Author's response

Manuscript:

Long-term hydro-sedimentary dynamics of the Ucayali River (Amazon Basin) revealed through combined observations, remote sensing, and SWAT-Amazon modelling

Hydrology and Earth System Sciences – egusphere-2025-4101

William Santini et al.

14 April 2026

We thank the Editor and the three reviewers for their thorough and constructive evaluation of our manuscript. The comments collectively identified five main areas requiring substantial revision:

(i) clearer separation between calibration and independent validation;
(ii) rigorous quantification and propagation of uncertainty and equifinality;
(iii) stronger validation of internal floodplain processes;
(iv) justification and sensitivity analysis of the two-fraction sediment class assumptions;
(v) and improved methodological transparency, including clearer research objectives and reproducibility.

Each of these points has been addressed in the revised manuscript, leading to substantial additions and restructuring. Most notably, a GLUE-based uncertainty propagation framework (2500 Latin Hypercube simulations) was implemented and is fully described in a new Supplementary Material section (S12); a dedicated equifinality analysis with dotty plots was added (Section 6.1.3 and Supplementary S11); the calibration/validation strategy was clarified and complemented by independent hold-out performance metrics and direct comparisons with gauging measurements (Supplementary S9); and the abstract, introduction, and conclusion were substantially rewritten to better articulate the research questions and key findings with their uncertainty bounds.

The structure of Section 6.3 was also reorganized into two subsections to clearly separate structural model limitations (6.3.1) from quantified uncertainty (6.3.2), and a new section (6.4.3) was added to discuss the broader implications of the floodplain-controlled sand sedimentation process identified in this study.

Minor revisions include figure color scheme improvements, clarification of the two-fraction sediment representation, and additional details on measurement protocols now consolidated in a new Supplementary section (S2).

Detailed point-by-point responses to all comments from Referee #1 (RC1), Referee #2 (RC2), and the Community Comment (CC1) are provided below.

Sincerely,
W. Santini et al.

Response to Referee Comment #1 (RC1)

General Comments

This manuscript presents an ambitious and innovative integration of long-term in situ observations, satellite remote sensing, and a modified SWAT framework to quantify multi-decadal hydro-sediment dynamics in the Ucayali River. The development of SWAT-Amazon and the explicit treatment of floodplain hydraulics and sand routing represent a substantial methodological advance for large, low-gradient tropical rivers, and the assembled observational dataset, particularly the long-term "super station" records and dedicated field campaigns, is a clear strength of the study. The results are compelling and potentially impactful for understanding sediment trapping, recycling, and floodplain controls in the Amazon Basin.

My comments below are intended to strengthen the robustness, generalisability, and interpretability of the findings. In particular, they focus on clarifying the separation between calibration and validation, quantifying uncertainty and equifinality in key inferred budgets, and providing additional evidence that internal floodplain processes are realistically represented rather than inferred solely from outlet behavior. I examined the Supplementary Material, while it provides valuable methodological detail, it does not address the validation, uncertainty, or equifinality issues raised below. Addressing these points would substantially increase confidence in the reported trapping and recycling fractions and enhance the value of the framework for application to other Amazonian basins.

Specific comments

1. The calibration period for discharge and sand overlaps almost entirely with the reported validation; I am wondering whether the authors could provide performance metrics for an independent hold-out period to demonstrate that the high NSE values are not artifacts of overfitting.

This is an important point that deserves clarification. We acknowledge that the wording in the original manuscript was ambiguous regarding the distinction between calibration and the reported performance periods, which may have given the impression that calibration and validation fully overlap. We also acknowledge that the available monitoring dataset limits the definition of a fully independent validation period for sand routing in the conventional sense. Nevertheless, model robustness is supported by: (i) consistent performance metrics over periods outside the calibration window (Table RC2.1), with reduced performance at Lagarto attributable to increased forcing and observational uncertainties rather than model degradation; and (ii) the comparison with direct gauging measurements (Fig. RC2.3), which constitute a quasi-independent validation as these observations are not explicitly used in calibration. Residual uncertainty is further addressed through the GLUE analysis presented in Section 6.

We address a complete response below:

For water routing (Section 5.1), the 2000–2016 period does not correspond to a calibration period, but rather to a common observation window used to compute performance metrics across stations, depending on data availability. Model calibration for water discharge (Table RC1.1) was conducted over the 2010–2015 period (except for Puerto Inca, for which the period was 2012–2014), selected based on data quality and availability. Furthermore, calibration does not rely on the optimization of a single time series, but on multiple hydraulic diagnostics, including water levels (Fig. 4a), velocities (Fig. 4b), and stage–discharge and stage–velocity relationships (Fig. 5). Moreover, independent hold-out periods were used to further evaluate model performance (Table RC1.1), showing consistent performance metrics across all periods.

Table RC1.1. Discharge model performance over calibration and independent validation periods. Calibration was performed over 2010–2015 (except for Puerto Inca, for which the period was 2012–2014), while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed daily discharge records.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Independent validation	01/2009–12/2009	287	0.9	0.94	-1.3
Lagarto	Calibration	01/2010–12/2015	2191	0.89	0.94	-1.7
Lagarto	Independent validation	01/2016–12/2019	1162	0.84	0.90	-1.7
Puerto Inca	Calibration	09/2012–08/2014	730	0.73	0.86	-1.1
Puerto Inca	Independent validation	09/2015–08/2016	731	0.69	0.83	-7.1*
Pucallpa	Independent validation	01/2000–09/2009	3653	0.94	0.94	-4.8
Pucallpa	Calibration	01/2010–12/2015	2161	0.92	0.89	1.1
Pucallpa	Independent validation	01/2016–09/2019	1389	0.91	0.94	-1.2
Requena	Independent validation	01/2000–09/2009	3653	0.92	0.93	-3.8
Requena	Calibration	01/2010–12/2015	2191	0.9	0.95	0.1
Requena	Independent validation	01/2016–12/2019	1461	0.85*	0.83*	-9.5*

* Lower performance during these periods is associated with reduced quality of observed discharge data and/or precipitation inputs.

For sand routing (Section 5.3), the 09/2009–08/2015 period shown in Fig. 6 corresponds to the interval with the most reliable and dense observations, as explained in the manuscript, although data quality remains heterogeneous depending on the period and sampling protocols. This period was therefore selected for model calibration. However, outside this interval, uncertainties in sand flux observations (derived from surface sediment concentration monitoring and sediment gauging; see Supplementary Material S3) increase, which precludes the definition of a robust and fully independent validation period. In particular, at Lagarto, the NSE for the validation period decreases to 0.45 (Table RC1.2). When excluding high waters (flood peaks), the NSE remains high (0.80), indicating robust model performance under low- and intermediate-flow conditions. The lower performance is therefore mainly associated with rapid Andean flood events and related uncertainties in rainfall forcing.

Table RC1.2. Sand routing model performance over calibration and independent validation periods. Calibration was performed over 2009–2015, while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed sand flux at surface concentration sampling time step.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Calibration	09/2009–08/2015	323	0.80	0.87	7.9
Lagarto	Independent validation	09/2015–08/2018	181	0.45	0.60	27.7
Requena	Calibration	09/2009–08/2015	404	0.86	0.92	-2.3
Requena	Independent validation	09/2015–08/2018	77	0.80	0.70	-24.0

Overall, reduced performance after the calibration period reflects both increased observational and forcing uncertainties rather than a degradation of model performance.

Therefore, to further evaluate model performance, we performed an additional comparison between daily simulations (discharge and suspended sand fluxes) and direct gauging measurements over the 09/2009–08/2016 period (Fig. RC1-1). This period starts with the implementation of improved sediment sampling protocols during field measurements, while simulations after 2016 were not considered due to increased uncertainties in rainfall inputs at

Requena (Fig. 3d). This comparison bypasses rating curves and provides a more direct evaluation of model performance. Although not fully independent, it offers a complementary and partially independent validation, as direct gauging measurements are not explicitly used in the calibration, even though they contribute to the construction of rating curves over the full observation period. The agreement between simulations and gauging measurements supports model robustness, with bias-corrected R^2 (bR^2) values of 0.90 for discharge and 0.88 for sand fluxes (Fig. RC1-1).

The corresponding results have been added in the Supplementary Material (Fig. S9) and are briefly discussed in the revised manuscript. We have also revised the text to explicitly distinguish between calibration, performance assessment, and the different sources of observational constraints.

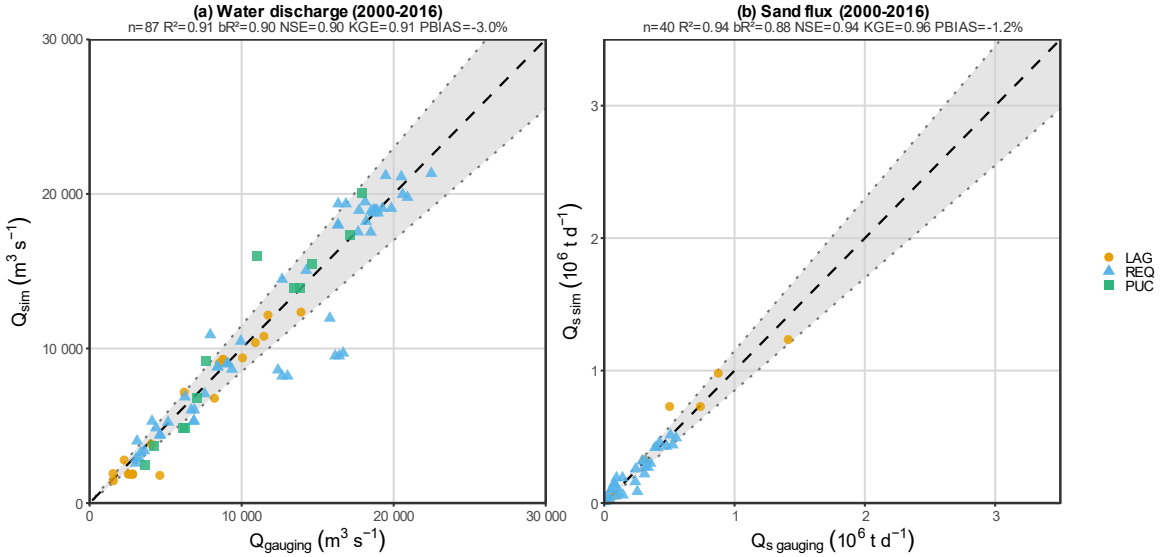

Figure RC1.1: Comparison of simulated and measured river discharge (a) and suspended sediment flux (b) at gauging stations in the Ucayali basin. Simulated values ($\boldsymbol{Q_{sim}}$, $\boldsymbol{Q_{s,sim}}$) correspond to daily outputs from the SWAT-Amazon model extracted at the dates of field measurements. Observed values ($\boldsymbol{Q_{gauging}}$, $\boldsymbol{Q_{s,gauging}}$) are direct field measurements, independent of rating curve derivation. The dashed line indicates the 1:1 relationship, and the shaded band represents the ±15% envelope. The Puerto Inca station was excluded, as its sub-daily flood pulse dynamics are not adequately captured at the model's daily time step. Axis ranges reflect the observed variability of $\boldsymbol{Q}$ and $\boldsymbol{Q_s}$ across all stations. Statistical indicators are computed over all stations pooled: n = number of paired observations; R^2 = coefficient of determination; bR^2 = bias-corrected coefficient of determination; NSE = Nash–Sutcliffe efficiency; KGE = Kling–Gupta efficiency; PBIAS = percent bias.

The manuscript was revised as follows:

Section 5.1

For water discharge, model calibration was performed over the 2010–2015 period based on multiple hydraulic diagnostics (water levels, velocities, and stage–discharge relationships). Model performance was further evaluated using independent hold-out periods and direct comparisons with gauging measurements that bypass rating curves, as detailed in the Supplement S9. These complementary evaluations show consistent performance across periods and support the temporal robustness of the model.

At a daily time step, SWAT-Amazon simulations at the watershed outlet show excellent performance (Fig. 3d): NSE (Nash-Sutcliffe Efficiency) = 0.92, KGE (Kling–Gupta efficiency) = 0.95, PBIAS (Percent Bias) = -1.8%, LogNSE (NSE on the logarithms of the series) = 0.92 over the 2000–2016 observation period used here for inter-station evaluation (see Moriasi et al. (2007) for details on these metrics).

Section 5.3

Calibration focused on the 09/2009–08/2015 period, when sediment monitoring protocols were enhanced, including higher sampling frequency at Requena between 11/2012 and 06/2013, where one sample was collected each two days plus three sampling repetitions each ten days. Beyond, sampling was conducted at five-day intervals during the wet period between July 2013 and September 2015. Additionally, the concentration gaugings were performed in all sites with a higher number of samples collected throughout the cross-section, particularly in the first half of the water column, to ensure more accurate sand concentration calculations. Outside this interval, uncertainties in sand flux observations increase, which complicates the definition of a robust and fully independent validation period (see Supplement S9 for a summary of model performance across periods). The lower performance outside the calibration period primarily reflects uncertainties in rainfall forcing and observations, particularly during rapid Andean flood events, rather than a degradation of model performance. To further evaluate model performance, direct comparisons between simulations and gauging measurements were performed (Fig. S9). These complementary evaluations support the temporal robustness of the model despite observational limitations and data heterogeneity.

Supporting analyses and details are now provided in the Supplementary Material (S9).

2. The daily timestep is justified qualitatively by slow, subcritical flow conditions, but I cannot find any formal numerical analysis; could the authors present Courant number ranges for representative reaches or demonstrate via timestep halving that peak attenuation and timing errors remain acceptably small?

We thank the reviewer for raising this important point regarding the numerical stability of the routing scheme. Although SWAT-Amazon provides model outputs at a daily time step, the routing of water and sediments is internally solved at a sub-daily time step determined dynamically according to the Courant–Friedrichs–Lewy (CFL) stability condition. Numerical stability requires that the computational time step remains smaller than the time needed for the fastest wave allowed by the governing equations to travel across a reach (Courant condition). Since reach lengths are fixed in the river network representation, the routing time step must therefore satisfy the CFL condition, which for a rectangular channel can be expressed as:

$$\Delta t \leq \alpha \, \Delta t_C = \alpha \, \frac{\Delta x}{\sqrt{g \, h}}, \qquad \text{Eq. RC1.1}$$

where Δt_C (s) is the flood-wave travel time, Δx (m) is the reach length, g (m s^{-2}) the gravitational acceleration, h (m) the flow depth, and $\alpha \leq 1$ a safety coefficient that can be parameterized for each reach in the SWAT-Amazon framework (default value = 0.5). Values in the range 0.2–0.7 are commonly recommended to ensure numerical stability (e.g., Bates et al., 2010; Pontes et al., 2017), and can be configured in SWAT-Amazon. For representative lowland reaches of the Ucayali River, the CFL-based time step evaluates to 3–14 h, yielding an internal routing time step of approximately 1.5–7 h with $\alpha = 0.5$. Consequently, the routing equations in SWAT-Amazon are computed in a loop at an intraday time step within the routing module, ensuring that the Courant condition is satisfied and preventing numerical instabilities. The simulated daily values correspond to the model state at the end of each simulation day, and wave propagation and attenuation are adequately resolved despite daily reporting of model outputs.

We added the following clarifications in the manuscript:

Section 4.1.3:

Although SWAT simulations are reported at a daily time step, river routing is internally solved at a sub-daily time step dynamically determined by the Courant–Friedrichs–Lewy (CFL) stability condition. The routing module automatically adjusts the internal time step to satisfy the CFL criterion and ensure numerical stability (Bates et al., 2010). Daily outputs therefore correspond to the model state at the end of each simulation day.

Section 4.2.5:

As for water routing, sand routing is internally solved at a sub-daily time step to satisfy the CFL stability condition.

