Response to reviewers’ comments

Response to Reviewer #1:

[Comment 1] In this paper, the authors hypothesize that the mismatch between modelled and measured global silicate-weathering rates are due to uncertainties in the underlying erosion models. The authors therefore present a global erosion-rate model trained against cosmogenic-nuclide derived erosion rates using a random forest approach. They show that this model can reproduce erosion rates much better than a basic stream-power-erosion approach. The authors then incorporate their predicted erosion patterns into an existing steady-state silicate-weathering model and demonstrate that the new product can predict both the total global weathering flux and the spatial distribution of weathering rates better than previous models.

I found the study to be very well motivated, well written, and convincingly argued. The analysis is sound and provides a significant contribution for modelling global silicate-weathering fluxes. The figures and text are of a high quality. In particular, I appreciate the explanations of complex terms and approaches (e.g. the SHAP work) for the unfamiliar reader. I only have a few relatively minor comments that may be interesting to consider before publication.

Response: Thank you very much for your thoughtful and constructive review of our manuscript. We are gratified by your positive assessment of the study's motivation and the clarity of our explanations. Your comments have helped us identify areas where our terminology was ambiguous or where the data presentation required more rigorous quantification.

[Comment 2] L478 – 496: You write that Fig. 9 “clearly illustrates the trade-off”, that the value in Fig. 9a is “far to the right of the observed range” and that its value within the square-band is “low”. To my eyes, the peak is just to the right of the optimum and there are still reasonable values within the grey band. Next, you write that in Fig. 9b, “peaks of all three metric curves are now located squarely within the grey observational band”. For the red and blue curves that is true, but the peak of the yellow curve is clearly not in the square band and actually looks quite close to the location in Fig 9b. Can you revise this text so that the text matches the data more clearly?

Response: We sincerely thank the reviewer for this rigorous and insightful evaluation of Figure 9. We appreciate the reviewer pointing out that our descriptions of the curves' positions relative to the observational grey band were not sufficiently precise.

We agree that the term "far to the right" in L478-496 is subjective. In Figure 9a (SPIM), the R_{log}^2 peak (representing the optimal fit to the spatial pattern of observations evaluated with the original metric Park et al. used in the paper) occurs at a global flux of $\sim 4.2 \times 10^{12}\ mol\ yr^{-1}$ which is approximately 1.7 times of the observation-based central estimate of $2.5 \times 10^{12}\ mol\ yr^{-1}$. While some R_{log}^2 values within the grey band are technically "reasonable" (0.41

compared to its optimal of 0.53), the model's best spatial performance (e.g., the peak) cannot be achieved without overestimating the global total flux.

The reviewer is correct that the peak of the yellow curve (R^2_{log}) in Figure 9b sits at $\sim 3.5 \times 10^{12}\ mol\ yr^{-1}$, which is indeed outside the $2.0 - 3.0 \times 10^{12}\ mol\ yr^{-1}$ grey band. Our previous statement that "all three metric curves" were "squarely within" the band was inaccurate.

We have revised the text to clarify that the significant improvement in MErSiM (Fig. 9b) is the convergence of the metrics. Specifically, the peak of the joint metric ($R^2 + R^2_{log}$, blue curve) and the standard R^2 (red curve) now fall within the observational range. The final parameter set we used for simulation is those who optimize $R^2 + R^2_{log}$.

[Comment 3] Perhaps, you can also explicitly quantify (1) the global weathering fluxes predicted by the two different models (MErSiM versu Park20) including an estimate of uncertainty for both, and (2) the difference between each one of these predictions and the measured total weathering fluxes.

Response: In the original manuscript, we have explicitly quantified the predicted flux in Section 3.5.1 and in Table 3 where the sensitivity test is presented, but the reviewer is right that this quantification could be mentioned in this section. Following the reviewer's suggestion, we have added a text summary explicitly comparing the model predictions to observations. To account for uncertainty in parameter selection, we consider all parameter sets whose metric value is within 0.05 of the maximum. The prediction uncertainty is then quantified by the variability of model outputs across these acceptable parameter sets.

