Response to RC2

We sincerely thank the Referee #2 for their constructive evaluation and for recognizing the scientific merit, novelty and potential value of our map. We deeply appreciate your rigorous critique of our initial uncertainty quantification and validation strategy. In response, we have implemented major revisions, most notably reframing our validation claims to explicitly acknowledge the shared-data limitations, adding new spatial pattern analyses, and introducing quantitative global uncertainty analysis using Latin Hypercube Sampling.

Below, we provide a detailed, point-by-point response to your comments. Your original comments are presented in **blue**, our responses are given in **black**, and the corresponding revisions added to the manuscript are provided in **red.**

This manuscript presents a new 1 km resolution solum thickness dataset for the Qinghai-Tibet Plateau (QTP) developed using a revised geomorphic mass balance model. The topic is highly relevant and timely. The methodological innovations represent a thoughtful and non-routine adaptation of classical mass balance models to the QTP's unique conditions. This work has strong potential value for the soil science, cryospheric, and Earth system modeling communities.

Overall, the scientific core has clear merit and novelty. However, several non-trivial weaknesses in validation strategy and uncertainty quantification prevent acceptance in the current form. With careful attention to the major and minor points below, a revised version has a good chance of acceptance in ESSD.

We thank the reviewer for this highly encouraging assessment of our manuscript's scientific significance. We fully acknowledge your concerns regarding the initial validation strategy and uncertainty quantification. As detailed in our point-by-point responses below, we have dedicated our efforts to directly resolving these two issues. We believe these major revisions have fundamentally strengthened the manuscript.

Major Comments

1 The performance assessment relies on 4-fold cross-validation using the same 552 soil profiles that overlap significantly with the data sources used to construct the benchmark maps (Shangguan and Liu). Therefore, the claims of outperforming existing products by 10–17% in RMSE and MRE are weakened by the lack of truly independent test data. Please provide a clearer acknowledgment of this shared-data limitation. Additionally, spatial pattern analysis (e.g., correlation of modeled thickness with topographic

curvature, topographic position index, or slope) might be helpful in enhancing assessment.

We thank the reviewer for this critical observation. We strongly agree that this shared-data limitation must be explicitly addressed.

To clarify the nature of the data structure: as detailed in Section 2.2, our 552 soil profiles are an integrated compilation of two historical archives: the Second National Soil Survey (1979–1985) and the National Soil Series Survey (2009–2019), and supplemented by additional independent regional samples. Because our reported performance metrics were generated using a 4-fold cross-validation on this entire compiled dataset, the test folds inevitably contained data shared with both the Shangguan map (which relied on the 1979–1985 data) and the Liu map (which relied on the 2009–2019 data).

While our dataset does contain a substantial number of independent samples relative to each individual baseline map, the reviewer is strictly correct that the final cross-validated metrics themselves lack true independence. However, compiling a massive, 100% independent dataset is currently impossible due to limited soil profile available on the extreme physically inaccessible QTP. Therefore, all large-scale mapping efforts have to utilize the same limited legacy archives.

To comprehensively address this limitation, we have implemented a two-part revision:

1 Acknowledging the share-data limitation. We have revised the Section 4.3 (Limitations and Future Directions) and the discussion about Table 5 to explicitly acknowledge this constraint and temper our performance claims. We now strictly frame the statistical comparison as a relative assessment under shared observational constraint rather than definitive proof of generalization to untested regions. In Section 4.3, we added:

"The performance of the model relies on a 4-fold cross-validation using 552 soil profiles. Consequently, our compiled dataset, drawn primarily from the Second National Soil Survey and National Soil Series Survey, overlaps significantly with the data sources used to construct the comparison maps (Shangguan and Liu maps). Because the test folds are not fully independent of these legacy datasets, the reported improvements in RMSE and MRE must be interpreted with caution. This comparison serves as a relative assessment under shared observational constraints, rather than definitive proof of generalization to untested regions."

A similar acknowledgment has been added to the Table 5 discussion.