3. The authors state that rectangular main channels and simple-reservoir floodplains were chosen after testing other configurations, but results from these tests are not shown; could the authors include a supplementary comparison of discharge, water level, and sand

Alternative channel geometries and floodplain parameterizations were tested during the preliminary development of the routing scheme using a prototype model implemented in R, as documented in Santini (2020, Chapter 5). These numerical experiments simulated the propagation of a characteristic piedmont hydrograph along a simplified 1500 km lowland river representative of the Ucayali system, without lateral inflows, in order to assess the respective effects of channel geometry, floodplain representation, and routing formulation. The main conclusions of these tests were as follows:

- Floodplain activation is the dominant control on flood-wave attenuation and travel time. When the floodplain is connected, rapid flood oscillations are strongly damped and the phase lag between upstream and downstream discharge can reach several weeks.
- Floodplain cross-section geometry (rectangular versus triangular) mainly affects the transition around bankfull conditions, with triangular sections producing smoother transitions. However, both geometries yield very similar large-scale propagation behaviour.
- Representing the floodplain as an explicit flowing compartment produces only minor differences compared with a simple storage-reservoir formulation, especially under the low floodplain velocities typical of the Ucayali lowlands. In practice, these small differences fall within the uncertainty associated with effective flow resistance and do not justify the additional model complexity and calibration burden.

Overall, these preliminary experiments showed that a rectangular main channel combined with a simple reservoir-type floodplain representation provides the most parsimonious and numerically robust configuration while preserving the first-order hydrodynamic behaviour relevant to this study, namely flood attenuation, phase lag, and realistic stage-discharge rating curves. Given that these tests are extensively documented in Santini (2020), we chose not to duplicate them in the Supplementary Material and have revised the manuscript accordingly.

The manuscript was revised as follows (Section 4.1):

The main channel's trapezoidal cross-section in SWAT was replaced with a rectangular one for consistency with the hydraulic equations used. Floodplains were represented using simplified rectangular or triangular cross-sections. Alternative geometries were tested during model development (Santini, 2020), and the selected configuration was retained as the most robust and parsimonious for large low-slope rivers.

4. Validation focuses on outlet hydrographs, rating curves, and sediment fluxes, yet floodplain processes drive the key findings; I am missing independent validation of internal floodplain behavior using, for example, SAR-derived inundation extent time series, floodplain lake stage records, or tracer-based residence time estimates.

We thank the reviewer for this important comment. We agree that an independent validation of internal floodplain dynamics would further strengthen the study. However, such observations remain extremely scarce at the basin scale in the Amazon, particularly for floodplain lake stages, residence times, or tracer-based constraints. Acquiring and integrating them at the spatial and temporal scales considered here would require dedicated observational efforts beyond the scope of the present study. We acknowledge that the absence of a dedicated independent validation of internal floodplain processes is the main unresolved limitation of the present study, and explicitly identify it as such in the revised Section 5.5. A fully independent validation of internal floodplain processes therefore remains an important direction for future work.

The objective of the present study is to constrain water and sediment fluxes along the main channel and to quantify long-term basin-scale mass balances. Accordingly, calibration and evaluation focus on elements that directly and robustly constrain the hydro-sedimentary dynamics of the system, namely discharge, water levels, velocities, rating curves, and sediment fluxes. Model performance is evaluated over independent time periods for discharge (Table RC1.1) and sand fluxes (Table RC1.2) and against extensive field measurements (Fig. RC1.1), providing strong constraints on the integrated behaviour of the river–floodplain system.

In the present framework, floodplains are represented using a simplified storage approach in which water levels are spatially aggregated at the reach scale. As discussed in the revised manuscript (Section 6.3.1), natural floodplains exhibit strong spatial heterogeneity (topographic gradients, secondary channels, lakes, varying hydraulic connectivity) such that local water levels may not directly reflect the large-scale storage dynamics represented in the model. Furthermore, existing satellite products lack the spatial and temporal resolution required to robustly constrain floodplain water levels and storage volumes across the diversity of water surface types present in Amazonian floodplains. For these reasons, the evaluation focuses on constraining water and sediment fluxes and storage volumes at the sub-basin scale through multi-source observations of main-channel dynamics, rather than on reproducing local floodplain water levels. In large lowland Amazon rivers, floodplain stages and storage are not independent from main-channel conditions: inconsistencies in floodplain exchanges would necessarily propagate to main-channel hydraulics through mass conservation, affecting discharge, water levels, velocities, and rating curves, which are all explicitly evaluated here.

Overall, model evaluation combines (i) discharge and sediment flux evaluation over independent time periods, (ii) comparison with an extensive in situ dataset of direct field measurements, and (iii) multiple hydraulic diagnostics constraining the coupled river–floodplain system.

The manuscript was revised as follows (Section 5.5):

The integration of remote sensing and hydrological modelling was validated at two super stations in the basin: Requena (Fig. 8) and Lagarto (not shown). Total sediment fluxes (Fig. 8c) were calculated by summing: (i) sand flux (Fig. 8a) from SWAT-Amazon simulations and (ii) fine sediment flux (Fig. 8b) from satellite data combined with SWAT-Amazon flows. The results align well with in situ flux measurements (daily NSE: 0.87 at Requena, 0.79 at Lagarto, monthly NSE: 0.87 at Requena, 0.86 at Lagarto), and suggests that both stations could be monitored in this way with a few calibration campaigns. This supports the definition of minimal observational requirements for transferring the method to other Amazonian basins (see Section 6.2.3). The evaluation focuses on variables that directly constrain main-channel hydrology and sediment transport, and thereby indirectly constrain floodplain processes. In the present framework, floodplains are represented using a simplified storage approach, in which water levels are spatially aggregated at the reach scale. However, natural floodplains exhibit strong spatial heterogeneity, including topographic gradients, secondary channels, lakes, and varying degrees of hydraulic connectivity. As a result, local water levels may not directly reflect the large-scale storage dynamics represented in the model. In addition, existing satellite products lack the spatial and temporal resolution required to robustly constrain floodplain water levels and associated storage volumes, particularly given the diversity of water surface types (e.g. secondary channels, connected lakes and isolated water bodies) and their distinct behaviours. Accordingly, local floodplain water levels are not explicitly reproduced; instead, the analysis focuses on constraining water and sediment fluxes and associated storage volumes at the subbasin scale using multi-source observations of main-channel dynamics. The increasing availability of satellite observations over time (Fig. 8b) further strengthens the observational constraints on these fluxes, with missing data mainly occurring at the beginning of the time series due to the progressive deployment of satellite sensors (Terra since 1999, Aqua since 2002, VIIRS since 2012).

5. The Sobol analysis identifies hf, Cnfp, kfp, and db as influential parameters, but does not explore whether multiple distinct parameter combinations yield equivalent outlet performance; could the authors present dotty plots or behavioral parameter ranges to assess equifinality and discuss how it affects the certainty of inferred trapping and recycling fractions?

This comment directly targets a key methodological gap in the original manuscript. We performed a dedicated equifinality analysis based on dotty plots (Fig. RC1.2) by sampling the main routing and sand transport parameters identified as influential in the Sobol analysis over wide physically plausible ranges, while keeping the other parameters fixed at their calibrated values. This analysis was used to assess whether distinct parameter combinations could produce similarly model performance, and to evaluate the implications for the robustness of the inferred floodplain trapping and recycling fractions.

The results show contrasted behaviours among parameters and reaches. The riverbed sand diameter d_b displays a clear optimum, consistent across reaches and close to the calibrated value, confirming that it is well constrained by the available observations. The water routing parameters (h_f, B, k_{fp}, C_{nfp}, n) all show identifiable optima to varying degrees depending on local hydraulic conditions, consistent with the Sobol analysis showing that all contribute to model performance. Since the key parameter controlling sediment budgets (d_b) is well constrained, and water routing parameters are collectively identifiable, the hydro-sedimentary budgets remain robustly constrained despite residual equifinality.

For d_b, the dotty plots also reveal secondary optimum(s) at unrealistically large values, which precludes sand suspension onset. This alternative mode(s) is not supported by field observations of bed-material grain size in the basin and was therefore excluded from the uncertainty analysis through physically informed priors, highlighting a known limitation of purely statistical calibration: without observational constraints, acceptable goodness-of-fit can be obtained for parameter values that are not physically meaningful.

For this reason, as suggested by the reviewer, the revised manuscript now propagates parameter uncertainty to the key budget indicators, so that the reported trapping and recycling fractions are presented together with uncertainty ranges rather than as single deterministic values.

New figures and details were added to the Supplementary Material (S11): dotty plots showing equifinality patterns and bimodality of d_b, with calibrated values indicated.

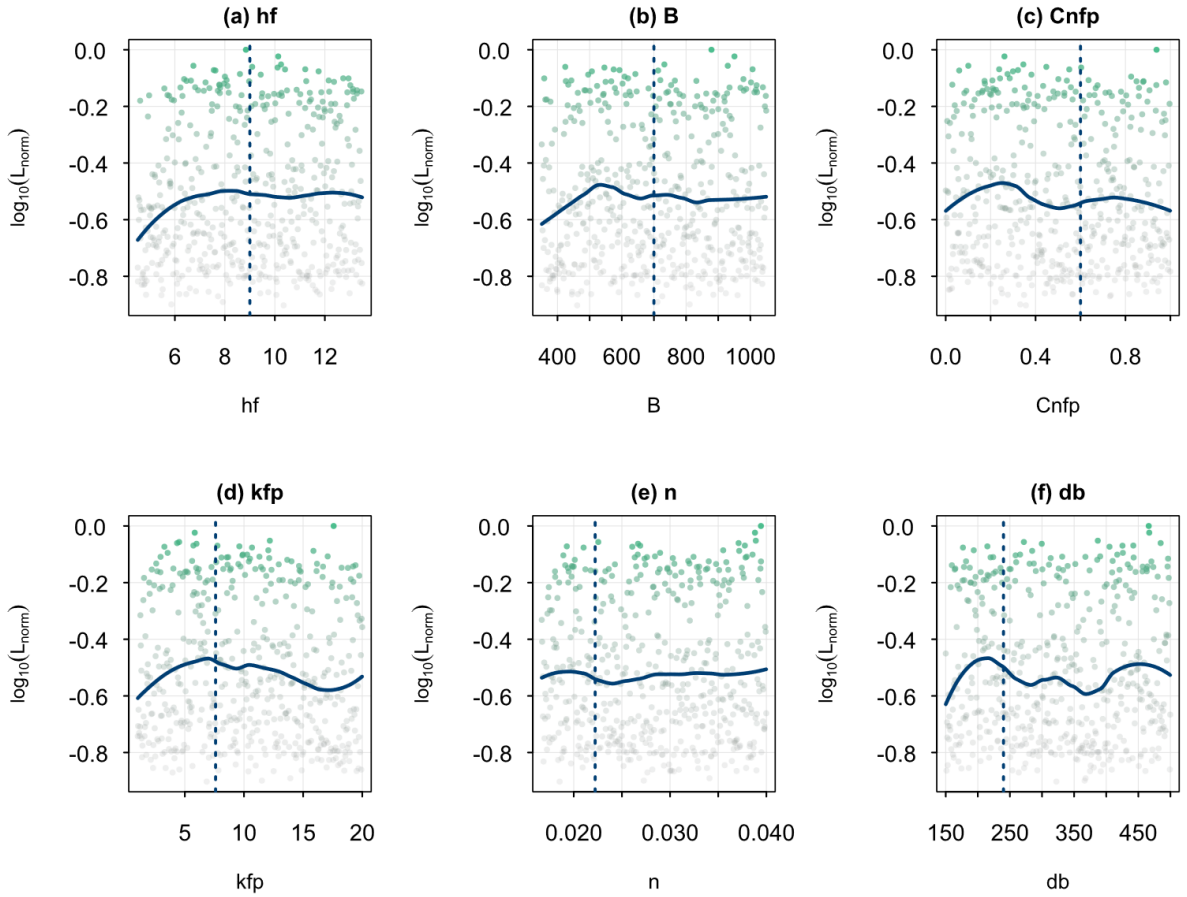

Figure RC1.3: Example of dotty plots for sub-basins 5, showing normalised GLUE likelihood versus parameter value for the Sobol-influential parameters. Green shading indicates higher likelihood. The blue curve is a loess smoother. Dashed blue vertical lines mark the calibrated values. Note the bimodal response of d_b, with a physical optimum near 240 µm and a spurious one around 400 µm.

The manuscript was revised as follows (new section):

6.1.3 Equifinality analysis

GLUE-based dotty plots were produced by sampling the Sobol-influential parameters over wide ranges (Supplement S11). The water routing parameters (h_f, B, k_{fp}, C_{nfp}, n) all show identifiable optima to varying degrees depending on the reach, consistent with the Sobol analysis. Critically, d_b exhibits a physical optimum near 220–250 µm matching observed grain sizes, and spurious coarser modes, precluding sand suspension. This bimodality highlights a well-known limitation of purely statistical calibration: without physical constraints, optimization can converge toward parameter values that are mathematically acceptable but physically meaningless. Overall, since d_b is well constrained and water routing parameters are collectively identifiable, the hydro-sedimentary budgets remain robustly constrained despite residual equifinality across all behavioural simulations (Supplement S12). These findings justify the use of expert measurement-informed priors for uncertainty propagation.

6. The reflectance–concentration relationship is calibrated on a limited number of samples and applied basin-wide for two decades; given the acknowledged potential for hysteresis and grain-size variability at the surface, I am wondering whether the authors could quantify the bias or uncertainty bounds under contrasting conditions such as rising versus falling limb or low-water resuspension events.

This is an important comment. We agree that potential hysteresis effects and grain-size variability at the water surface could introduce uncertainty in the reflectance–concentration relationship, particularly under contrasting hydrological conditions such as rising versus falling limbs or low-water resuspension events.

As noted in the manuscript, previous studies in the Amazon Basin have indeed reported hysteresis when relating single-band reflectance (e.g., NIR) to total suspended sediment concentration (Pinet, 2017). However, this effect is strongly reduced when using spectral band ratios such as NIR/RED, which are specifically designed to minimize variability related to illumination conditions and particle-size effects (Martinez et al., 2015; Pinet, 2017). In addition, field observations from several Amazonian white rivers indicate a relative stability of surface particle-size distributions across hydrological conditions (Martinez et al., 2015). This stability is generally attributed to hydrodynamic sorting processes that preferentially maintain the finest particles in suspension near the water surface, thereby limiting grain-size-induced variability in optical properties.

Nevertheless, Santini (2020) suggested that a residual hysteresis may persist, based on interannual variability in satellite-derived reflectance (MODIS, VIIRS). However, the magnitude of this variability was found to be of the same order as the intrinsic uncertainty of satellite reflectance measurements (MAE = 132 mg l^{1}) and therefore neglected in the study. Furthermore, the dataset used to validate the relationship (matchups between satellite-derived reflectance ratio and fine sediment concentrations monitored at gauging stations, averaged at a monthly time step) spans a wide range of hydrological conditions, including major floods and droughts and rising/falling limbs and doesn't show, at first order, any hysteresis beyond the level of intrinsic uncertainties. Finally, mass balances are computed at daily to yearly time scales, which further reduces the impact of short-term hysteresis effects through temporal integration. As a result, the influence of these processes on long-term sediment flux estimates is expected to remain limited relative to other sources of uncertainty and to the overall magnitude of sediment transport.

A dedicated quantification of uncertainty under contrasting hydrological conditions (e.g., rising vs. falling limbs) would nevertheless require targeted high-frequency radiometric field campaigns and is beyond the scope of the present study.

The manuscript was revised as follows (Section 6):

However, Pinet (2017) noted hysteresis in Madeira River relationships due to variations in grain diameter at the water surface, and Santini (2020) suggested that a fraction of fine sediments might also be sensitive to turbulence-induced lift variations. This may introduce limited variability in the relationship during specific hydrological phases, particularly resuspension events and low-water conditions. However, previous studies have shown that such effects are strongly reduced when using the spectral band ratio (NIR/RED) (Martinez et al., 2015; Pinet, 2017), with remaining variability comparable to satellite reflectance uncertainty. In addition, evaluation against satellite–in situ matchups across a wide range of hydrological conditions does not reveal any systematic hysteresis at the monthly time scale. As a result, the impact on long-term sediment flux estimates is expected to remain limited relative to other sources of uncertainty and to the overall magnitude of sediment transport.