Measured target: $2.5 \pm 0.5 \times 10^{12}\ mol\ yr^{-1}$

Park20 (GM09-SPIM): $4.54 \pm 0.4 \times 10^{12}\ mol\ yr^{-1}$ (Error: $2.04 \times 10^{12}\ mol\ yr^{-1}$, 82% bias).

MErSiM (Revised): $3.11 \pm 0.3 \times 10^{12}\ mol\ \ yr^{-1}$ (Error: $0.61 \times 10^{12}\ mol\ yr^{-1}$, 24% bias).

[Comment 4] Finally, in the section L478 – 496 you claim that these results speak to the trade-off between matching total weathering fluxes versus the pattern of weathering fluxes. I did not understand how Figure 9 relates to (or informs) the spatial pattern of silicate weathering fluxes.

Response: We appreciate the reviewer's request for clarification on the relationship between the metrics in Figure 9 and the "spatial pattern" of weathering. In our evaluation framework, the performance metrics (R^2 and R^2_{log}) shown in Figure 9 are not calculated for a single location, but represent the agreement between modeled and observed weathering fluxes across 81 major river basins distributed globally. Because these 81 basins encompass a vast range of tectonic settings (from active orogens like the Andes to stable cratons like the Amazon) and climatic zones (from the tropics to high latitudes), a high R^2 value indicates that the model successfully reproduces the observed spatial heterogeneity and relative differences in weathering rates among these diverse regions. Therefore, the "model performance" plotted on the y-axis of Figure 9 is a direct proxy for how well the model captures the global spatial pattern of silicate weathering, while the value on the x-axis indicates the performance in matching the total weathering flux. A tradeoff is thus obtained if the maximum y value is obtained for an x value within the grey band. We have clarified this definition in the revised manuscript to avoid ambiguity.

[Comment 5] L497 – 504: Similar to above, the text here seems a bit more pushed than warranted based on Fig 10. I agree that the errors are clearly smaller in the RF model, but there are still quite large errors, in particular in the tropical regions. When you write things like “large” errors and is “much less” severe, can you explain quantitatively what you understand by a large and small error. How much were errors reduced on average etc.?

Response: We thank the reviewer for this constructive critique. We agree that using qualitative descriptors like "large" or "much less severe" without clear benchmarks can be subjective. In the revised manuscript, we have explicitly defined "large errors" as those exceeding $5 \times 10^{10}\ mol\ yr^{-1}$.

To provide a clearer comparison, we have calculated the following metrics across the 81 major basins. First, MErSiM reduced the Mean Absolute Error (MAE) across all 81 basins by approximately 35% compared to the original GM09-SPIM ($1.77 \times 10^{10}\ mol\ yr^{-1}$ compared to $2.72 \times 10^{10}\ mol\ yr^{-1}$). Second, the number of basins exhibiting "large" positive errors ($>5 \times 10^{10}\ mol\ yr^{-1}$) dropped from 10 in GM09-SPIM to 4 in MErSiM. Third, specifically in the tropics (basins located between 23S-23N), where GM09-SPIM errors often exceeded $7 \times 10^{10}\ mol\ yr^{-1}$, MErSiM reduced the average basin-scale bias by 90% ($0.41 \times 10^{10}\ mol\ yr^{-1}$ compared to $4.12 \times 10^{10}\ mol\ yr^{-1}$). We have added these quantitative description in the explanation paragraph of Fig. 10.

We appreciate the reviewer for pointing out that significant errors remain in the tropics. We have added a paragraph noting that these persistent biases likely stem from uncertainties in the cation concentrations of sedimentary and metamorphic rocks, which remain as adjustable bulk parameters in the current framework.

[Comment 6] I was confused by the discussion about the effect of CO2 increases. You explain your models small response to the CO2 increase by widespread chemostasis and a resulting runoff insensitivity of the weathering flux. My understanding is exactly the opposite: To me, chemostasis means that concentrations remain constant with changes in runoff – contrary to what is suggested in L568 (e.g. Godsey et al., 2019; Godsey et al., 2009). Under these conditions (common in active mountains (Godsey et al., 2019)) the weathering fluxes should strongly increase with runoff. Hence, the system should have a high sensitivity to increased CO2 and runoff. Something is off here, and maybe it is just about the term chemostasis? I guess your model also has a negative temperature response, does that matter here?