2 Deepening the spatial pattern analysis. We now explicitly use the topographic correlation (Slope Position r = 0.25, TRI r = 0.24, and Slope r = 0.25 from Figure 4)

and the Hengduan Mountains transect (Figure 5) as additional spatial validation to verify that the model correctly reproduces physically plausible, terrain-driven differentiation at the hillslope scale. Key revision in Section 3.3:

"...These correlations verify that terrain attributes successfully drive spatial differentiation within our model framework. [...] As demonstrated in the transect analysis across the Hengduan Mountains (Fig. 5), the simulated thickness varies systematically with terrain position: steep ridges act as sediment-source areas where active erosion restricts soil accumulation, whereas gentler slopes and valley bottoms function as depositional zones."

2 The steady-state assumption is also a critical simplification on a tectonically active and permafrost-degrading plateau. While limitations are discussed, there is no quantitative propagation of uncertainty into the final dataset. I suggest the authors can add some sensitivity analysis. For example, vary key parameters ±20–50% or using Monte Carlo sampling for weighting coefficients and present spatial uncertainty estimates, then discuss how uncertainty varies across clusters or geomorphic zones.

To comprehensively quantify the uncertainty, we applied parameter perturbations to the distinct sources of uncertainty in our framework.

1 Spatial sensitivity analysis (± 20% perturbation). First, to address the sensitivity of our internal model parameters, we conducted a spatially explicit One-At-A-Time (OAT) sensitivity analysis. Based on the calibrated baseline map, we applied a relative perturbation of ±20% to the key erosion weighting coefficients (α, β, γ). We applied ±20% perturbations in accordance with the reviewer's suggestion to vary key parameters by ±20–50%. This range is appropriate given that the calibrated coefficients are already small in most clusters.

As demonstrated in our newly added spatial map and summary statistics (Fig. R1 and R2 / Table R1), the parameter sensitivity is strongly partitioned by regional geomorphic regimes:

- In the extreme relief of the Hengduan Mountains (Cluster 1), the model exhibits the highest elasticity to gravitational weights (γ).
- In the kinetically limited interior permafrost zones (Cluster 7), sensitivity is predominantly driven by aeolian weights (β).
- However, the overall plateau-wide elasticities remain robustly low (α = -0.01, β = -0.02, γ = -0.08), indicating that the model's internal structure is physically stable and not overly sensitive to these mathematical balancing coefficients. The low sensitivities are partly attributable to the calibration assigning small weighting

coefficients in most clusters, which effectively down-weights the empirical erosion inputs.