---

7. The bimodal fine/sand assumption is conceptually justified, but I cannot find any sensitivity test of budget estimates to the threshold grain size separating the two classes; could the authors show how trapping and recycling percentages change if the cutoff is shifted?

We thank the reviewer for this relevant suggestion. In the present modelling framework, the fine–sand cutoff does not explicitly appear in the transport equations. Sand transport capacity is computed using the formulations of Camenen and Larson (2005, 2008), which rely on a single representative grain diameter (d_s) for the transported sand fraction. As a result, shifting the cutoff between fine and sand fractions does not introduce a new transport behaviour, but only modifies the representative diameter assigned to the sand class. The effect of changing the cutoff is therefore equivalent to perturbing d_s.

This effect was already explored through the Sobol sensitivity analysis, in which d_s was varied over a wide range (60–120 µm), and through the uncertainty propagation (see response to comment 11), which quantifies its impact on sand budgets. These analyses show that: (i) the model is moderately sensitive to d_s, although less than d_b, which is the most sensitive parameter for sand routing and transport; and, (ii) the resulting variations in sand budget indicators remain roughly on the order of ±10–20% when all routing and sand transport parameters are varied within plausible ranges. This indicates that the specific contribution of d_s alone (and therefore of the fine–sand cutoff) is limited to a few percent. Overall, its influence can be considered second order compared to the dominant controls related to floodplain processes.

A dedicated sensitivity test based on an explicit threshold shift was therefore not performed, as it would be fully redundant with the existing analyses. This point has not been added to the manuscript to avoid unnecessary lengthening.

8. The paper advocates extending the approach to other Amazonian basins but does not specify minimum data requirements; could the authors indicate the minimum gauging frequency at a super station, the required number of sediment samples for calibrating the reflectance model, and whether performance degrades gracefully as data density decreases?

We agree that assessing how the framework performs under reduced data availability would be highly informative to evaluate its transferability to other large tropical basins. A controlled data-degradation experiment at a super station, consisting in progressively subsampling discharge, sediment, PSD, and radiometric observations, would provide a rigorous way to quantify how model performance and inferred sediment budgets evolve as data density decreases. However, because the impact of reduced data availability cannot be evaluated independently from measurement uncertainties and data quality, such an experiment constitutes a substantial extension of the present work, which we consider as a perspective for future research.

These considerations place strong emphasis on the role of long-term observatories such as HyBAm, where sustained collaboration between researchers and operational hydrological services enables the production of high-quality, well-documented datasets supported by FAIR data principles and open data and open science practices, including DOI attribution, standardized metadata, and clear data governance frameworks ensuring traceability and reproducibility. In this context, close interaction between modelling and field measurements is crucial. Model calibration and interpretation rely on a detailed understanding of measurement protocols, data limitations, and associated uncertainties. Strengthening data governance practices (e.g. QA/QC procedures, metadata standardization, and interoperability frameworks such as the OGC SensorThings standard) and promoting knowledge transfer from research to operational technicians, as implemented in the present framework through its operational deployment within hydrological services, are therefore key elements to ensure the robustness and reproducibility of hydro-sediment budget estimates in large river basins.

Based on the long-term experience of the CZO HyBAm network, we can nonetheless provide the following practical guidelines. In large Amazonian rivers, stage–discharge relationships are generally stable at the decadal scale at mainstem stations (channel-controlled conditions). The main difficulty is therefore not frequent re-calibration of the rating curve, but rather documenting the full range of hydraulic conditions, especially in floodplain reaches where stage–discharge relations are often non-bijective because of backwater effects and floodplain storage. In practice, our experience suggests that a limited number of well-distributed gauging campaigns, typically on the order of 4–8, covering the main phases of the annual hydrograph (rising limb, high water near and beyond h_f, falling limb, and low water), is sufficient to constrain discharge, water level,

and velocity relationships at a virtual station. For a super station, 2–4 gauging by year are necessary.

For sediment observations, fine sediments are relatively well constrained through index concentration sampling at the water surface or through detailed sampling campaigns. In contrast, sand particles are more difficult to monitor due to their strong sensitivity to flow fluctuations, making it essential to complement observations with robust suspended sediment gaugings well distributed across the annual hydrograph and carried out using standardized and hydraulic-based protocols (Santini et al., 2019). The required effort also differs spatially: in the upper Amazon foreland, the main modes of the suspended and bed-material particle-size distribution remain relatively stable along a given reach, so that one to two detailed PSD sampling campaigns per site may be sufficient for a first-order characterization. Downstream of major confluences in the central and lower Amazon, stronger spatial heterogeneity driven by contrasting major tributary inputs (e.g. Madeira, Negro) requires a more extensive PSD characterization, on the order of four to five detailed campaigns per site distributed across the main hydrological conditions and across the river section.

Regarding the reflectance–concentration relationship for fine suspended sediments, previous work has shown that the use of the NIR/RED ratio substantially reduces sensitivity to varying optical conditions in Amazonian white-water rivers. On this basis, we consider that 2–3 radiometric calibration campaigns spanning contrasted hydrological conditions (low, intermediate, and high water) would provide a reasonable minimum dataset for calibrating the relationship at basin scale, provided that the sampled sites are representative of mainstem conditions.

Overall, performance is expected to degrade progressively rather than abruptly as data density decreases, as long as the calibration dataset still accurately captures the main hydrological stages of the annual cycle. The framework is relatively robust to moderate data scarcity, but its reliability decreases when key parts of the hydrograph or major sedimentary contrasts are not sampled. Satellite altimetry and reflectance data are essential complements, because they help compensate for sparse field observations and support extension from super stations to virtual stations. In this regard, the increasing availability of satellite data, in particular from the SWOT mission for water levels and from the growing constellation of optical sensors, is expected to make observational requirements less restrictive in the near future, especially if field campaigns are synchronized with satellite overpasses.

These developments call for a rethinking of monitoring strategies in large river basins, with a redeployment of human resources and budgets towards targeted calibration efforts and high-quality field measurements, a sustained effort to maintain super stations over the long term, and

strengthened collaboration between researchers, universities, and operational hydrological services.

The manuscript was revised as follows:

Section 5.5

The results align well with in situ flux measurements (daily NSE: 0.87 at Requena, 0.79 at Lagarto, monthly NSE: 0.87 at Requena, 0.86 at Lagarto), and suggests that both stations could be monitored in this way with a few calibration campaigns. This supports the definition of minimal observational requirements for transferring the method to other Amazonian basins (see Section 6.2.3).

Section 6.2

6.2.3 Minimum observational requirements

Based on the long-term monitoring experience of the CZO HyBAm network and the results obtained in this study, the framework can be applied to other Amazonian basins provided that a minimal observational dataset is available at a limited number of super stations and complementary virtual stations. For the last ones, this includes: (i) approximately four to eight discharge gauging campaigns covering the main stages of the annual hydrograph (rising limb, high water, falling limb, and low water), (ii) a comparable number of suspended sediment gaugings distributed along the hydrograph and carried out using standardized and hydraulic-based protocols (Santini et al., 2019), (iii) one to two detailed PSD surveys in relatively homogeneous upstream reaches, increasing to four to five in more heterogeneous downstream settings (e.g. below major confluences), and (iv) two to three radiometric calibration campaigns spanning contrasted hydrological conditions.

9. Channel geometry parameters are held constant over the full simulation period in a dynamically meandering system; could the authors test stationarity of the calibrated relationships using moving-window recalibration or change-point analysis to confirm that morphological evolution has not introduced time-varying bias?

We agree that freely meandering rivers are morphodynamically active and that lateral migration of meanders can locally modify channel geometry. Reported lateral migration rates in large Amazonian rivers commonly reach several tens of meters per year, with local maxima exceeding 100 m yr^{-1} in some reaches (Abad et al., 2012; 2013; Quintana-Cobo et al., 2015). If model calibration relied on a single, morphologically active cross-section, such changes could potentially affect the stability of stage–discharge relationships.

However, morphological adjustments were not explicitly simulated in the present modelling framework. In SWAT-Amazon, the channel geometry parameters represent effective, reach-averaged hydraulic geometry for very large Amazonian rivers rather than the instantaneous geometry of a single cross-section. Lateral migration in meandering rivers mainly redistributes erosion and deposition within the meander belt and does not necessarily translate into systematic changes in the effective hydraulic control at the scale of very large reaches over decadal time scales. Furthermore, the floodplain trapping and recycling fluxes estimated in this study implicitly integrate these lateral migration processes at the reach scale, which is fully consistent with the budget framework adopted here. Consistent with this reach-scale perspective, long-term observations from the CZO HyBAm network indicate that stage-discharge relationships at the mainstem stations remain very stable over time, with no significant rating shifts. This suggests that flow conditions are primarily controlled by reach-scale channel geometry, which supports the use of time-invariant hydraulic geometry parameters over the simulation period.

The manuscript was revised as follows (section 4.1):

The main channel's trapezoidal [...]. In this framework, channel geometry parameters represent effective hydraulic conditions averaged at the reach scale rather than the instantaneous geometry of a single cross-section. Although large Amazonian rivers are morphodynamically active and may exhibit lateral migration rates of several tens of meters per year, long-term observations from the HyBAm network indicate that stage-discharge relationships remain remarkably stable over time at the mainstem stations. This suggests that flow conditions are primarily controlled by reach-scale channel geometry over the decadal time scales considered here.

10. The MGB model is critiqued for underestimating Ucayali sediment loads, but no controlled comparison using equivalent forcing and calibration effort is provided; I would encourage the authors to clarify whether this difference reflects model structure or calibration intensity, or alternatively defer detailed benchmarking to future work while softening the current language.

The objective of this section was not to provide a controlled benchmarking between models but to place the present results in the context of previous large-scale modelling efforts. We agree that differences between model estimates may reflect both model structure and calibration strategies, and that a rigorous comparison would require applying the models under identical forcing and calibration frameworks, which is beyond the scope of this study.

However, the discussion highlights a potential structural limitation related to the representation of sand transport. In the MGB framework, sand transport is represented as bedload, whereas field observations in large Amazonian rivers indicate that sand can represent a substantial fraction of the suspended sediment load. This conceptual difference likely explains a large part of the discrepancy observed in sediment load estimates.

The manuscript was therefore revised to soften the wording and clarify that this interpretation is indicative rather than a formal model evaluation.

The manuscript was revised as follows (Section 6):

Previous large-scale efforts with the MGB (Modelo de Grandes Bacias) model (Fagundes et al., 2021, 2023), contributed significantly to understanding sediment transport across South America. However, challenges remain in representing sand transport, particularly its suspension dynamics. The MGB model assumes that sand transport is predominantly bedload, whereas field observations indicate that sand can account for a substantial fraction of the suspended sediment load in large Amazonian rivers, reaching up to ~70%. Rouse numbers in the range 0.2–0.8 further indicate transport in graded suspension rather than intermittent transport (Santini et al., 2019; Martinelli, 2022). This conceptual difference likely leads to underestimates of sediment load in the Ucayali Basin in previous MGB applications, with reported estimates being up to nearly three times lower than observations, although its contribution would need to be confirmed through a controlled inter-model comparison.

11. Key budget figures such as floodplain trapping, recycling contribution, and sand capture at peak discharge are presented without confidence intervals; I strongly encourage the authors to propagate observational uncertainty and parameter uncertainty through to these final percentages using Monte Carlo or ensemble methods.

We thank the reviewer for this important suggestion. We agree that providing uncertainty bounds for the main budget indicators significantly strengthens the robustness and interpretability of the results. In the revised manuscript, we implemented a GLUE (Generalized Likelihood Uncertainty Estimation) based uncertainty propagation framework (Beven and Binley, 1992) to quantify the impact of parameter uncertainty on the reported hydro-sedimentary budgets.

Scope of the uncertainty analysis

The analysis focused on parameters controlling the routing and sand transport modules developed in this study. These include (i) main channel and floodplain (h_f, C_{nfp}, k_{fp}, n, B) and (ii) sand transport and routing parameters (d_b, d_s, β_s, k_{bed}, C_{bk}, η). SWAT's default hydrologic parameters (rainfall–runoff and groundwater) were held fixed at their calibrated values to avoid sampling an excessively large parameter set, which would be beyond the scope of this study but represents a natural extension of the framework. The reported uncertainty bounds should therefore be interpreted as reflecting uncertainty in routing and suspended sand transport processes.

Parameter sampling strategy

Parameter uncertainty was represented using parameter-specific prior distributions (Table RC1.3). Measured or quasi-measured parameters (n, d_b, d_s) were assigned truncated normal distributions centred on calibrated values, with standard deviations consistent with field measurement uncertainty. Parameters well constrained by calibration against multiple hydraulic diagnostics (B, k_{fp}, C_{nfp}) were assigned narrow uniform ranges (±5–10%). Parameters controlling floodplain activation (h_f) and bed erosion (k_{bed}) were assigned absolute offsets (±0.50 m and ±0.20, respectively) reflecting their physical interpretation. Less identifiable parameters (β_s, η, C_{bk}) were assigned wider relative ranges (±15–20%) to reflect their lower calibration constraint. This strategy reflects the balance between parameter sensitivity and calibration constraint: strongly constrained parameters are associated with lower uncertainty, whereas weakly influential parameters are assigned wider ranges due to their limited identifiability.

The joint parameter space was sampled using Latin Hypercube Sampling (LHS) with 2500 realizations. Parameter deviations were applied independently for each reach, allowing spatially

heterogeneous parameter combinations. The magnitude of parameter uncertainty was assumed uniform across reaches, reflecting generic measurement and calibration uncertainty.

Table RC1.3. Prior distributions used for uncertainty propagation of SWAT-Amazon routing parameters.

Parameter	Unit	Distribution	Sampling method	Proposed prior / range	Basis for the proposed range
h_f	(m)	Uniform	LHS	[$h_{f,cal}$- 0.5; $h_{f,cal}$ +0.5]	An uncertainty of approximately ±0.50 m was inferred from rating curve analysis.
B	(m)	Uniform	LHS	[0.95 B_{cal}; 1.05 B_{cal}]	B is well constrained by the velocities and water levels (time series and direct measurements)
k_{fp}	(–)	Uniform	LHS	[0.9 $k_{fp,cal}$; 1.1 $k_{fp,cal}$]	k_{fp} is well constrained by the discharges, velocities and water levels.
n	(s m$^{-1/3}$)	Truncated normal	LHS	μ=n_{cal}, σ =0.0005 s m$^{-1/3}$ Truncated to [1/50;1/40]	In the lowland plain, calibrated values were consistently close to 1/45 s·m$^{-1/3}$. With σ = 0.0005 s·m$^{-1/3}$, 95% of sampled values fall within ±0.001 of the mean, consistent with the narrow range of calibrated values across reaches.
C_{nfp}	(–)	Uniform	LHS	[$C_{nfp,cal}$ - 0.1; $C_{nfp,cal}$ + 0.1]	C_{nfp} is strongly constrained by calibration against water levels and velocities, and identified as highly influential in the Sobol analysis.
d_s	(μm)	Truncated normal	LHS	μ=$d_{s,cal}$ σ = 3 Truncated to [0.85μ, 1.15μ]	d_s is measurement-informed. The manuscript reports typical suspended-sand diameters around 80–120 μm, with calibrated values near 80 μm and about 98 μm at Lagarto.
d_b	(μm)	Truncated normal	LHS	μ=$d_{b,cal}$ σ = 3 Truncated to [0.85μ, 1.15μ]	d_b is also measurement-informed. Calibrated values (220–252 μm) match observations, which supports a Gaussian prior. d_b is strongly constrained by calibration (most sensitive parameter for Qs).
β_s	(–)	Uniform	LHS	[0.85 $\beta_{s,cal}$; 1.15 $\beta_{s,cal}$]	β_s is weakly constrained by calibration against sand flux and was identified as having limited influence in the Sobol analysis. A wider ±15% range was therefore adopted to reflect its lower identifiability.
K_{bed}	(–)	Uniform	LHS	[$K_{bed,cal}$ – 0.2; $K_{bed,cal}$ + 0.2]	K_{bed} is weakly constrained by calibration against sand flux and was identified as having limited influence in the Sobol analysis. A wider range was therefore adopted to reflect its lower identifiability.
C_{bk}	(t m^{-3})	Uniform	LHS	[0.8 $C_{bk,cal}$; 1.2 $C_{bk,cal}$]	C_{bk} is poorly constrained and represents an effective source term rather than a directly measurable quantity. A uniform prior was therefore adopted.
η	(–)	Uniform	LHS	[0.85 η_{cal}; 1.15 η_{cal}]	η is moderately constrained by gauging measurements during flood conditions.