Response: We are very grateful to the reviewer for pointing out the potential confusion regarding the term "chemostasis." We agree that in the hydrological community (e.g., Godsey et al., 2009, 2019), chemostasis typically refers to a state where solute concentrations remain nearly constant despite changes in discharge, which implies that weathering fluxes should scale linearly with runoff.

In the original manuscript, we used "chemostatic" in the context of the Damköhler number framework (Maher and Chamberlain, 2014) to describe a state of weathering flux saturation. We acknowledge that this usage is ambiguous and contradicts the definition favored by the reviewer.

In our revised model, under greenhouse conditions (), the system enters a regime where the fluid travel time (τ_f) is much shorter than the reaction equilibration time (τ_{eq}). This leads to a "sluggish" response, causing the total weathering flux to plateau despite in higher temperature scenario.

The reviewer correctly intuited the role of the negative temperature response. In our data-driven erosion module, warming in high-latitude or high-altitude regions can move the environment out of the optimal "frost-cracking window (Anderson and Anderson, 2010; Zhao et al., 2026)," leading to a localized decrease in erosion rates. This reduction in mineral supply effectively decouples the weathering flux from the kinetic acceleration expected from rising temperatures.

We have revised the text to replace "chemostatic" with "kinetically limited" and added a discussion on the coupling between the negative temperature-erosion response and the attenuated weathering feedback.

[Comment 7] Your paper demonstrates that an erosion model based on 7 input parameters performs equally well to a model based on 14 parameters. That makes me wonder: would a model with even fewer parameters also work? For example, if I understand Fig. 6 correctly, the lithology may play a minor role. Would it be feasible to progressively eliminate parameters from your model (starting with the least important) and see when the model starts breaking down?

Response: We thank the reviewer for this insightful suggestion regarding model robustness. Achieving a balance between model simplicity and predictive accuracy is particularly important for deep-time applications where boundary conditions are less certain.

While our SHAP analysis (Fig. 6) suggests that lithology (ROCK) has a lower GLOBAL importance compared to other predictors, it remains a fundamental factor in geomorphology as it dictates rock strength and initial cation concentrations (Bluth and Kump, 1994; Bufe et al., 2022). When modeling LOCAL erosion rates, lithology sometimes serves as the only other predictor other than slope (e.g., Zondervan et al., 2023). Removing this variable could lead to systematic spatial biases, especially in regions with distinct lithological controls.

To address this, we have previously conducted experiments by progressively removing variables. When ROCK was eliminated (leaving SLOPE, ELEVATION, MAT, LAI, RUNOFF, and PWET), the erosion model's validation R^2 decreased to 0.78 (from 0.81 in the baseline), and the weathering model's maximum $R^2 + R_{log}^2$ dropped to 0.72 (from 0.86 in the 7-parameter version). Although the performance decline is not "catastrophic," the 7-parameter model consistently provides a superior fit to both erosion and weathering datasets. So are the results of experiments eliminating other 6 predictors.

To further evaluate the necessity of these 7 variables, we have performed a Permutation Importance Test. Unlike SHAP, which measures local contributions, this test assesses global importance by randomly shuffling a single feature's values and measuring the resulting change in the model's Mean Absolute Error (MAE).

In this test, all seven variables yielded a positive MAE change, meaning the exclusion of any single variable increases the model's prediction error (Figure R1). In terms of its impact on global error (MAE change), ROCK was found to be more significant than both RUNOFF and LAI. This confirms that while lithology may not be the primary driver of local variation in all cases, it is essential for the model's overall stability and global accuracy. Therefore, we believe the 7-parameter configuration represents the "optimal" set of predictors for MErSiM v1.0.