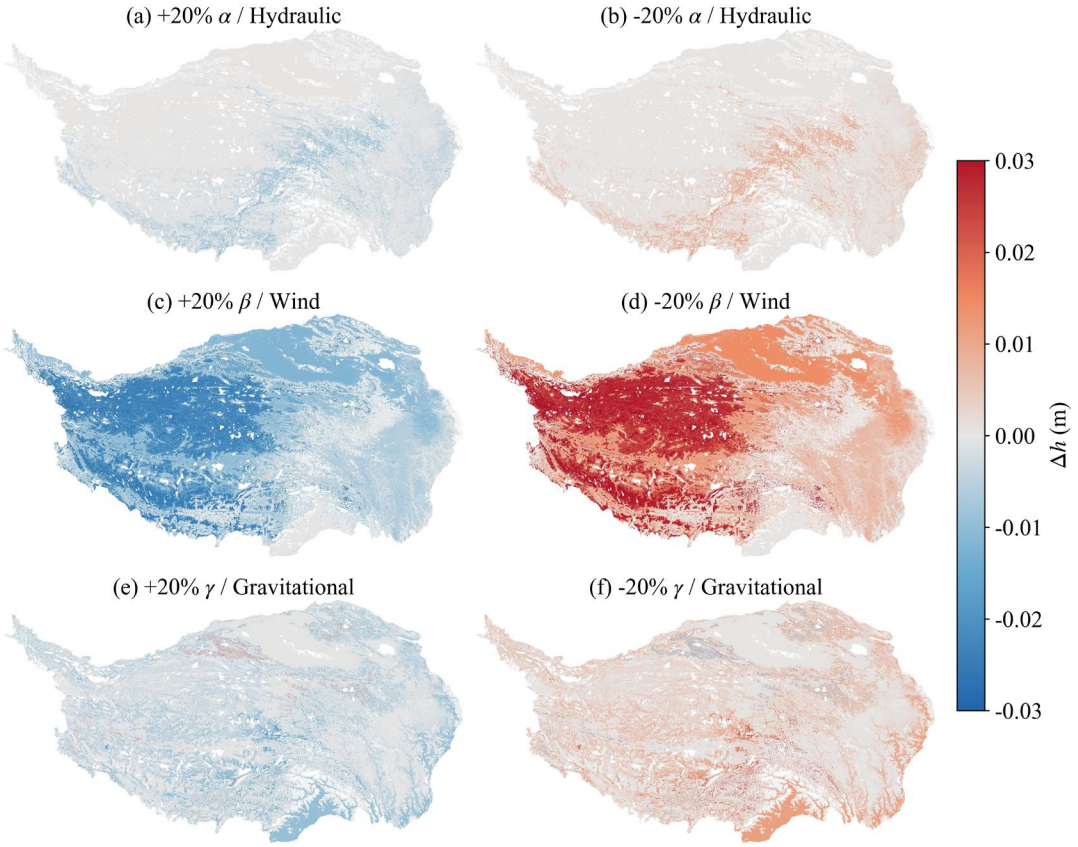


Figure R1. Spatial distribution of solum thickness sensitivity to ±20% perturbations in erosion weighting coefficients. The legend shows 98% range of values.

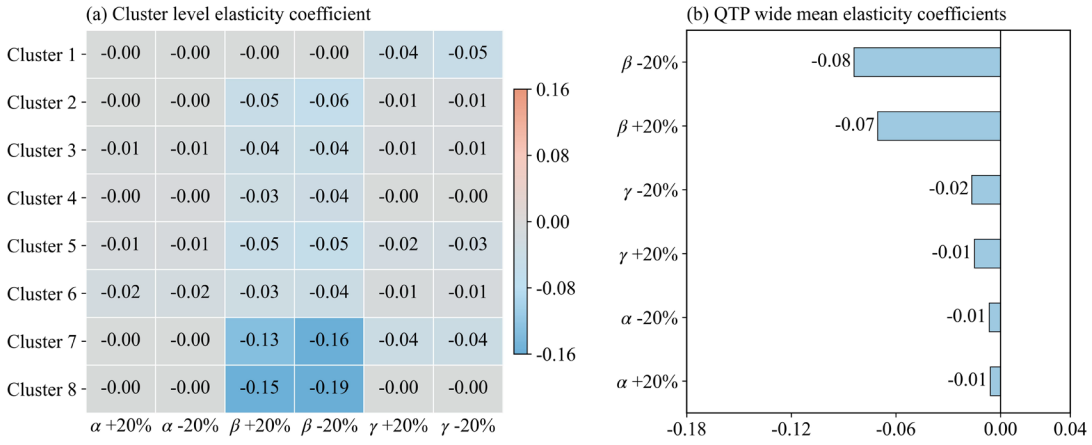


Figure R2. Elasticity coefficients of simulated solum thickness. (a) Cluster-level elasticity coefficients showing the relative sensitivity of solum thickness to ±20%

perturbations of α, β, and γ for each environmental cluster. (b) QTP-wide mean elasticity coefficients summarizing the overall sensitivity of solum thickness across the plateau. Elasticity is negative for both +20% and –20% perturbations because solum thickness responds inversely to erosion rate (Eq. 9).

Table R1. Summary statistics of the OAT sensitivity analysis, detailing mean absolute and relative changes, mean elasticity and area with >10% change per cluster and across the entire QTP.