Behavioural simulations

Simulations were selected based on their ability to reproduce multiple observed variables at the Ucayali outlet (Requena), including both time series and discrete measurements of discharge, water level, flow velocity, and suspended sand flux (Table RC1.4). Selection was performed using a GLUE likelihood function applied to log-transformed residuals. Variable-specific weights were assigned according to data quality and physical relevance, with discharge and direct gauging measurements providing the strongest constraints. For comparison and validation of the GLUE-weighted approach, a threshold-based (binary) selection using standard performance criteria (NSE, KGE, PBIAS) was also applied, but was not retained for the final budget estimates (Fig. RC1.3).

Table RC1.4. Relative weights assigned to each variable in the GLUE-based uncertainty analysis. Data types distinguish between continuous time series (derived, measured, or satellite-based) and direct gauging measurements. Data quality is indicated qualitatively (+++ high, ++ moderate, + lower). Raw weights represent relative contributions and are normalized within the likelihood function.

Subbasin	Variable	Data type	Available?	Quality	Weight	Justification
21	Q	Time Series (derived)	Yes	+++	0.20	Long continuous record, robust rating curve; primary hydraulic constraint
21	h	Time Series (measured)	Yes	+++	0.10	Directly measured but largely redundant with Q
21	u	Time Series (derived)	Yes	+++	0.10	Secondary hydraulic constraint; correlated with Q
21	Q_s	Time Series (derived)	Yes	+	0.15	Derived from surface concentration monitoring and gauging curve extrapolation; high uncertainty
21	Q	Gauging	Yes	+++	0.20	Direct ADCP measurement; highest individual weight as primary discharge constraint
21	h	Gauging	Yes	+++	0.10	Directly measured at same campaign; weighted to jointly constrain hydraulic conditions determining sand transport capacity
21	u	Gauging	Yes	+++	0.10	Directly measured at same campaign; weighted to jointly constrain hydraulic conditions determining sand transport capacity
21	Q_s	Gauging	Yes	++	0.15	Direct sediment measurement; key budget constraint, moderate weight reflecting inherent sampling uncertainty

Multi-site consistency constraint

To reduce equifinality and ensure spatial coherence, a multi-site consistency filter was applied using all station with data (super, low data and virtual stations). Simulations inconsistent with observations at any site were discarded. This step removed approximately 20% of the behavioural simulations (Fig. RC1.3).

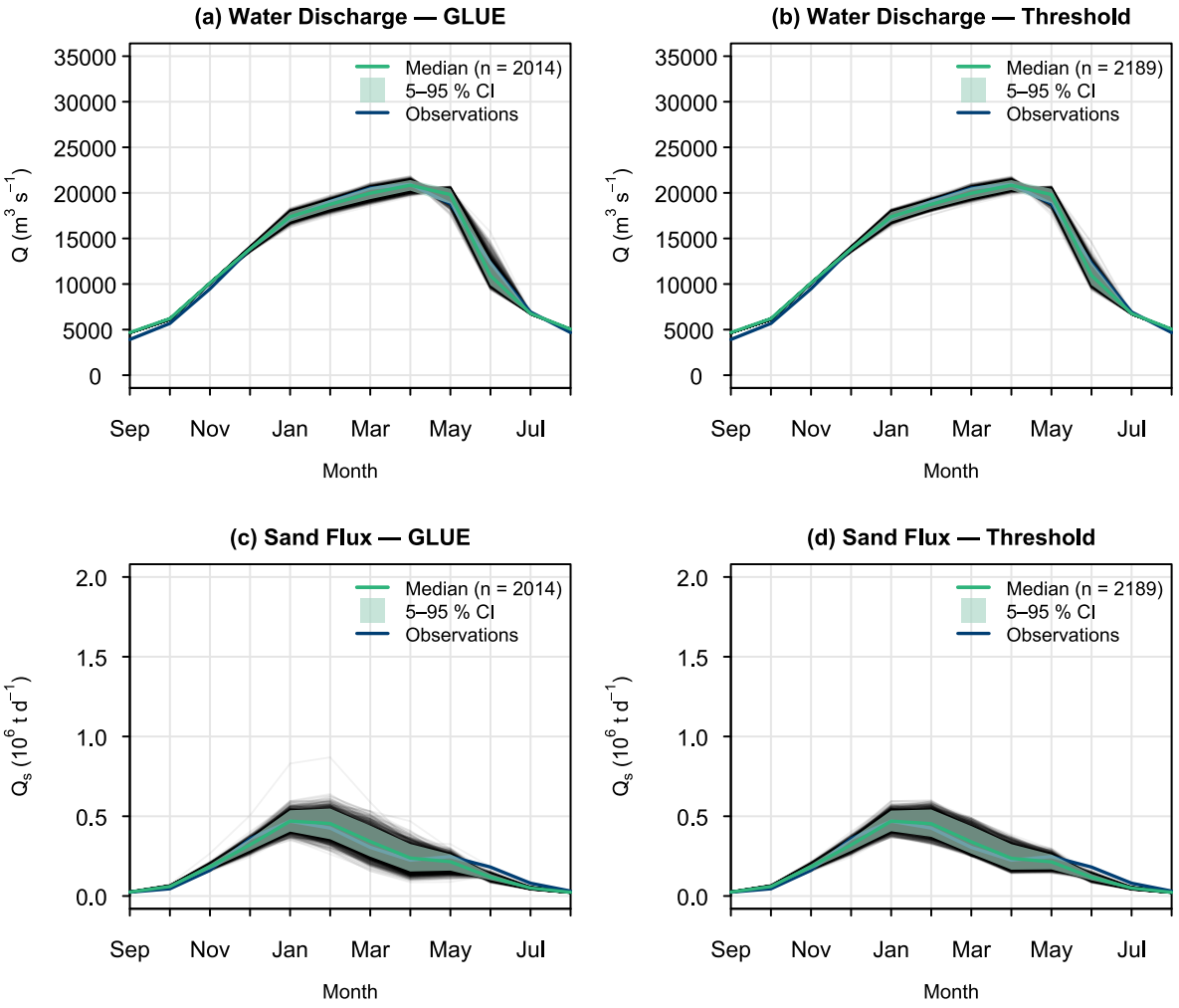

Figure RC1.3: Interannual cycles of water discharge and sand flux at the Ucayali outlet (subbasin 21, Requena station), comparing GLUE and threshold-based uncertainty propagation methods. (a, b) Water discharge Q (m^3 s^{-1}) and (c, d) suspended sand flux Qs (10^6 t d^{-1}), computed over the September–August hydrological year for the calibration period. Left panels (a, c) show results from the GLUE method with likelihood-weighted quantiles; right panels (b, d) show results from the threshold selection with uniform weights. In each panel, the green line represents the weighted median simulation, the green shaded area indicates the 5–95 % confidence interval, grey lines show individual behavioural runs, and dark blue circles with connecting line represent observed values derived from conventional monitoring.

Propagation to mass balances

For each retained simulation, the main budget metrics (sand trapping, floodplain recycling, resuspension, and sand capture at peak discharge) were recomputed. The resulting ensemble was used to derive weighted 5–95% confidence intervals for all reported budget figures (Fig. RC1.4). Observational uncertainty is implicitly accounted for within the GLUE framework, as the likelihood function allows for deviations between simulations and observations that are consistent with measurement uncertainty. Simulations falling within these error margins are retained as behavioural, so that the resulting uncertainty bounds on budget indicators reflect the influence of observation quality on parameter identifiability. Routing parameters were also perturbed in Andean reaches to evaluate the contribution of upstream uncertainty propagation to the

variability of budget indicators. The results show that this contribution is negligible in reaches without active floodplains. Therefore, the dominant source of uncertainty in the routing and in the reported mass balances arises from floodplain activation processes and sand transport dynamics in lowland reaches.

Overall, this approach provides a first-order estimate of uncertainty associated with water and sediment routing processes. A full propagation including hydrological parameters, structural uncertainty, and remote sensing errors remains an important perspective for future work.

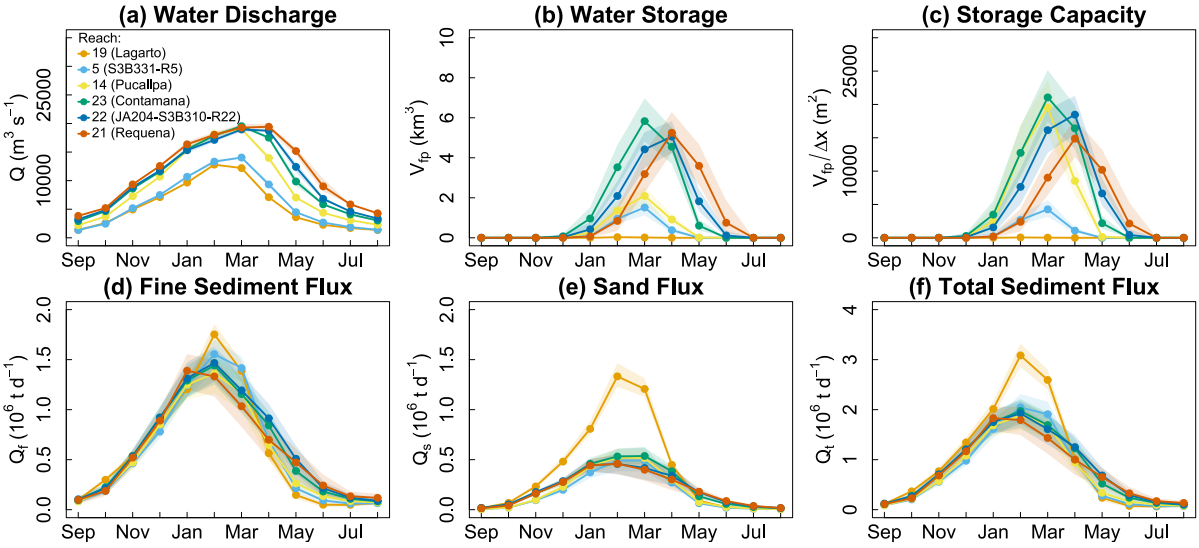

Figure RC1.4: Interannual Figure 11: Interannual water (1983¬2019) and sediment (2000¬2019) balances along the Ucayali River. Lines show interannual monthly means and shaded bands represent the 5th–95th percentile envelope of the behavioural GLUE ensemble (2500 runs). (a) Water Discharge (b) Floodplain water storage ($\boldsymbol{V_{fp}}$) in km3. (c) Normalization of $\boldsymbol{V_{fp}}$ by the reach's length ($\boldsymbol{\Delta x}$) for cross-sub-basin comparison. (d) Fines suspended sediments. (e) Suspended sand fraction. (f) Total suspended sediment load. See Supplementary Material (Section S12) for full uncertainty propagation details.

Mean uncertainty on V_{fp} (~±20%) are much wider than those on Q (~±4%) or h (~±6%), reflecting the structural non-linearity of the floodplain activation threshold: small perturbations of h_f generate disproportionately large changes in V_{fp} when $h \approx h_f$. The narrow envelope on Q_s at reach 19 (Lagarto) reflects limited floodplain activation ($h < h_f$) throughout most of the hydrological cycle, which precludes h_f, the most influential parameter, from affecting sand transport at this reach. Uncertainty envelopes on Q remain narrow by construction, as the LHS sampling targets routing and sand transport parameters only, conditional on calibrated SWAT rainfall-runoff forcing; they therefore represent a lower bound on total predictive uncertainty.

The manuscript was revised accordingly. The main changes are summarized below.

Abstract:
The uncertainty ranges (absolute) were added to all key budget.

Section 6.3 was restructured into two subsections (6.3.1 and 6.3.2) to clearly separate structural model limitations from quantified uncertainty.

6.3 Framework limitations and uncertainty

6.3.1 Model structural limitations

6.3.2 Framework uncertainty

A GLUE-based uncertainty analysis was conducted over a Latin Hypercube ensemble of 2500 simulations (Supplement S12), scoped to routing and sand transport parameters, conditional on the hydrological forcing, with physically informed priors (Section 6.1).

For discharge and sand flux, GLUE-weighted envelopes (5–95th percentiles) remain narrow relative to the interannual amplitude, indicating that parameter uncertainty propagates weakly compared to the magnitude of hydrological variability and that the dominant dynamics are robustly constrained. Envelope width depends jointly on data availability and measurement quality, which act as coupled controls on parameter identifiability rather than independent sources of uncertainty. Floodplain water storage uncertainty in particular is strongly controlled by h_f*, emphasizing the need for accurate stage–discharge rating curves to constrain this parameter reliably (Section 6.2.3). Importantly, because the budget indicators (trapping and recycling fractions) depend primarily on* h_f*,* B*, and* d_b*, which are well-constrained parameters (Section 6.1.3), they remain robust across the behavioural ensemble despite the wider prior ranges assigned to less identifiable floodplain parameters. This structural stability is confirmed by the GLUE-propagated budget envelopes (Supplements S11, S12).*

Fine sediment fluxes achieve low bias and good performance metrics (Fig. 8b), yet exhibit wider uncertainty envelopes. This apparent performance is largely driven by the dominant seasonal signal rather than true predictive skill. Satellite retrieval errors cancel over long-term integration, masking short-term variability: the method reliably captures the mean seasonal cycle, but not its anomalies. Satellite radiometry is therefore robust at the climatological scale but not at the event scale.

Both conclusions rest on a common foundation: physically meaningful uncertainty quantification requires prior knowledge, which itself requires long-term, multi-variable observations, as provided

by networks such as the CZO HyBAm. Uncertainty is therefore not only a modelling issue, but fundamentally a data-structure constraint. Model calibration and interpretation depend on sustained fieldwork and on a detailed understanding of measurement protocols and data limitations. As a result, rigorous data governance, FAIR practices, and strong research–operational links are essential for reproducibility and long-term continuity.

Section 6.4.1: budgets:

All trapping and recycling fractions now include GLUE-based uncertainty intervals (absolute).

Section 7: Conclusion:

Key budgets updated with (absolute) uncertainty ranges.

Supplementary Material S12 (new section):

Full description of the GLUE uncertainty framework: LHS sampling (n=2500), parameter priors (Table S12.1), behavioural selection criteria, multi-site consistency filter, and propagation methodology for budget indicators including lateral inputs. The Table RC1.4 was also added.

Table RC1.4. Summary of hydro-sedimentary budget indicators with GLUE-based uncertainty bounds.

Compartment / Reach	Indicator	Central	IC	Unit	Rel. IC
Basin inlet	Q_t Andean input	455	[410, 500]	10^6 t yr^{-1}	±10%
21 (Requena)	Q_t basin export	290	[235, 345]	10^6 t yr^{-1}	±19%
Lagarto – Pucallpa	ΔQ_s	−76	[−89, −63]	10^6 t yr^{-1}	±17%
Lagarto – Pucallpa	ΔQ_f	−51	[−67, −35]	10^6 t yr^{-1}	±31%
Lagarto – Pucallpa	ΔQ_t	−127	[−130, −124]	10^6 t yr^{-1}	—
Pucallpa – Contaya Arch	ΔQ_s	+4	[+1, +7]	10^6 t yr^{-1}	—
Pucallpa – Contaya Arch	ΔQ_f	+10	[+2, +17]	10^6 t yr^{-1}	—
Pucallpa – Contaya Arch	ΔQ_t	+14	[+3, +24]	10^6 t yr^{-1}	—
Contaya Arch – Requena	ΔQ_s	−11	[−13, −9]	10^6 t yr^{-1}	±17%
Contaya Arch – Requena	ΔQ_f	−16	[−23, −10]	10^6 t yr^{-1}	—
Contaya Arch – Requena	ΔQ_t	−27	[−36, −19]	10^6 t yr^{-1}	±31%
Contaya Arch – Requena	Sand capture (peak)	14%	[10%, 20%]	%	—
Basin	V_{fp} peak (Mar)	19.1	[15.3, 22.9]	km³	±20%
Lagarto – Pucallpa	V_{fp} peak (Mar)	1.5	[1.1, 1.9]	km³	—
Pucallpa – Contaya Arch	V_{fp} peak (Mar)	10.4	[8.3, 12.5]	km³	—
Contaya Arch – Requena	V_{fp} peak (Apr)	9.4	[7.7, 11.1]	km³	—

12. Several figures use cyan for SWAT-Amazon simulations and marine blue for observations, which are difficult to distinguish; some figures also use slightly different shades of cyan for multiple plots. I suggest that the authors consider using higher-contrast colour pairs or different line styles to improve readability.