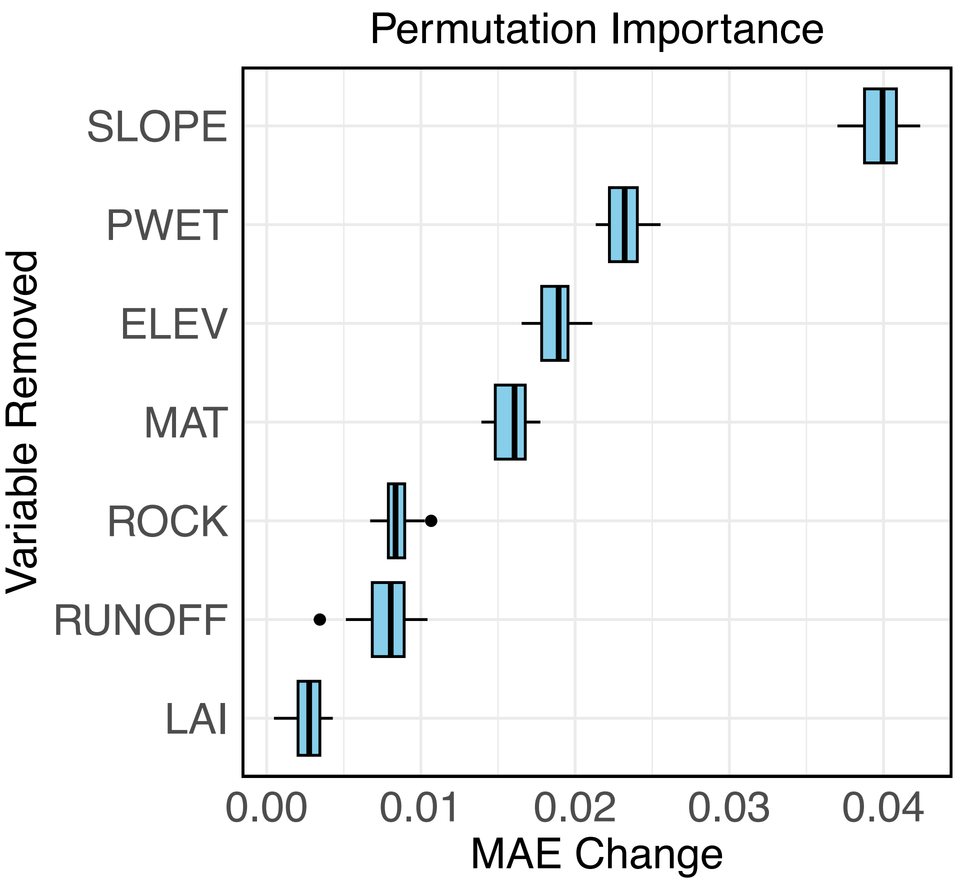

Figure R1 Mean Absolute Error (MAE) change of randomly shuffling a single feature's value

[Comment 8] L10: Which “widely used weathering model”- can you be specific? There are several out there.

Response: We referred to the GM09 framework. The rephrased L10 is “The widely-used weathering model framework of Gabet and Mudd (2009), when driven by stream power erosion laws…”

[Comment 9] L51: became

Response: Fixed to "became." Thanks.

[Comment 10] L60: How can a model “bias [be] supported by […] data”? Wasn’t your point that the bias is a bias because it doesn’t match global fluxes?

Response: Thanks. We have rephrased. The "bias" is supported by data in the sense that independent proxies (like Osmium isotopes) also suggest the model's spatial distribution of weathering is incorrect. The rephrased L60 is “It stems from a systematic overestimation of weathering fluxes specifically in tropical regions, a spatial mismatch corroborated by osmium isotope data (Rugenstein et al., 2021)."

[Comment 11] L102 – 136: Please systematically define all variables (even if they are obvious); e.g. missing t in equation 1 (definition comes only for equation 2) and many of the parameters in the Arrhenius relationship (equation 6).

Response: We have defined all parameters in the manuscript. The rephrased explanation is "The core of the GM09 model describes the change as a function of time (t, in years) in regolith thickness (h, in meters) as a balance between…" for equation (1) and "Where E_a is the activation energy (J/mol), R is the ideal gas constant (8.314 J/(mol·K)), T_0 is the reference temperature (K), and k_{rp} is a proportionality constant" for equation (6).