Parameter	Scenario	Cluster	Mean Relative Change (%)	Mean Elasticity	Area with >10% Change (%)
α	-20%	1	0.00	0.00	0
	-20%	2	0.03	0.00	0
	-20%	3	0.11	-0.01	0
	-20%	4	0.08	0.00	0
	-20%	5	0.19	-0.01	0.14
	-20%	6	0.37	-0.02	0
	-20%	7	0.03	0.00	0
	-20%	8	0.07	0.00	0
	-20%	Plateau wide	0.13	-0.01	0.01
α	+20%	1	0.00	0.00	0
	+20%	2	-0.03	0.00	0
	+20%	3	-0.10	-0.01	0
	+20%	4	-0.08	0.00	0
	+20%	5	-0.19	-0.01	0.14
	+20%	6	-0.32	-0.02	0
	+20%	7	-0.03	0.00	0
	+20%	8	-0.06	0.00	0
	+20%	Plateau wide	-0.12	-0.01	0.01
β	-20%	1	0.00	0.00	0
	-20%	2	1.16	-0.06	0
	-20%	3	0.89	-0.04	0
	-20%	4	0.71	-0.04	0
	-20%	5	1.08	-0.05	0.90

	-20%	6	0.74	-0.04	0
	-20%	7	3.12	-0.16	0
	-20%	8	3.73	-0.19	0
	-20%	Plateau wide	1.68	-0.08	0.07
β	+20%	1	0.00	0.00	0
	+20%	2	-0.96	-0.05	0
	+20%	3	-0.75	-0.04	0
	+20%	4	-0.59	-0.03	0
	+20%	5	-0.95	-0.05	0.61
	+20%	6	-0.63	-0.03	0
	+20%	7	-2.63	-0.13	0
	+20%	8	-3.06	-0.15	0
	+20%	Plateau wide	-1.41	-0.07	0.05
γ	-20%	1	0.92	-0.05	0
	-20%	2	0.14	-0.01	0
	-20%	3	0.21	-0.01	0
	-20%	4	0.04	0.00	0
	-20%	5	0.54	-0.03	0.72
	-20%	6	0.13	-0.01	0
	-20%	7	0.77	-0.04	0
	-20%	8	0.00	0.00	0
	-20%	Plateau wide	0.33	-0.02	0.06
γ	+20%	1	-0.76	-0.04	0
	+20%	2	-0.12	-0.01	0
	+20%	3	-0.19	-0.01	0
	+20%	4	-0.04	0.00	0
	+20%	5	-0.46	-0.02	0.68
	+20%	6	-0.12	-0.01	0
	+20%	7	-0.72	-0.04	0
	+20%	8	0.00	0.00	0
	+20%	Plateau wide	-0.30	-0.01	0.05

2 Uncertainty propagation and sensitivity analysis. We added an uncertainty propagation and sensitivity analysis to explicitly assess the effects of uncertainty in the empirical erosion inputs. To quantify uncertainty propagation from the external empirical erosion inputs, we independently perturbed E_{water} and E_{wind}, derived from RUSLE and RWEQ, respectively, within ±40% of their baseline values using Latin Hypercube Sampling (LHS). This ±40% range is justified by the known limitations of these temperate-derived empirical models when extrapolated to high-altitude permafrost environments, where uncertainties are typically reported in the 30–60% range.

This generated 150 ensemble simulations, from which we characterized the spatial uncertainty in simulated solum thickness using the coefficient of variation (CV), standard deviation, the 5th–95th percentile range, and the relative difference between the ensemble mean and the baseline simulation (Fig. R3a–d). To further identify the dominant sources of uncertainty, we performed a separate Sobol total-order sensitivity analysis. The required samples were generated using the Saltelli sampling scheme, with E_{water} and E_{wind} sampled within the same ±40% ranges. A total of 512 model evaluations were used to calculate the Sobol total-order indices, which quantify the overall contributions of E_{water} and E_{wind} to output uncertainty, including both their direct and interaction effects (Fig. R3e, f).