We thank the reviewer for this helpful suggestion. We have revised the color scheme throughout the manuscript to improve contrast and readability.

The figures were redrawn as follows:

Water discharge observations are now shown in dark blue (#003E73), SWAT-Amazon simulations in green (#2FB47C), and the default SWAT Muskingum simulations in orange (#E69F00). These colors provide higher contrast and are used consistently across all figures.

Captions were also modified accordingly.

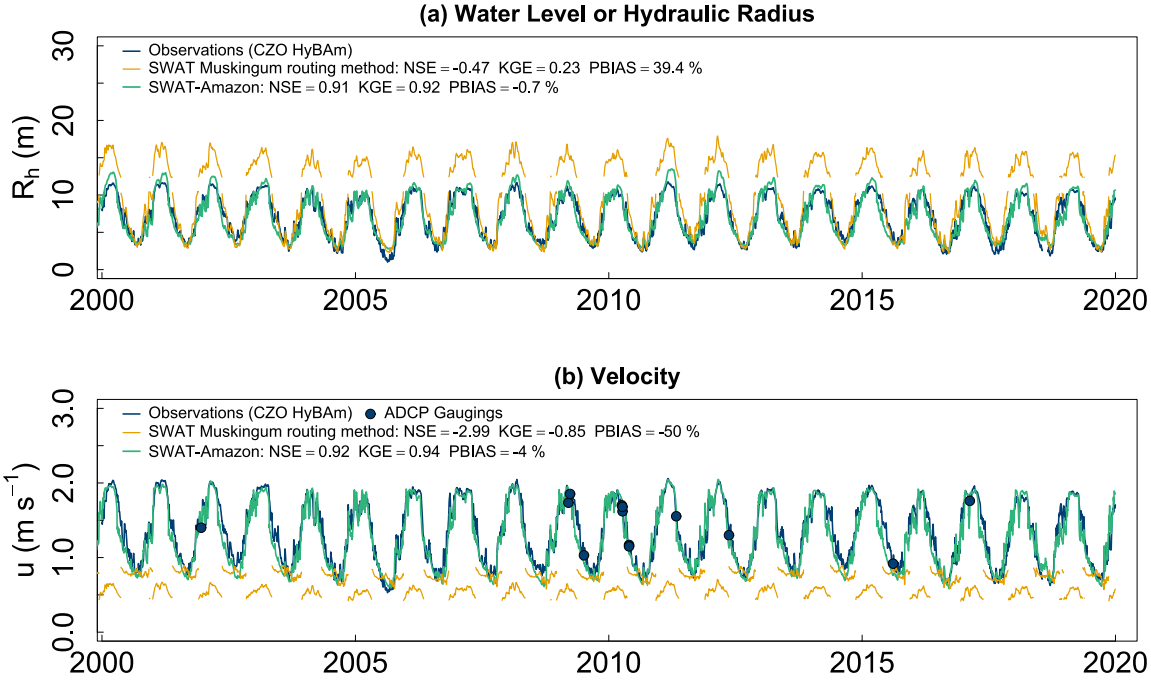

Figure RC1.5: Figure 4, modified with the new color palette.

Additional references

Abad, J. D., Montoro, H., Frias, C., Paredes, J., and Peralta, B.: The meandering Ucayali River, a cyclic adaptation of cutoff and planform migration, River Flow 2012, 141–156, 2012.

Abad, J. D., Vizcarra, J., Paredes, J., Montoro, H., Frias, C., and Holguin, C.: Morphodynamics of the upper Peruvian Amazonian rivers, implications into fluvial transportation, Congreso Internacional IDS2013 – Amazonía, Iquitos, Perú, http://hdl.handle.net/10077/8822 (last access: 30 March 2026), 2013.

Bates, P. D., Horritt, M. S., and Fewtrell, T. J.: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling, Journal of Hydrology, 387, 33–45, https://doi.org/10.1016/j.jhydrol.2010.03.027, 2010.

Beven, K. J. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrological Processes, 6, 279–298, https://doi.org/10.1002/hyp.3360060305, 1992.

Quintana-Cobo, I., Moreira-Turcq, P., Cordeiro, R. C., Aniceto, K., Crave, A., Fraizy, P., Moreira, L. S., Duarte, M. A., and Turcq, B.: Dynamics of floodplain lakes in the Upper Amazon Basin during the late Holocene, Comptes Rendus Géoscience, 350, 55–64, https://doi.org/10.1016/j.crte.2017.10.004, 2018.

Marinho, T., Filizola, N., Martinez, J. M., and Armijos, E.: Suspended sediment variability at the Solimões and Negro confluence between May 2013 and February 2014, Geosciences, 8, 265, https://doi.org/10.3390/geosciences8070265, 2018.

Response to Referee Comment #2 (RC2)

General Comments

The manuscript proposes a novel framework integrating remote sensing data with hydrological and hydraulic modelling to monitor hydro-sedimentary fluxes in the Amazon River.

The manuscript provides a clear advancement in hydro-sedimentary modelling of large rivers and it could be considered for publication on HESS after clarifications and revisions. I have listed below some points asking the Authors to address them to improve the current manuscript.

Major comments

1. Line 123/132. The Authors have considered two groups of sediments: fine sediments (10-20 μm, fine and medium silt) and sands (80-120 μm, very fine sands). What is the potential effect in model accuracy of having neglected the material with an intermediate diameter?

We thank the reviewer for this relevant comment.

In the present framework, the two sediment classes should be understood as a simplified representation of a continuous particle-size distribution that is predominantly bimodal in Amazonian white rivers, with a fine mode (silts and clays) and a very fine sand mode (Fig RC2.1). The reported diameter ranges (10–20 µm and 80–120 µm) therefore correspond to mean values of these dominant groups rather than strict granulometric bounds.

As a result, intermediate particle sizes (e.g. coarse silts in the 20–63 µm range) are not neglected, but are implicitly included within these two fractions depending on their physical behaviour. Coarse silts may either behave as wash-load when transported as aggregates or cohesive particles, or close to graded suspension when non-cohesive. Given the typical hydraulic conditions of large Amazonian rivers, their Rouse numbers remain low, so that they predominantly behave as fully suspended material and contribute only marginally to uncertainty arising from the two-fraction representation.

In the transport capacity equations, this representation relies on a single mean diameter for the sand fraction. Therefore, accounting for intermediate particle sizes is simply equivalent to perturbing this representative diameter. This effect has already been explored through the Sobol sensitivity analysis and uncertainty propagation, which show that its influence on sediment budget indicators remains limited (on the order of a few percent to ~10% when combined with other parameters). Overall, the omission of an explicit intermediate class is not expected to significantly affect model accuracy, while avoiding an unnecessary increase in model complexity and parameter uncertainty.

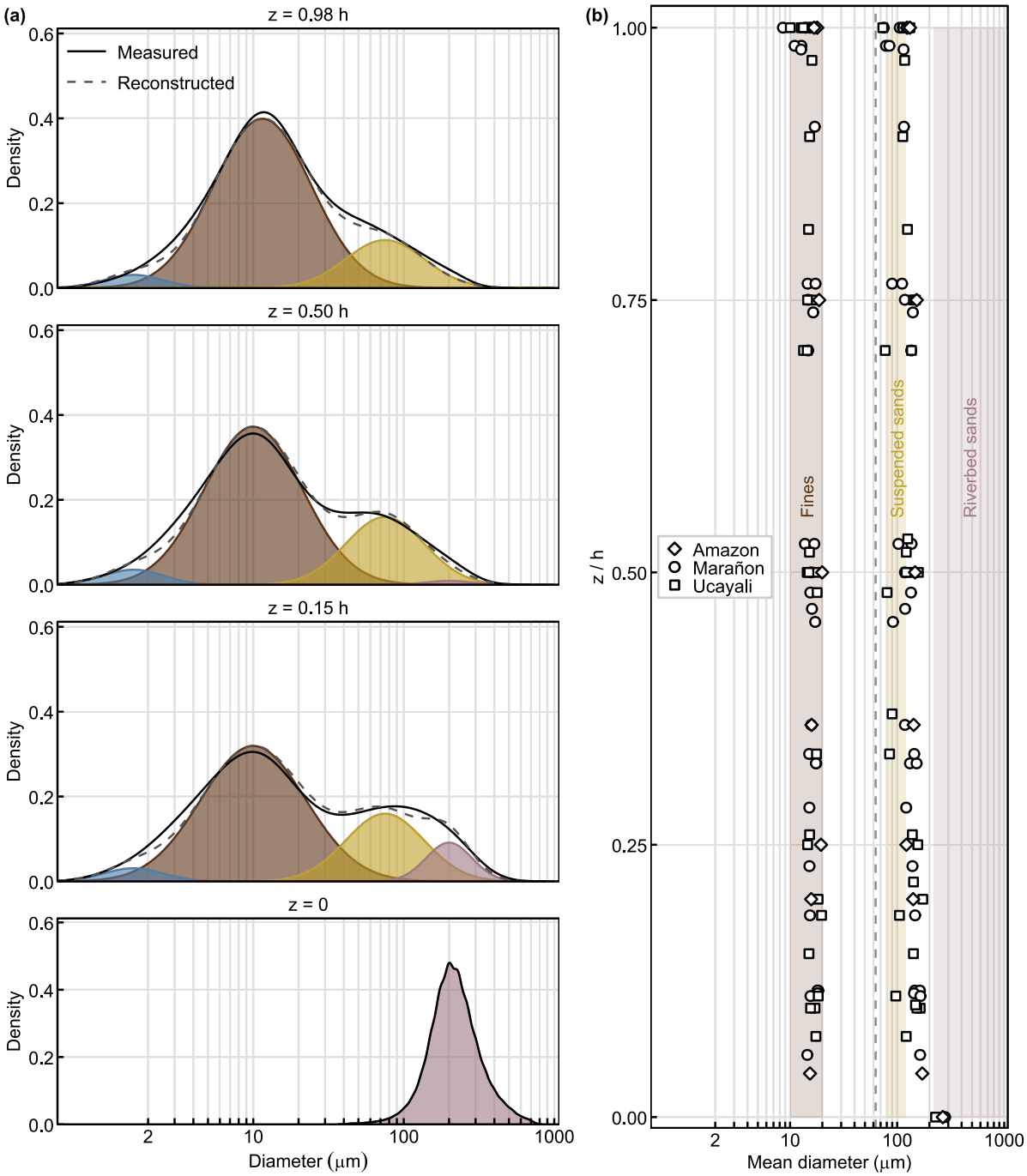

Figure RC2.1: Typical PSD of large white-water rivers in the Amazon Basin (adapted from Santini et al., 2019). (a) Particle-size distributions (PSD) measured in suspension at different relative depths and in the riverbed (Requena station, Ucayali River, 16 March 2015). The grey line shows the measured distribution, while the dashed line represents a reconstructed PSD obtained from a mixture of log-normal components corresponding to clay, silt, and sand fractions. (b) Vertical profiles of representative grain diameters for the fine and sand fractions. Symbols correspond to measurements at different stations. The dashed vertical line indicates the conventional 63 μm boundary between silt and sand.

We have added the following paragraph in Section 2 to clarify this point:

These ranges of mean diameter should be understood as representative values of the dominant PSD modes, rather than strict grain size distribution bounds. Intermediate particle sizes (e.g. coarse silts in the 20–63 µm range) are included within the fine sediment fraction defined by the 63 µm threshold used in the HyBAm monitoring protocol (see Section 3.1). Depending on their physical properties, coarse silts may either behave as wash load (e.g. aggregates) or exhibit transport dynamics closer to those of very fine sands when non-cohesive. This supports a first-order bimodal representation of suspended sediment transport at the basin scale, consistent with PSD measurements and vertical grain-size profiles presented in the Supplementary Material (Fig. S1).

Additional details have been added in the Supplementary Material.

---

2. Line 172. The Authors mentioned "two radiometric calibration-validation campaigns" for fine sediment concentration. My understanding is that the first campaign (November 2011) was for calibration and the second campaign (February 2017) was for validation. I suggest to better clarify this in the text. Furthermore, the Authors should short discuss about the extendibility to the Ucayali Basin of the data collected in few locations.

We thank the reviewer for pointing out this ambiguity. The wording in the manuscript was indeed misleading. The two campaigns were not used as independent calibration and validation datasets; rather, both campaigns contributed to the establishment of the radiometric relationship between reflectance and fine suspended sediment concentration. The text has been clarified accordingly.

We also added a short clarification regarding the applicability of the relationship across the Ucayali Basin.

The manuscript was revised as follows (section 3.3.2):

Two radiometric calibration-validation campaigns […] -> Two radiometric field campaigns […]

[…]

Relying on this dataset, a unique model for all the Ucayali Basin was fitted between fine sediment concentration at the water surface and the ratio of radiometer reflectance in the NIR (841–876 nm, according to the satellite sensor bands) and red bands (620–670 nm) (see Section 5.4). This single relationship is considered applicable along the mainstem of the Ucayali River in the lowland plain because surface reflectance is largely controlled by fine silts of Andean origin, whose grain-size characteristics and optical properties are relatively homogeneous along the river continuum

(Martinez et al., 2015; Santini, 2020). In addition, the use of spectral band ratios (NIR/RED), commonly applied to reduce the influence of these potential variations in optical conditions (Doxaran et al., 2002; Martinez et al., 2015; Pinet, 2017), supports the applicability of this relationship at the basin scale considered in this study.

Figure 7 was also modified:

Calibration – validation with field radiometers -> Calibration with field radiometers

---

3. Section 3.3.3. The Authors should provide more details about this important point. Have any field measurements been made? In which sections of the basin? Have the cross-sectional variations in sediment concentration along the vertical been considered?

We thank the reviewer for this comment. Field measurements of vertical concentration profiles of fine and sand suspended sediments have been collected at several stations of the HyBAm observatory across the Amazon Basin, including the Ucayali River. In particular, 73 field measurements were conducted at Requena (44), Lagarto (28), and Pucallpa (1) using a point-sampling method to characterize vertical concentration gradients. Details on the HyBAm dataset and methodology are provided in Santini et al. (2019) and Santini (2020) and in the Supplement S2. The estimation of α_f relies on the models proposed by Santini et al. (2019), which use Rouse-type formulations constrained by these observed concentration profiles to describe the vertical distribution of suspended sediments as a function of hydraulic conditions. These models were applied using hydraulic variables simulated by SWAT-Amazon to estimate α_f along the river network. Further details on the formulation and validation of the approach are provided in Santini et al. (2019).

The manuscript was revised as follows (section 3.3.3):

Due to the considerable depth of Amazonian rivers and the vertical sediment concentration gradient near the surface, the ratio α_f, relating the channel mean concentration to the surface index concentration retrieved by satellite, ranges from 1 to 1.8 according to the CZO HyBAm database (1.1 to 1.2 in the Ucayali). ***These values are derived from field measurements of vertical suspended sediment concentration profiles collected at several stations of the HyBAm observatory across the Amazon Basin, including the Ucayali River (e.g. Santini et al., 2019; Santini, 2020). To estimate α_f along the river network, the models proposed by Santini et al. (2019) were applied. These models use Rouse-type formulations constrained by observed concentration profiles to describe the vertical distribution of suspended sediments as a function of hydraulic conditions. They were parameterized using hydraulic variables simulated by SWAT-Amazon.***

Additional details have been added in the Supplementary Material (S2).