[Comment 12] L140: I guess you could cite West (2012) here who explores that steady-state expression that you show

Response: Thanks. We have added the citation here.

[Comment 13] L149: You have the variable name R already in the Arrhenius relationship. It would be useful to choose a different name for runoff – for example qw.

Response: We have renamed the runoff variable from R to q to avoid confusion with the universal gas constant.

[Comment 14] L263: By "this compilation" you refer to the Gaillardet et al. (1999) compilation, right? Maybe specify to clarify sentence

Response: We will specify this compilation as "the compilation of Gaillardet et al. (1999) and HYBAM"

[Comment 15] Figure 3: Please define the abbreviations in the figure caption

Response: We have added definitions for all abbreviations (e.g., PWET, LAI) and units (e.g., mm/kyr, °C) to the caption. Thanks for bringing this to our attention.

[Comment 16] Figure 8: Can you indicate the units of these variables?

Response: We have added the units in the caption.

Reference

Anderson, R.S., Anderson, S.P., 2010. Geomorphology : The Mechanics and Chemistry of Landscapes.

Bluth, G., Kump, L., 1994. Lithologic and Climatologic Controls of River Chemistry. Geochimica Et Cosmochimica Acta 58, 2341–2359. https://doi.org/10.1016/0016-7037(94)90015-9

Bufe, A., Cook, K.L., Galy, A., Wittmann, H., Hovius, N., 2022. The effect of lithology on the relationship between denudation rate and chemical weathering pathways – evidence from the eastern Tibetan Plateau. Earth Surface Dynamics 10, 513–530. https://doi.org/10.5194/esurf-10-513-2022

Coogan, L.A., Caves Rugenstein, J.K., 2025. Regulation of the carbon cycle on geological timescales, in: Treatise on Geochemistry. Elsevier, pp. 419–465. https://doi.org/10.1016/B978-0-323-99762-1.00060-7

Dessert, C., Dupré, B., Gaillardet, J., François, L.M., Allègre, C.J., 2003. Basalt weathering laws and the impact of basalt weathering on the global carbon cycle. Chemical Geology, Controls on Chemical Weathering 202, 257–273. https://doi.org/10.1016/j.chemgeo.2002.10.001

Gabet, E.J., Mudd, S.M., 2009. A theoretical model coupling chemical weathering rates with denudation rates. Geology 37, 151–154. https://doi.org/10.1130/G25270A.1

Gaillardet, J., Dupré, B., Louvat, P., Allègre, C.J., 1999. Global silicate weathering and CO2 consumption rates deduced from the chemistry of large rivers. Chemical Geology 159, 3–30. https://doi.org/10.1016/S0009-2541(99)00031-5

Maher, K., Chamberlain, C.P., 2014. Hydrologic Regulation of Chemical Weathering and the Geologic Carbon Cycle. Science 343, 1502–1504. https://doi.org/10.1126/science.1250770

Moon, S., Chamberlain, C.P., Hilley, G.E., 2014. New estimates of silicate weathering rates and their uncertainties in global rivers. Geochimica et Cosmochimica Acta 134, 257–274. https://doi.org/10.1016/j.gca.2014.02.033

Rugenstein, J.K.C., Ibarra, D.E., Zhang, S., Planavsky, N.J., Von Blanckenburg, F., 2021. Isotope mass-balance constraints preclude that mafic weathering drove Neogene cooling. Proc. Natl. Acad. Sci. U.S.A. 118, e2026345118. https://doi.org/10.1073/pnas.2026345118

Zhao, J., Liu, Y., Li, G., Zuo, H., 2026. Thresholds in the controls of denudation rates: A global analysis of tectonic, climatic and biological factors based on machine learning. Earth and Planetary Science Letters 674, 119750. https://doi.org/10.1016/j.epsl.2025.119750

Zondervan, J.R., Hilton, R.G., Dellinger, M., Clubb, F.J., Roylands, T., Ogrič, M., 2023. Rock organic carbon oxidation CO2 release offsets silicate weathering sink. Nature 1–5. https://doi.org/10.1038/s41586-023-06581-9