This CV map explicitly quantifies how the uncertainty of using these erosion models in permafrost-dominated zones propagates into our final solum thickness estimates. Notably, the uncertainty is highest in the plateau's permafrost core, directly highlighting where aeolian and cryogenic interactions deviate most from the empirical baselines.

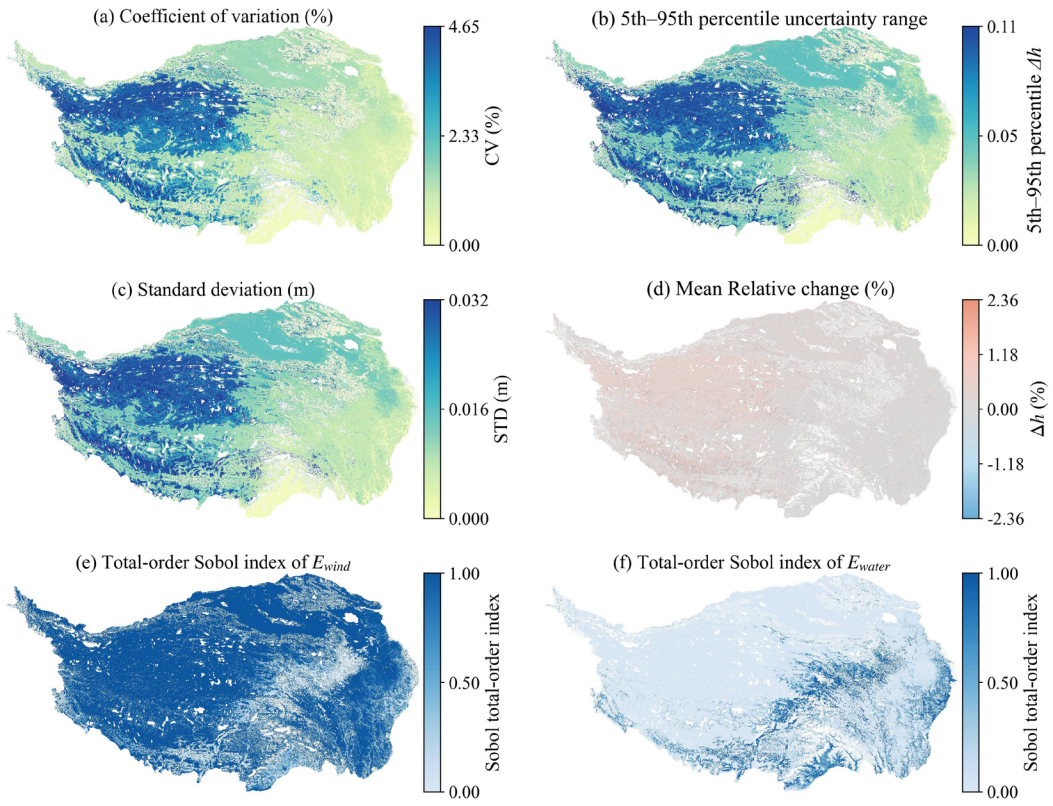


Figure R3. Spatial distribution of uncertainty and sensitivity in the simulated solum thickness across the QTP due to uncertainties in the empirical wind and water erosion inputs (RUSLE and RWEQ). (a) Coefficient of variation (CV, %), (b) 5th–95th percentile range. (c) Standard deviation of the 150 ensemble simulations. (d) Relative difference between ensemble mean and baseline. (e-f) Sobol total-order sensitivity indices for E_{wind} and E_{water}, respectively.