---

4. Section 4.1. It is unclear if (and how) morphological adjustements of channel cross-section over the simulation period were considered.

We agree that freely meandering rivers are morphodynamically active and that lateral migration of meanders can locally modify channel geometry. Reported lateral migration rates in large Amazonian rivers commonly reach several tens of meters per year, with local maxima exceeding 100 m yr^{-1} in some reaches (Abad et al., 2012; 2013; Quintana-Cobo et al., 2015). If model calibration relied on a single, morphologically active cross-section, such changes could potentially affect the stability of stage–discharge relationships.

However, morphological adjustments were not explicitly simulated in the present modelling framework. In SWAT-Amazon, the channel geometry parameters represent effective, reach-averaged hydraulic geometry for very large Amazonian rivers rather than the instantaneous geometry of a single cross-section. Lateral migration in meandering rivers mainly redistributes erosion and deposition within the meander belt and does not necessarily translate into systematic changes in the effective hydraulic control at the scale of very large reaches over decadal time scales. Furthermore, the floodplain trapping and recycling fluxes estimated in this study implicitly integrate these lateral migration processes at the reach scale, which is fully consistent with the budget framework adopted here. Consistent with this reach-scale perspective, long-term observations from the CZO HyBAm network indicate that stage-discharge relationships at the mainstem stations remain very stable over time, with no significant rating shifts. This suggests that flow conditions are primarily controlled by reach-scale channel geometry, which supports the use of time-invariant hydraulic geometry parameters over the simulation period.

The manuscript was revised as follows (section 4.1):

The main channel's trapezoidal [...]. In this framework, channel geometry parameters represent effective hydraulic conditions averaged at the reach scale rather than the instantaneous geometry of a single cross-section. Although large Amazonian rivers are morphodynamically active and may exhibit lateral migration rates of several tens of meters per year, long-term observations from the HyBAm network indicate that stage-discharge relationships remain remarkably stable over time at the mainstem stations. This suggests that flow conditions are primarily controlled by reach-scale channel geometry over the decadal time scales considered here.

5. Line 343. I would ask the Authors to shortly discuss how largely Eq.(28) is modifying Manning coefficient using the correction factor.

Eq. (28) increases flow resistance as a function of the relative water level $Y = (h - h_f)/h$ (–), which represents the stage at which water exchanges between the main channel and the floodplain begin to influence flow velocity. The parameter C_{nfp} (–) is a calibration coefficient controlling the strength of this effect when floodplain interactions occur ($Y > 0$).

In the calibrated model, C_{nfp} values along the mainstem range between 0.3 and 1. For typical flood conditions, Y varies approximately between 0 and 0.3 with a maximum value of about 0.5 during extreme floods. These extreme values ($Y = 0.5$) correspond to situations where simulated peak stages largely exceed observed values.

Under typical flood conditions, the correction remains limited to about 0–6% ($0.94 \leq \zeta_n \leq 1$) **(Fig. RC2.2)**. Larger corrections, reaching up to about 20%, occur only in rare situations where simulated peak stages substantially exceed observations. This range of values for the flow-resistance correction factor ζ_n is consistent with values reported in the literature for compound channel hydraulics and floodplain–channel momentum exchange (e.g. Nicollet and Uan, 1972; Smart, 1992; Shiono and Knight, 1996; Bousmar and Zech, 1999).

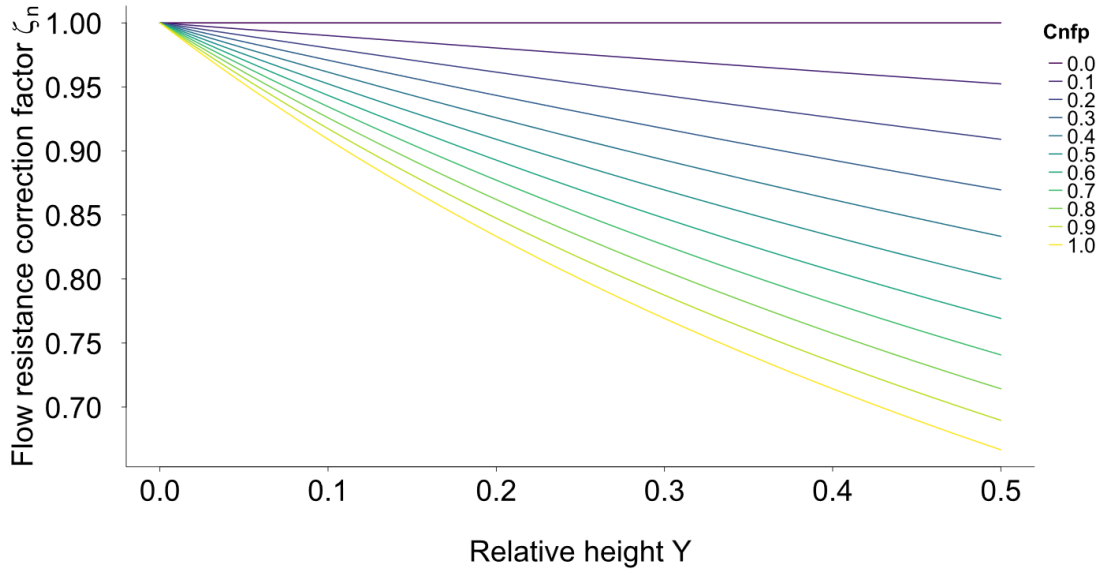

Figure RC2.2: Effect of the coefficient C_{nfp} on the Manning coefficient when the floodplain is active.

Additional details have been added in the Supplementary Material (S5).

The section 4.3.1 was revised, after Eq. (28):

In the calibrated model, C_{nfp} values along the mainstem range between 0.3 and 1. Under typical flood conditions, Y varies between 0 and about 0.3, resulting in an increase of the effective Manning coefficient of approximately 0–6% (see Supplement S5). Larger corrections, up to about 20%, may occur only during extreme simulated floods when Y approaches 0.5. Such variations are consistent with resistance changes reported in studies of compound channel hydraulics and floodplain–channel momentum exchange (e.g., Nicollet and Uan, 1972; Smart, 1992; Shiono and Knight, 1996; Bousmar and Zech, 1999). The coefficient C_{nfp} was calibrated during the model calibration phase to reproduce observed stage–discharge relationships at the mainstem gauging stations.

---

6. Line 364/365. I would ask the Authors to shortly provide details about the calibration of Eq.(32).

The parameters of Eq. (32), namely C_{nch} and h_{ch}, were calibrated together with the other hydraulic parameters of the routing scheme during the model calibration phase. The calibration aimed at reproducing the observed relationships between discharge or velocity and water level at the mainstem gauging stations, with particular attention to low-flow conditions where bed roughness exerts a stronger influence on flow resistance. Calibrated values of h_{ch}, range between ~ 35% and 75% of the maximum water level across the mainstem reaches, and C_{nch} between 0.12 and 0.4. The resulting correction increases the effective Manning coefficient by approximately 10–30% at low water, decreasing progressively as stage rises h_{ch} (See Fig. S7).

This behaviour is consistent with the well-known stage-dependence of bed roughness in sand-bed rivers and the two parameters remain identifiable as h_{ch} controls the threshold below which the bed roughness correction activates, reflected by an inflection in the observed stage–discharge relationship at low stage, while C_{nch} controls the amplitude of the resulting Manning coefficient increase.

The manuscript was revised as follows (section 4.3.2):

where C_{nch} (–) is a coefficient and h_{ch} (m) the water height below which the additional bed roughness correction activates. Calibrated values of h_{ch} range between 35% and 75% of the maximum simulated water level across the mainstem reaches, and C_{nch} between 0.12 and 0.4, resulting in an increase of the effective Manning coefficient of approximately 10–30% at low stage, decreasing progressively as stage rises above h_{ch} (Fig. S7). Both parameters were calibrated to reproduce the observed stage–discharge and stage–velocity relationships at the mainstem gauging stations.

7. Line 384. Yamazaki et al (2011) suggested to apply the diffusive wave equation to include backwater effect on water surface elevation. I would ask the Authors present and discuss the difference between the results from the diffusive wave and those from the kinematic wave in a Supplementary Material.

We thank the reviewer for this comment. The routing scheme implemented in the modified SWAT version used in this study already relies on the diffusive wave formulation, which was specifically selected to represent backwater effects on water surface elevation in large low-slope rivers. The limitations of the kinematic wave approximation for representing backwater effects in large rivers have been documented in the literature (e.g. Yamazaki et al., 2011).

Because the simulations presented in this study already rely on the diffusive wave formulation, a direct comparison with the kinematic wave approach was not considered necessary for the objectives of this work. Additional comparisons between routing formulations were performed during the development phase of the SWAT-Amazon framework and are documented in Santini (2020). These analyses confirmed the expected limitations of the kinematic wave approximation for representing backwater effects in large low-slope floodplain rivers.

To avoid ambiguity, the manuscript has been revised to explicitly clarify that the routing scheme used in this study is based on the diffusive wave formulation.

The manuscript was revised as follows (section 4.1.2):

SWAT-Amazon enables reach-specific selection between kinematic wave ($S_f = S_b$), suitable for steep Andean reaches, and diffusive wave approximation ($S_f = S_b + S_w$), preferred for low-slope floodplain reaches where backwater effects may occur (e.g. Yamazaki et al., 2011), with S_f (m m^{-1}) the energy gradient (or friction slope), S_b (m m^{-1}) the bed slope and S_w (m m^{-1}) the water-surface slope.

We also add the following sentence after Equation 4 (section 4.1.2)

In the present study, the diffusive wave formulation was applied throughout the river network in order to account for backwater effects in the low-slope reaches of the Ucayali River.

8. Section 5.3. It seems that the model was applied (Fig.6) for the same period considered for calibration (09/2009–09/2015). The Authors should better clarify this point.

This is an important point that deserves clarification. We acknowledge that the wording in the original manuscript was ambiguous regarding the distinction between calibration and the reported performance periods, which may have given the impression that calibration and validation fully overlap. We also acknowledge that the available monitoring dataset limits the definition of a fully independent validation period for sand routing in the conventional sense. Nevertheless, model robustness is supported by: (i) consistent performance metrics over periods outside the calibration window (Table RC2.1), with reduced performance at Lagarto attributable to increased forcing and observational uncertainties rather than model degradation; and (ii) the comparison with direct gauging measurements (Fig. RC2.3), which constitute a quasi-independent validation as these observations are not explicitly used in calibration. Residual uncertainty is further addressed through the GLUE analysis presented in Section 6.

We address a complete response below:

For water routing (Section 5.1), the 2000–2016 period does not correspond to a calibration period, but rather to a common observation window used to compute performance metrics across stations, depending on data availability. Model calibration for water discharge (Table RC1.1) was conducted over the 2010–2015 period (except for Puerto Inca, for which the period was 2012–2014), selected based on data quality and availability. Furthermore, calibration does not rely on the optimization of a single time series, but on multiple hydraulic diagnostics, including water levels (Fig. 4a), velocities (Fig. 4b), and stage–discharge and stage–velocity relationships (Fig. 5). Moreover, independent hold-out periods were used to further evaluate model performance (Table RC1.1), showing consistent performance metrics across all periods.

Table RC1.1. Discharge model performance over calibration and independent validation periods. Calibration was performed over 2010–2015 (except for Puerto Inca, for which the period was 2012–2014), while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed daily discharge records.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Independent validation	01/2009–12/2009	287	0.9	0.94	-1.3
Lagarto	Calibration	01/2010–12/2015	2191	0.89	0.94	-1.7
Lagarto	Independent validation	01/2016–12/2019	1162	0.84	0.90	-1.7
Puerto Inca	Calibration	09/2012–08/2014	730	0.73	0.86	-1.1
Puerto Inca	Independent validation	09/2015–08/2016	731	0.69	0.83	-7.1*
Pucallpa	Independent validation	01/2000–09/2009	3653	0.94	0.94	-4.8
Pucallpa	Calibration	01/2010–12/2015	2161	0.92	0.89	1.1
Pucallpa	Independent validation	01/2016–09/2019	1389	0.91	0.94	-1.2
Requena	Independent validation	01/2000–09/2009	3653	0.92	0.93	-3.8
Requena	Calibration	01/2010–12/2015	2191	0.9	0.95	0.1
Requena	Independent validation	01/2016–12/2019	1461	0.85*	0.83*	-9.5*

* Lower performance during these periods is associated with reduced quality of observed discharge data and/or precipitation inputs.

For sand routing (Section 5.3), the 09/2009–08/2015 period shown in Fig. 6 corresponds to the interval with the most reliable and dense observations, as explained in the manuscript, although data quality remains heterogeneous depending on the period and sampling protocols. This period was therefore selected for model calibration. However, outside this interval, uncertainties in sand flux observations (derived from surface sediment concentration monitoring and sediment gauging; see Supplementary Material S3) increase, which precludes the definition of a robust and fully independent validation period. In particular, at Lagarto, the NSE for the validation period decreases to 0.45 (Table RC2.1). When excluding high waters (flood peaks), the NSE remains high (0.80), indicating robust model performance under low- and intermediate-flow conditions. The lower performance is therefore mainly associated with rapid Andean flood events and related uncertainties in rainfall forcing.

Table RC2.1. Sand routing model performance over calibration and independent validation periods. Calibration was performed over 2009–2015, while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed sand flux at surface concentration sampling time step.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Calibration	09/2009–08/2015	323	0.80	0.87	7.9
Lagarto	Independent validation	09/2015–08/2018	181	0.45	0.60	27.7
Requena	Calibration	09/2009–08/2015	404	0.86	0.92	-2.3
Requena	Independent validation	09/2015–08/2018	77	0.80	0.70	-24.0

Overall, reduced performance after the calibration period reflects both increased observational and forcing uncertainties rather than a degradation of model performance.

Therefore, to further evaluate model performance, we performed an additional comparison between daily simulations (discharge and suspended sand fluxes) and direct gauging measurements over the 09/2009–08/2016 period (Fig. RC2-3). This period starts with the implementation of improved sediment sampling protocols during field measurements, while simulations after 2016 were not considered due to increased uncertainties in rainfall inputs at Requena (Fig. 3d). This comparison bypasses rating curves and provides a more direct evaluation of model performance. Although not fully independent, it offers a complementary and partially independent validation, as direct gauging measurements are not explicitly used in the calibration, even though they contribute to the construction of rating curves over the full observation period. The agreement between simulations and gauging measurements supports model robustness, with bias-corrected R^2 (bR^2) values of 0.90 for discharge and 0.88 for sand fluxes (Fig. RC2-3).

The corresponding results have been added in the Supplementary Material (Fig. S9) and are briefly discussed in the revised manuscript. We have also revised the text to explicitly distinguish between calibration, performance assessment, and the different sources of observational constraints.

