quantity of different components of water budget (subsurface and/or surface runoff) that are difficult or impossible to be measured directly.

Response: Thanks. We adopted this reviewer’s suggestion by adding a similar statement in the text.

L134: The authors choose one of the largest loess area watershed, but did you address the human activities and their impacts on hydrological analysis? My understanding is that CLM does not explicitly address those activities that makes the evaluation of model difficult in the large scale watershed.

Response: Yes, the impact of the human’s activities on the runoff is always important in this region, but it is quite a challenge to incorporate those activities into the physical models such as CLM. Thus, to avoid human’s factors in our runoff simulations, we intentionally selected a study period of 1956-1969 when human activities were minimal (Jiao et al., 2017). We have also explained this in Section 6.1.

Jiao, Y., Lei, H., Yang, D., Huang, M., Liu, D., and Yuan, X.: Impact of vegetation dynamics on hydrological processes in a semi-arid basin by using a land surface-hydrology coupled model, J. Hydrol., 551, 116-131, https://doi.org/10.1016/j.jhydrol.2017.05.060, 2017.

L228: If I understand the authors correctly, this study only evaluates the sensitivity of layer thickness and/or the number of layers, etc. Would the soil properties more important, such as water capacity, or other soil parameters in CLM? Did you use the default options? What about the runoff generation parameters (there are a lot in CLM)?

Response: As we have mentioned above, this study focused mostly on examining the model structures in simulating the runoff. After we adjusted the deficient model structures to more closely reflect the reality, the model can reasonably simulate the runoff, implying that the model parameters related to the runoff simulations should be within the reasonable range except for the soil evaporation parameter as discussed in Section 6.5. Without tuning those parameters, our final results show that the model can realistically capture the observed runoff.

L233: Did you use the same spin-up for all sensitivity analysis cases, or each sensitivity case uses their own spin-up?

Response: Each sensitivity case has its own spin-up, and we also clarified the issue in Section 5.2.

L244: A few details are missing here for how the simulated surface and subsurface runoff are compared to the streamflow gauge observation at the outlet of the watershed. Did the authors use the runoff at the grid cell of the watershed outlet for comparison, or use the total flux over the watershed for the comparison? If the latter option is used, the watershed is relatively large so did you expect any delay in hydrological response time?

Response: It is a good question. During the early stage of this study, we compared the runoff simulations with and without river routing at a monthly scale, but did not see a significant difference (Figures not shown). In fact, for our study watershed that can easily fit into a 200 km by 200 km box, a month time mostly should be sufficient for the river flow to come out to the outlet from the farthest point in this watershed with elevation ranging from nearly 2,000 m to about 500 m. In addition, we can see in Figure 8 that there are no systematic phase shifts between the runoff observations and simulations, further indicating that a river routing may not be necessary. Thus, we used the simulated total flux over the watershed for the comparison with observations in this study and explained this in Section 4.

L249: Also shown in Figure 2 and 3, the total water budget or accumulated runoff are way more greater than the observation. I think it will be helpful to plot their ratio against the total precipitation.

Response: We have plotted the ratio of both observed and simulated runoff to observed precipitation and found that the ratio over the summer rainy season is less than 1 (Figure S2). However, the ratio is much larger than 1 during the winter and spring dry seasons. This is because the runoff comes mostly from the groundwater during the dry period when precipitation was very weak. During this period, the magnitude of runoff is often greater than that of precipitation. Figure S2 shows that the simulated runoff to precipitation ratio is larger than the observed value, resulting from the underestimated evaporation as discussed in Section 6.5.

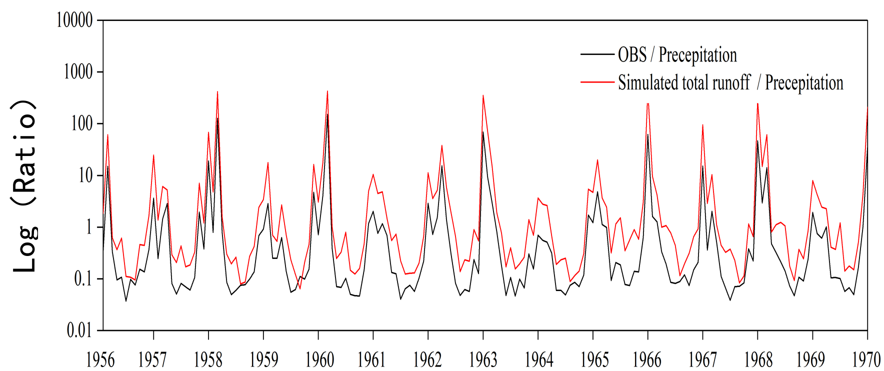

Figure S2. The ratios of the observed (dark line) and simulated (red line) runoff (default model) to observed precipitation for the WRB over the period of 1956-1969 based on the default options of CLM.

L298: With more layers, the RMSE still looks great or the simulated total runoff still does fit well to observation. Maybe the number of layers is not a significant variable? Should you examine other parameters or model setup?

Response: In fact, the model with more soil layers did not give great benefits to improve the runoff simulations. The key factors for the improvements of the simulated runoff are the addition of the river channels and the adjustment of the soil evaporation parameter. However, even though the model with more soil layers did not generate much better runoff simulations, the simulated vertical soil moisture showed much smoother patterns than that with less soil layers (Figure 5 in the revised manuscript). Due to the homogenous textures of the loess soil, these smoother patterns may be more reasonable when compared with those with less soil layers. We also will need to further verify the soil moisture profile simulations when observations are available.

L323: Again the NSE value of all of those results are smaller than the default option of CLM. I'm still confused about how this could happen (why the results after changing/improving conditions and model setup) are still worse than the first trial? It seems simulated total runoff are way higher than the observation.

Response: After we improved the model structures (model layering and river channels), the NSE actually increased slightly from -12 to about -10. However, the R-squared dramatically increased from 0.02 to 0.52 with a P_{gr} of 0.15, resulting from the improvement in the simulated runoff variability. Due to the underestimated evaporation, the simulated total runoff was still way higher than observations, which was discussed and fixed in Section 6.5.

L376: Now I understand that the authors adjust ET parameters to improve the runoff results. But this still doesn't explain why the default options of CLM actually simulate the runoff relatively well? Is the simulated ET also underestimated with the default options of CLM?

Response: In fact, the default options of CLM had a poor performance in simulating the runoff for our study watershed with an R-squared of 0.02 and an NSE of -12.34. We can see that both phases and magnitudes of the simulated runoff did not match with observations. The addition of the river channels to CLM significantly improved the runoff variability simulations, and the adjustment of the soil evaporation parameter improved the runoff magnitude simulations, which are the key findings in this study. These improvements drastically increased the R-squared to 0.62 and the NSE to 0.61.

Yes, the default options of CLM significantly underestimated ET as discussed in Section 6.5.

L394: If I understand the paper correctly, the authors adjusted the Pgu values to represent the number of river channels in the watershed, but do not explicitly use river network in their model. I feel this sentence can be misleading by letting readers assume they use the river routing model.

Response: The detailed definition of the river channels in this study are given in Section 3.2 where the soil depths were generated through the elevation comparison between the 90 and 5 km resolution DEMs. However, the river network is generated through the river routing module in CLM5 based only on one DEM. We added a simple statement to Section 3.2 as follows to clarify the issue:

"This is different from river routing that is based only on one DEM."