3 Theoretical reframing to the steady-state assumption. We have revised the text to explicitly connect these quantitative findings to the steady-state assumption. We now clarify that the weighting coefficients act as "effective scaling factors" that absorb the LHS-quantified uncertainties in the transient inputs. In section 4.3:

"To partially reduce the bias associated with the steady-state assumption, we introduced adjustable weighting coefficients (α, β, γ) to scale the relative contributions of different erosion processes… These coefficients should be interpreted as effective scaling factors rather than direct proxies for deviations from geomorphic equilibrium. They absorb multiple sources of uncertainty, including the biases in empirical erosion estimates (E_{water}, E_{wind}) quantified in our LHS analysis, unresolved transient dynamics, and scale mismatches. Therefore, they do not explicitly resolve true non-steady-state mass

imbalances, but rather ensure a physically constrained, quasi-steady representation of the regional baseline."

Minor Comments

Abstract

- Lines 24–25: Specify the metric(s) in the performance claim (e.g., "by approximately 10–17% in RMSE and MRE").

In the revised abstract, we have specified the exact metrics achieved by the model (RMSE of 0.34 m and MRE of 0.78). Furthermore, consistent with our revisions regarding the shared validation data (detailed in our response to Major Comment #1), we have carefully rephrased this performance claim. Rather than claiming absolute independent outperformance, we explicitly state these metrics represent a relative reduction under the available cross-validation framework.

Study Area and Data

- Line 105: Briefly mention known limitations of the Shangguan et al. (2017) DTB dataset in the QTP interior when it is first introduced.

We thank the reviewer for this helpful suggestion. We have revised the text in Section 2.1 to explicitly note the limitations of the Shangguan et al. (2017) dataset regarding the QTP interior. In section 2.1:

"...To ensure physical realism, we integrated the global depth-to-bedrock (DTB) dataset developed by Shangguan et al. (2017) as a critical structural constraint. This dataset, generated at a 1km resolution using an ensemble of machine learning algorithms, defines the total available weathered mantle. However, it should be noted that uncertainties remain in the interior regions of the QTP due to highly sparse field observations and a heavy reliance on model-based extrapolation."

Methods

- Lines 165–170: Justification for excluding human activities is too brief. Note that effects are assumed negligible at 1 km resolution.

In the revised manuscript, we have expanded this explanation to explicitly clarify that, given the 1 km spatial resolution and the sparsely populated nature of the QTP, geomorphic and climatic factors overwhelmingly dominate large-scale erosion patterns.

"...Other processes, including freeze–thaw and human activities, are not treated as independent erosion categories. Freeze–thaw processes primarily act as a weathering mechanism that generates loose material through frost action, which is subsequently transported by fluvial, aeolian, and gravitational processes (Guo et al., 2015a). Furthermore, given the 1 km spatial resolution and the vast area with sparse population of the QTP, geomorphic and climatic controls dominate large-scale erosion patterns, while the effects of human activities are comparatively negligible at this scale (Sun et al., 2020)."

- Table 1: Move "Parent Material of Soil Formation" to a new "Geological" category.

Thanks. We've updated Table 1 in response to your suggestion.

Results

- Table 3: Add "(–)" for the unitless NDVI column.

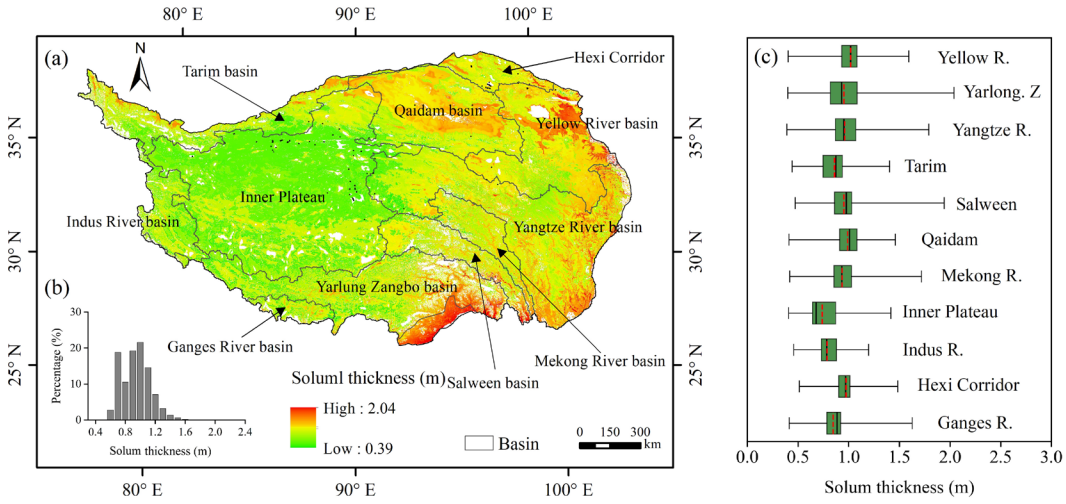
Many thanks and we've fixed this formatting problem.