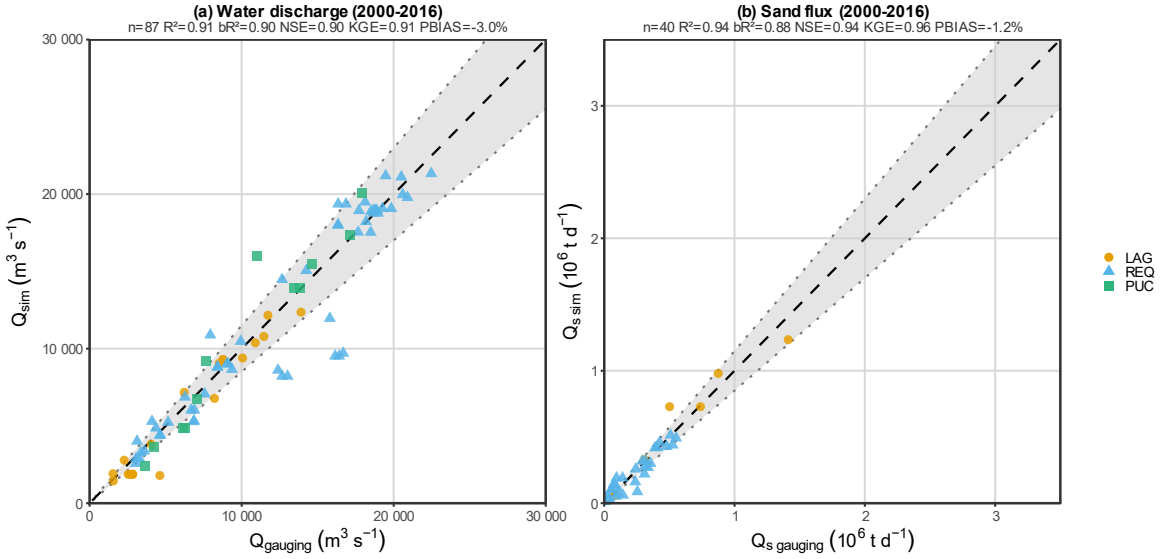

Figure RC2.3: Comparison of simulated and measured river discharge (a) and suspended sediment flux (b) at gauging stations in the Ucayali basin. Simulated values ($\boldsymbol{Q_{sim}}$, $\boldsymbol{Q_{s,sim}}$) correspond to daily outputs from the SWAT-Amazon model extracted at the dates of field measurements. Observed values ($\boldsymbol{Q_{gauging}}$, $\boldsymbol{Q_{s,gauging}}$) are direct field measurements, independent of rating curve derivation. The dashed line indicates the 1:1 relationship, and the shaded band represents the ±15% envelope. The Puerto Inca station was excluded, as its sub-daily flood pulse dynamics are not adequately captured at the model's daily time step. Axis ranges reflect the observed variability of $\boldsymbol{Q}$ and $\boldsymbol{Q_s}$ across all stations. Statistical indicators are computed over all stations pooled: n = number of paired observations; R^2 = coefficient of determination; bR^2 = bias-corrected coefficient of determination; NSE = Nash–Sutcliffe efficiency; KGE = Kling–Gupta efficiency; PBIAS = percent bias.

The manuscript was revised as follows:

Section 5.3

Calibration focused on the 09/2009–08/2015 period, when sediment monitoring protocols were enhanced, including higher sampling frequency at Requena between 11/2012 and 06/2013, where one sample was collected each two days plus three sampling repetitions each ten days. Beyond, sampling was conducted at five-day intervals during the wet period between July 2013 and September 2015. Additionally, the concentration gaugings were performed in all sites with a higher number of samples collected throughout the cross-section, particularly in the first half of the water column, to ensure more accurate sand concentration calculations. Outside this interval, uncertainties in sand flux observations increase, which complicates the definition of a robust and fully independent validation period (see Supplement S9 for a summary of model performance across periods). The lower performance outside the calibration period primarily reflects uncertainties in rainfall forcing and observations, particularly during rapid Andean flood events, rather than a degradation of model performance. To further evaluate model performance, direct comparisons between simulations and gauging measurements were performed (Fig. S9). These complementary evaluations support the temporal robustness of the model despite observational limitations and data heterogeneity.

Supporting analyses and details are now provided in the Supplementary Material (S9).

Minor comments

9. Line 135. The quality of Fig.1 should be improved.

Thank you for this comment. The lower quality of Fig. 1 results from the resolution used in the EGUsphere discussion paper. The figure is already provided at full resolution (300 dpi) and will appear in high quality in the final manuscript.

10. Line 325. On compound channels hydrodynamics see also: Proust, S., & Nikora, V. (2020). Compound open-channel flows: effects of transverse currents on the flow structure. Journal of Fluid Mechanics, 885, A24. https://doi.org/10.1017/jfm.2019.973

Thank you for this suggestion. The reference has been added in the revised manuscript (section 4.3.1) to complement the discussion on compound channel hydrodynamics.

11. Line 342. How was the coefficient Cfnp calibrated ?

The coefficient C_{nfp} was calibrated together with the other hydraulic parameters of the routing scheme during the model calibration phase. The calibration aimed at reproducing the observed relationships between discharge and water level at the mainstem gauging stations. The resulting values range between 0.3 and 1 along the mainstem, reflecting spatial variations in floodplain interaction intensity. This clarification has been added in Section 4.3.1 of the revised manuscript (see response to Referee Comment #5).

12. Line 667/668. I suggest to further elaborate the discussion on the Kbed coefficient of erosion susceptibility. You might provide details on that at line 289.

We thank the reviewer for this suggestion. A short clarification has been added in Section 4.2.3 to explain the role of the coefficient K_{bed} , which represents the susceptibility of the riverbed to erosion when the simulated sand transport capacity exceeds the available sand load. This parameter therefore controls the potential entrainment of bed material into suspension and regulates the contribution of riverbed erosion to the simulated suspended sand flux. It was calibrated together with the other sediment routing parameters during the model calibration phase.

Proposed revision in the manuscript:

Section 4.2.3, after Eq. 18:

where K_{bed} (–) is a coefficient ($0 \leq K_{bed} \leq 1$) representing the susceptibility of the channel bed to erosion when the simulated sand transport capacity exceeds the available sand load. This parameter governs the potential entrainment of bed material into suspension and thus the possible contribution of riverbed erosion to the simulated suspended sand flux.

Section 4.2.4:

Riverbed erosion, E_{bed} (Eq. 18) and two deposition terms (D_{ovbk}, D_{lat}), all expressed in (t d^{-1}), are also considered. Riverbed erosion is therefore explicitly represented in the model formulation so that its potential contribution to the suspended sand flux can be evaluated during calibration.

Section 6.3:

Lastly, the distinction between E_{bk} and E_{bed} is partly supported by the calibration experiments. When bed erosion is activated ($k_{bed} > 0$), the model tends to generate rapid and abrupt peaks in simulated sand flux once the transport capacity exceeds the available sand load, whereas adjusting C_{bk} produces a smoother and more progressive increase in sand concentration that better reproduces the secondary peaks observed during the recession phase. The calibrated values of K_{bed} remain very small ($K_{bed} \ll 1$), suggesting that riverbed erosion contributes only weakly to the simulated suspended sand flux in the main stem, although a minor contribution cannot be fully excluded.

---

13. Line 785. List of model parameters. Water kinematic viscosity has two different symbols.

Thank you for pointing out this inconsistency. The notation of the water kinematic viscosity has been corrected to ensure a consistent symbol throughout the manuscript and in the list of model parameters.

Additional references

Bousmar, D. and Zech, Y.: Momentum transfer for practical flow computation in compound channels, Journal of Hydraulic Engineering, 125, 696–706, https://doi.org/10.1061/(ASCE)0733-9429(1999)125:7(696), 1999.

Doxaran, D., Froidefond, J. M., Lavender, S., and Castaing, P.: Spectral signature of highly turbid waters: Application with SPOT data to quantify suspended particulate matter concentrations, Remote Sensing of Environment, 81, 149–161, https://doi.org/10.1016/S0034-4257(01)00341-8, 2002.

Shiono, K. and Knight, D. W.: Turbulent open-channel flows with variable depth across the channel, Journal of Fluid Mechanics, 222, 617–646, https://doi.org/10.1017/S0022112091001246, 1991.

Response to Community Comment #1 (CC1)

General Comments

Thank you for the opportunity to review this manuscript. The paper presents an integrative framework to quantify fine and sand-sized sediment fluxes in a large floodplain river system by combining discharge modelling, floodplain-aware routing, and satellite-derived suspended sediment estimates. The topic is timely and potentially high impact. I have listed my section-based comments below.

Abstract

1- *While I understand the large goal of the paper is to highlight the role of floodplain in sediment transport, and the paper is attempting to build an integrated approach using discharge estimates to calculate fine sediment load and land surface model to simulate sand load. I cannot understand from the latter half of the Abstract what critical research questions this paper attempts to solve, except the role of floodplain in sediment transport.*

2- *The abstract misses in providing uncertainty or targeted application of this study, as the last three lines reads very generic. Until the reader reads the full paper, he/she may not understand the latter part of Abstract.*

We thank the reviewer for these helpful comments. We agree that the key research questions were not sufficiently explicit in the original abstract. The abstract has therefore been revised to clearly state the main objectives of the study, namely the development of an integrated framework combining long-term observations, remote sensing, and hydrological–hydraulic modelling to quantify multi-decadal hydro-sedimentary budgets and analyze the role of floodplains in sediment dynamics in large Amazonian rivers, and to better specify the associated uncertainties and intended applications of the approach.

The revised abstract reads as follows:

The Amazon basin is undergoing increasing environmental changes, potentially approaching a climatic tipping point in the coming decades. Understanding how these changes affect water and sediment fluxes is key for constraining large-scale biogeochemical cycles, yet conventional hydrological networks lack the spatial and temporal resolution required to accurately quantify hydro-sedimentary budgets.

To address this limitation, we develop an integrated, physically constrained framework combining long-term observations, remote sensing, and hydrological–hydraulic modelling (SWAT-Amazon) to quantify multi-decadal hydro-sedimentary budgets and investigate how floodplain inundation controls sediment dynamics in large Amazonian rivers. Focusing on the Ucayali River, a major foreland tributary of the Amazon, this study provides the first detailed, long-term hydro-sedimentary budgets for the Upper Amazon, distinguishing fine sediment fluxes from sand loads.

Results reveal a previously undocumented floodplain-controlled sand sedimentation process: during high waters, large floodplain water storage (up to 19.1 [15.3, 22.9] km^3, ~38% of discharge) reduces main-channel transport capacity, capturing up to 14% [10%, 20%] of the sand flux at peak discharge, while recycling during recession contributes 22% of the total suspended load at the basin outlet. This dual control partially decouples sediment transport from water discharge. The Andean Ucayali exports 455 [410, 500] × 10^6 t yr^{-1} of suspended sediment (40% sand), of which 36% is trapped within the floodplain, predominantly as sand (65% of total deposition). The river delivers 290 [235, 345] × 10^6 t yr^{-1} to the Amazon River (26% sand), making it the dominant sediment source among the Andean foreland tributaries. Uncertainty analysis combining Sobol indices and GLUE simulations shows that, despite substantial equifinality among secondary floodplain parameters, sediment fluxes and associated trapping and recycling fractions remain stable across all behavioural simulations. Budget accuracy is therefore controlled by long-term, multi-variable, multi-source observations rather than by parameter calibration or model structure alone.

These findings demonstrate that floodplains control hydro-sedimentary fluxes in large river systems and act as dynamic regulators of sediment transport, storage, and recycling, with major implications for biogeochemical cycles.

Introduction

3- ***Section 1.4:*** *What are the key research questions that the authors are trying to address in this research? Role of floodplain in sediment transport? The paper should state 2–3 explicit objectives/hypotheses early (end of Abstract + end of Intro). Kindly add a 1–2 sentence novelty statement distinguishing this from prior Amazon basin sediment studies*

We thank the reviewer for this suggestion. The end of the introduction (Section 1.4) has been revised to explicitly state the main objectives of the study and to clarify its novelty compared with previous Amazon sediment studies:

Revised text:

However, to date, no study has yet combined remote sensing with modelling to investigate sediment dynamics in detail in the Amazon. Building on long-term CZO HyBAm conventional observations, this study introduces an integrated framework coupling field calibration campaigns, satellite remote sensing, and hydraulic–hydrological modelling to derive process-based hydro-sedimentary budgets for a major foreland tributary of the Upper Amazon: the Ucayali River. Satellite altimetry and satellite-derived fine sediment estimates are used to constrain a hydrological–hydraulic modelling scheme (SWAT-Amazon) simulating water and sand fluxes, with floodplains represented using a simplified reservoir approach.

The objectives are to (i) quantify long-term hydro-sedimentary budgets at sub-basin scale, (ii) investigate the impact of floodplain inundation on water and sediment fluxes, thereby identifying key processes controlling sediment transport, storage, and recycling that remain poorly constrained by conventional river monitoring approaches, and (iii) distinguish the respective contributions of fine sediment fluxes, associated with organic matter and pollutant transfer, and sand loads related to river dynamics.

Unlike previous basin-scale sediment studies in the Amazon, which largely relied on sparse gauging networks or large-scale modelling approaches, this framework provides process-based hydro-sedimentary budgets through multi-source data integration. Given the key role of sediment dynamics in biogeochemical cycles, it also contributes to improving the understanding of the Amazon's role in global material fluxes and of the potential impacts of environmental changes on its hydrology and sediment transport.

Integrative strategy

4- Figure 1 is scattered. How a), b) & c), and d) are linked with each other is not easy to comprehend with the figure. Kindly sync them

The figure caption has been clarified to better explain the logical connection between panels.

Modified figure caption:

Figure 1: General schematic overview of the proposed methodology. *Panels (a–c) illustrate the integrative approach.* ***(a) Types of stations. (b) Typical bimodal particle size distribution*** *(PSD) in the large Amazonian rivers, identifying two main size groups: 1- fine sediments that can be monitored by satellite but not modelled; 2- fine sands in graded suspension, invisible to satellites but whose transport capacity can be modelled.* ***(c) Integrated approach*** *combining remote sensing, modelling, and calibration campaigns.* ***(d) SWAT-Amazon****, a tailored version of the SWAT model for simulating water and sand fluxes. This modelling framework consists of a Fortran-based executable (SWAT-Amazon.exe) and an R notebook (Run-SWAT-Amazon.Rmd) used for model runs, simulation analysis, interactive visualization, sensitivity analysis and calibration with the SWATrunR package (Schürz et al., 2019).*

5- I doubt the assumption of assuming mean diameter for fine sediment in the range of 10-20 micrometer. This because Firstly, are there any records of PSD profile of the basin. Secondly, 10- 20 micrometer sits at the lower range of very fine silt. Thus, underrepresenting coarse silt or large composition of fine sediments in the range of 20- 100 Micrometer. The sediment transport of both the group large differs, with former influenced by flocculation, organic matter, and high suspension. And latter having conventional particle dynamics. To simplify, why don't the authors use at least 3 particle size to better represent the PSD, clay, silt and sand.

We thank the reviewer for this insightful comment. The representation adopted in this study is supported by both field observations and methodological considerations. Indeed, particle-size distribution (PSD) measurements available for suspended sediments in the Ucayali River and other white-water rivers of the Amazon Basin (Fig. CC1.1a) consistently show a predominantly bimodal structure, with a dominant fine mode composed mainly of silts and clay aggregates and a sand mode centred in the very fine sand range. This typical bimodal structure has been documented in several previous studies (e.g. Vanoni, 1979, 1980; Bouchez et al., 2011; Armijos et al., 2016; Santini et al., 2019; Martinelli, 2022). Thus, the mean diameter ranges reported in the manuscript (10–20 µm for fine sediments and 80–120 µm for sands) should be understood as order-of-magnitude estimates of the mean diameters of these dominant grain-size classes, rather than strict granulometric bounds of the same groups.

The separation between these two dominant size groups is defined by the sieving at 63 µm of the suspended sediment samples (see the new section in the Supplementary Material (Section S2),

where the HyBAm monitoring protocol is now detailed). This threshold provides a consistent and reproducible basis for sediment flux measurements across the basin. Therefore, intermediate particle sizes of the PSD continuum, in particular the coarser silts (~20–63 µm), are not neglected, but are represented by one of these two mean diameters.

For the fine sediment group, several previous studies in the Amazon Basin have demonstrated that their concentration and flux can be robustly quantified from satellite reflectance (e.g. Martinez et al., 2015; Espinoza-Villar et al., 2012; Park and Latrubesse, 2014). Despite processes such as flocculation, aggregation with organic matter, or variations in particle properties (shape, density, mineral composition, or Si/Al ratios), the bulk of fine sediments generally behaves at the reach scale as a wash-load fraction. In the remaining subgroup of the coarse silts size (~20–63 µm), as described by the reviewer, **t**wo distinct particle behaviours may occur**.** First, aggregates, which have a lower density than non-cohesive particles, tend to behave as passive scalars in the flow. A smaller fraction of non-cohesive silt particles with higher density may exhibit graded suspension behaviour; however, their contribution to the uncertainty in fine sediment fluxes is likely negligible compared to the bulk of the fine fraction. Moreover, given the range of shear velocities in large Amazonian rivers (roughly 0.01–0.4 m s^{-1}), the Rouse number of non-cohesive silts close to the 63 µm threshold is expected to remain low (typically ~0.01–0.2 depending on particle density and flow conditions), consistent with a wash-load to fully suspended regime. Therefore, even if a very small clay fraction can be detected in PSD measurements, and even if this subgroup dominates satellite reflectance according to Mie scattering theory, the fine sediment group can be robustly represented as a single wash-load fraction at the scale considered in this study.