- Table 4: Replace "-" for Cluster 8 RMSE with "N/A (n=4 insufficient for cross-validation)". Add a footnote explaining *n*.

We have updated Table 4 in the revised manuscript to replace the "-" with "N/A". We explained n as the number of soil profile observations used for calibration in each cluster. N/A: the sample size of 4 is insufficient for reliable cross-validation.

- Fig. 3: Consider adding a histogram inset or basin-specific summary statistics.

In the revised manuscript, we have updated Figure 3 to include a new histogram inset (panel b) and a basin-level boxplot of solum thickness (panel c). Furthermore, we have updated the main text to explicitly interpret this distribution, highlighting that the model predominantly produces moderately developed soils, with thin and thick extremes cleanly partitioned by geomorphic regimes. We also added a description of the differences in solum thickness among different basins. For example, the Yarlung Zangbo Basin in southeastern QTP exhibits the greatest solum thickness, but also the largest standard deviation due to its complex terrain. In contrast, the permafrost-dominated basins in the inner QTP show the smallest solum thickness and the lowest internal variability.



Revised Figure 3. Spatial pattern and frequency distribution of simulated solum thickness across the QTP. (a) Spatial distribution of simulated solum thickness on the QTP. The delineated areas for the Mekong, Salween, and Yarlung Zangbo (Brahmaputra) rivers represent the catchment sections located within China. White patches indicate non-soil areas (glaciers and water bodies) where solum thickness was not simulated. (b) The inset histogram shows the frequency distribution of simulated solum thickness across the QTP. (c) Boxplot showing the distribution of solum thickness across different basins of the QTP.

Supplement table. Statistics of solum thickness at the basin level across the QTP

Basin	Mean (m)	Min (m)	Max (m)	SD (m)
Ganges River	0.85	0.41	1.63	0.15
Hexi Corridor	0.97	0.51	1.48	0.12
Indus River	0.79	0.46	1.20	0.12
Inner Plateau	0.74	0.41	1.42	0.12
Mekong River	0.94	0.42	1.72	0.13
Qaidam Basin	0.99	0.41	1.46	0.13
Salween Basin	0.95	0.47	1.94	0.17
Tarim Basin	0.86	0.44	1.40	0.16
Yangtze River	0.96	0.39	1.79	0.17
Yarlung Zangbo	0.95	0.40	2.04	0.22
Yellow River	1.02	0.40	1.59	0.16

- Lines 330–340: explain why curvature shows near-zero correlation at plateau scale yet drives local differentiation in the model.

We thank the reviewer for highlighting this apparent paradox. we have addressed it as part of our broader revision to Section 3.3. In the revised text, we explain that general curvature fluctuates rapidly between positive (divergent ridges) and negative (convergent valleys) at the local scale. When aggregated across the immense 2.5 million km² area of the plateau, this high-frequency spatial signal statistically averages out to near-zero, masked by the dominant, low-frequency climatic gradients. However, at the hillslope scale, the physical mass balance model correctly preserves and responds to these local terrain features. In section 3.3:

"...Conversely, the near-zero plateau-scale correlation with general curvature (r = -0.01) should not be interpreted as evidence of negligible topographic influence. Rather, it indicates that curvature-related mass redistribution fluctuates rapidly around zero at the local scale, and this high-frequency signal is statistically masked by the immense, low-frequency climatic gradients when aggregated across the plateau.

At the hillslope scale, however, this topography-driven spatial redistribution becomes clearly visible. As demonstrated in the transect analysis across the Hengduan Mountains (Fig. 5) …simulated thickness varies systematically with terrain position: steep ridges (divergent curvature) act as sediment-source areas… whereas gentler slopes and valley bottoms (convergent curvature) function as depositional zones."

Discussion and Comparison

- Table 5: Explicitly note the shared validation data limitation. explain the higher MAE.

We thank the reviewer for this important suggestion. We have addressed both points in the revised text accompanying Table 5.

1 Consistent with our comprehensive revisions for Major Comment #1, we have updated the text introducing Table 5 to explicitly acknowledge the overlapping nature of the validation profiles. We now strictly frame these metrics as a relative assessment under shared observational constraints rather than a fully independent validation.

2 We have expanded the discussion to explain why our model exhibits a slightly higher Mean Absolute Error (MAE of 0.13 m) compared to the Shangguan map (0.06 m).

The legacy statistical maps tend to produce spatially smoothed estimates. While statistical smoothing mathematically minimizes average absolute errors, it artificially

masks sharp, terrain-driven physical extremes. In contrast, our physical mass balance model intentionally preserves this high-frequency spatial heterogeneity (e.g., thin regolith on steep ridges and thick accumulations in valleys). Therefore, the slightly higher MAE is a natural byproduct of capturing a wider, more physically realistic range of solum thicknesses. In Section 4.1:

"It should be noted that these 552 observations partially overlap with the source data used to generate the reference maps. Therefore, this comparison must be interpreted as a relative assessment under shared observational constraints, rather than as a fully independent external validation.

Under this framework... our model exhibited a slightly higher MAE (0.13 m) than the Shangguan map (0.06 m). This difference reflects the contrasting spatial behavior of the datasets. The reference maps tend to produce spatially smoothed estimates, which mathematically reduce average absolute errors but underrepresent sharp, terrain-driven variations. In contrast, the revised mass balance model preserves a much wider and more physically realistic range of solum thickness, from thin regolith on steep ridges to thicker accumulations in valley bottoms (Fig. 3 and Fig. 5). Thus, the higher MAE should be interpreted together with the lower RMSE and MRE, indicating that our model sacrifices some statistical smoothness to better capture first-order physical spatial structures."

- Lines 470–485: Strengthen discussion of potential RUSLE/RWEQ biases in permafrost areas and suggest future use of cryogenic-specific erosion models.

In the revised manuscript, we have significantly strengthened the discussion regarding potential biases in RUSLE and RWEQ estimates. We explicitly acknowledge that these models may not fully account for the complex interactions of seasonal freeze-thaw cycles and ice-rich permafrost layers on sediment mobility. Furthermore, we have added a dedicated recommendation for the future development and integration of QTP-specific cryogenic parameterizations for erosion and weathering to reduce these systematic uncertainties. In Section 4.2:

"A primary source of uncertainty lies in the spatially explicit erosion rates derived from the empirical RUSLE and RWEQ models. Originally developed for low-altitude temperate zones, these models may have limited adaptability to the cryogenic environment of the QTP, where freeze-thaw cycles and permafrost conditions significantly alter sediment transport dynamics. This may introduce potential biases in erosion estimation due to the absence of cryogenic-specific processes in these frameworks. ... Future research should aim to incorporate QTP-specific cryogenic

parameterizations that explicitly account for freeze-thaw dynamics and permafrost degradation, thereby improving the representation of cold-region pedogenic and erosional processes.”

- Lines 500–510: Rephrase the sentence about weighting coefficients mitigating steady-state bias to “partially mitigated”.

We thank the reviewer for this important correction. In the revised manuscript, we have rephrased this section to explicitly state that the weighting coefficients only "partially" reduce the biases associated with the steady-state assumption.