From the 63 µm threshold, non-cohesive coarse silts are considered equivalent to very fine sands transported in a graded suspension regime, and represented by a single effective mean diameter in the transport capacity equations. Even if some coarse aggregates may temporarily resist turbulent shear, their contribution to the suspended sand mass remains negligible. For non-cohesive particles, given the quadratic dependence of settling velocity on particle size, the graded suspension regime is rapidly reached, and the uncertainty induced by the small volume of particles behaving closer to wash load than to graded suspension is expected to be negligible.

Overall, field observations indicate a predominantly bimodal PSD structure, suggesting that a separation into three size groups is not warranted in this context. Such a division would increase model complexity and parameter uncertainty without clear improvement in budget robustness. The two-fraction representation is therefore considered sufficient, with only second-order impacts on simulated fluxes.

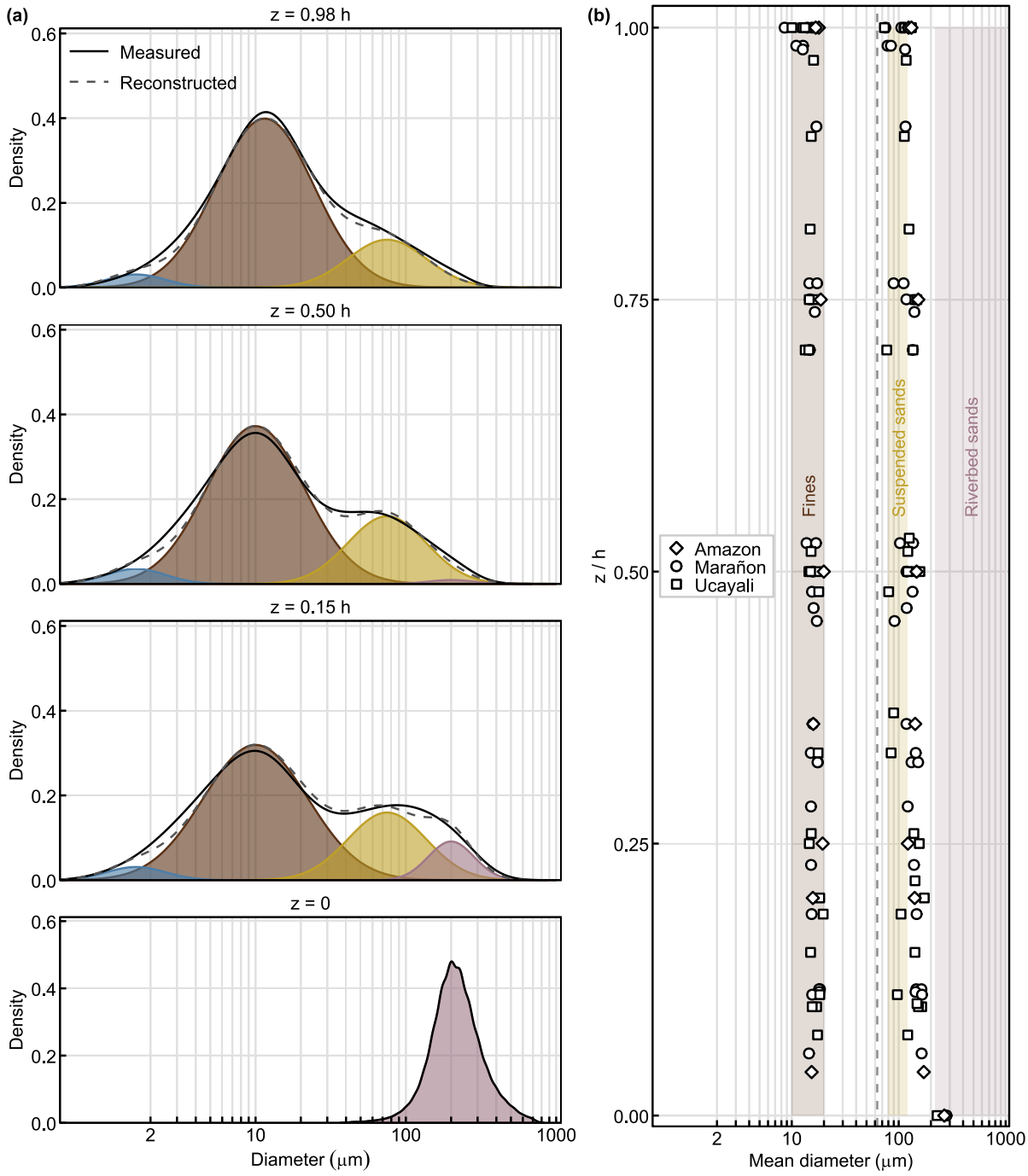

Figure CC1.1: Typical PSD of large white-water rivers in the Amazon Basin (adapted from Santini et al., 2019). (a) Particle-size distributions (PSD) measured in suspension at different relative depths and in the riverbed (Requena station, Ucayali River, 16 March 2015). The grey line shows the measured distribution, while the dashed line represents a reconstructed PSD obtained from a mixture of log-normal components corresponding to clay, silt, and sand fractions. **(b)** Vertical profiles of representative grain diameters for the fine and sand fractions. Symbols correspond to measurements at different stations. The dashed vertical line indicates the conventional 63 µm boundary between silt and sand. This figure illustrates the predominance of two main modes in the PSD, supporting the two-fraction representation adopted in this study.

We have added the following paragraph in Section 2 to clarify this point:

These ranges of mean diameter should be understood as representative values of the dominant PSD modes, rather than strict grain size distribution bounds. Intermediate particle sizes (e.g. coarse silts in the 20–63 µm range) are included within the fine sediment fraction defined by the 63 µm threshold used in the HyBAm monitoring protocol (see Section 3.1). Depending on their physical properties, coarse silts may either behave as wash load (e.g. aggregates) or exhibit transport dynamics closer to those of very fine sands when non-cohesive. This supports a first-order bimodal representation of suspended sediment transport at the basin scale, consistent with PSD measurements and vertical grain-size profiles presented in the Supplementary Material (Fig. S1).

6- I feel the current setup may underrepresent sand load in the model. Could you please clarify and provide support?

We respectfully do not find evidence supporting this concern and provide several lines of justification below:

First, in the present framework, suspended sand fluxes are directly constrained by long-term field observations, including cross-sectional sediment sampling, and are therefore not solely determined by the model structure. Model simulations were calibrated and evaluated against these observations and successfully reproduce both the magnitude and temporal variability of measured sand fluxes at the main gauging stations, which indicates that no systematic underestimation of sand load is observed.

Second, from a sediment transport perspective, the use of a representative grain diameter is consistent with the transport formulation adopted in the model. The Camenen and Larson equations are designed to use a representative diameter for the sand fraction and are known to provide robust and reliable estimates under a wide range of hydraulic conditions. Observations of particle-size distributions in the Amazon Basin indicate that suspended sediments exhibit a predominantly bimodal structure, with a clear separation between fine sediments and a single dominant suspended sand mode, which supports the use of a representative sand diameter in the model.

Third, the calibrated representative diameter ($d_s \approx 80$ µm), close to the silt–sand transition, likely integrates part of the coarsest non-cohesive silts, whose hydraulic behaviour approaches that of very fine sands. This suggests that the effective transported fraction represented in the model may be slightly broader than the strict sand class, further reducing the risk of underestimating sand fluxes.

Finally, a secondary coarser mode may exist close to the riverbed under mixed suspension–bedload conditions. However, this fraction is largely associated with bed material and contributes only intermittently to the suspended load, likely through short hydraulic bursts linked to bedform dynamics. In addition, flow velocities decrease in the near-bed region, and this coarser fraction remains concentrated within the near-bed layer (0–0.2h), limiting its contribution to depth-integrated fluxes. Therefore, its contribution to the depth-integrated suspended sand flux is expected to remain minor at the temporal and spatial scales considered in this study.

Taken together, these elements indicate that the two-fraction representation does not lead to an underestimation of sand load but rather provides a physically consistent and observation-constrained approximation of suspended sediment transport at the basin scale.

Conventional data

7- ***Section 3.1:*** *Kindly introduce the term "super stations" in this section as you talk about long data stations, the virtual stations and low data station in the next station.*

The term "super stations" is already introduced and defined in Section 2 (Integrative strategy), where the three types of stations used in the framework (low-data stations, virtual stations, and super stations) are described. No modification to the manuscript was considered necessary.

8- ***Figure 2:***

- The blue color interferes with the color used in the map. Kindly use some other color to mark the stations. Perhaps red. It is not needed to write the years in the map.

*- "The text box details observation periods for water level (**h**) discharge (**Q**), and suspended sediment concentration (**C**)." It is not easy to comprehend.*

- River mapping is not clear

The figure has been revised to improve readability: station markers now include a bolder border to increase visibility, wetland colors have been lightened to improve contrast with other map elements, and river network line width has been increased from 1 to 3 points. The observation periods were retained in the figure as they provide important contextual information on the temporal coverage of the datasets; the abbreviated notation (h, Q, C) is defined in the figure caption, as space constraints preclude spelling out full variable names directly on the map.

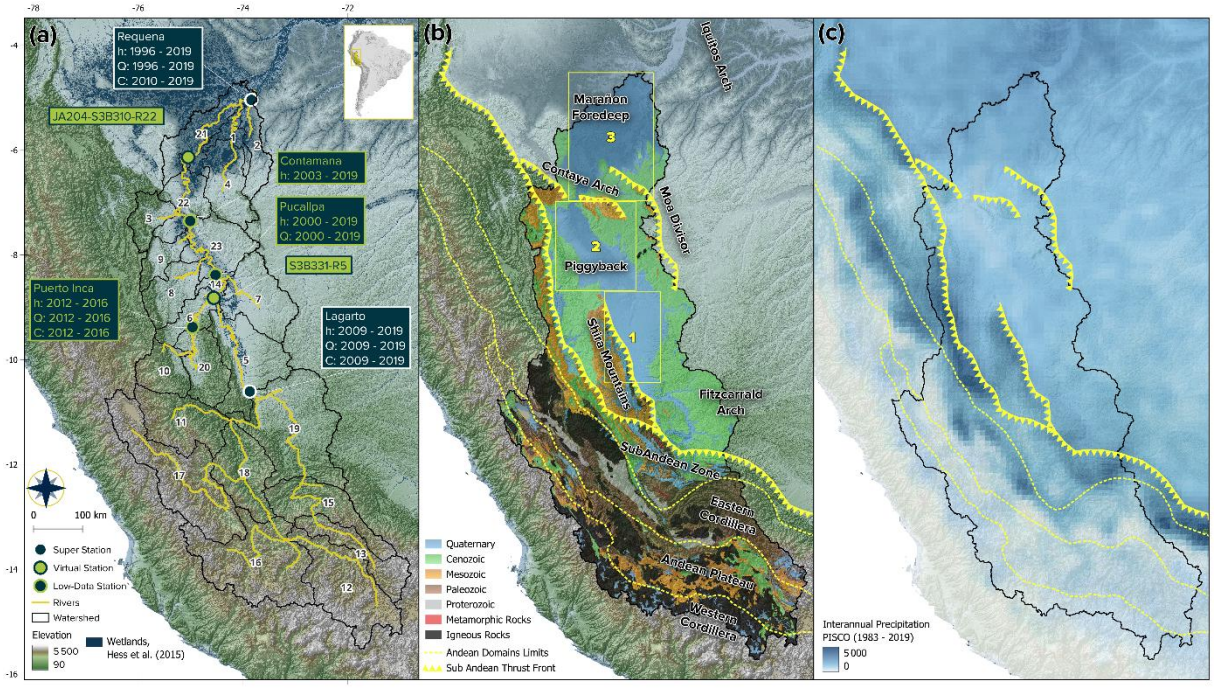

Figure CC1-2: Revised Figure 2.

9- ***Section 3.3.2:*** *Could you please clarify, how total, fine, and sand concentration was determined. Were the samples dry or wet sieved. Was gravimetric filtration used or else?*

10- ***Section 3.3.3:*** *How the vertical gradient in sediment concentration was considered in the field measurements?*

11- *Does the site also have lateral change in the suspended sediment concentration? Was the mean concentration representative vertically and/or laterally? How would author try to convince this?*

This point is clarified in the revised manuscript, and additional details on the measurement protocols are now provided in the Supplementary Material S2. A reference to this new supplementary section has been added in Section 3.1 (Conventional data).

Suspended sediment concentrations were determined using standard methods. After field collection, water samples were transported to the HyBAm laboratory in Lima, where the sand fraction was separated from the fine fraction using a 63 µm sieve following ASTM D3977 procedures. Both fractions were then filtered through 0.45 µm cellulose acetate filters, dried at 50 °C for 24 h, and weighed to determine sediment concentrations.

Vertical concentration gradients were measured during dedicated field campaigns conducted at the HyBAm gauging stations. Suspended sediment samples were collected using a point-sampling method along three verticals across the river section. The sampling protocol evolved during the monitoring period. Early campaigns typically included three samples per vertical (near-bed, mid-

depth, and near-surface). Later campaigns increased the vertical resolution, with sampling levels typically located at 0.98 h, 0.75 h, 0.5 h and 0.25 h above the riverbed (occasionally including additional levels). As a result, between 9 and 15 samples per cross-section were generally used to estimate cross-sectionally averaged concentrations of total suspended sediments, fine sediments and sands.

Lateral variability was addressed by sampling along three verticals distributed across the channel, allowing cross-sectional mean concentrations to be derived from multiple vertical profiles. These measurements were used to establish empirical relationships between surface index concentrations and cross-sectionally averaged concentrations.

The dataset used in this study relies on more than two decades of systematic monitoring within the HyBAm observatory and 82 dedicated campaigns conducted between 2003 and 2019, representing an unprecedented hydro-sedimentary observational effort for large Amazonian rivers.

Model development

12- ***Section 4.1:*** *Please explain how reach-wise inputs are derived and how this can be reproduced in other basins. Additionally, given floodplain activation is event-scale, please discuss how a daily routing framework affects peak timing/magnitude and floodplain exchange, and whether it biases sediment flux estimation.*

Reach-wise inputs follow the standard SWAT routing framework, where each reach receives routed inflows from upstream reaches through the predefined reach network structure. In the sand routing module, this incoming routed sediment flux represents the sediment supply entering the reach. The effective sand flux exported from the reach is subsequently recalculated according to the local transport capacity. The resulting sand dynamics are governed by the mass balance formulated at the reach scale (Eq. 25), which accounts for riverbed erosion when transport capacity exceeds the incoming supply, as well as exchanges with the floodplain through overbank deposition and floodplain-derived inputs. Consequently, the exported sand flux is not imposed directly by the incoming routed flux but results from the balance between sediment supply, transport capacity, and floodplain–channel exchanges as described in Section 4.2.4.

We add to the manuscript the following sentence (section 4.2.5):

The routed sediment inflow provided by the SWAT-Amazon reach network represents the sediment supply entering each reach and is incorporated into the reach-scale sand mass balance described in Eq. (25).

Regarding the routing time step, the reviewer raises an important point. Although SWAT-Amazon simulations are reported at a daily time step, the routing equations for water and sediments are internally solved at a sub-daily time step dynamically determined according to the Courant–Friedrichs–Lewy (CFL) stability condition. This ensures that the computational time step remains smaller than the time required for the fastest wave allowed by the governing equations to travel across a reach, thereby maintaining numerical stability.

Consequently, the routing module iterates internally over intraday time steps to satisfy the CFL criterion, and the daily outputs correspond to the model state at the end of each simulation day. This adaptive routing ensures that flood-wave propagation and attenuation remain correctly resolved despite the daily reporting of model outputs. In large low-gradient Amazonian rivers, where flood-wave propagation occurs over time scales of several days to weeks, the daily reporting time step therefore does not significantly affect peak timing, floodplain exchange, or sediment flux estimates.

We added the following clarifications in the manuscript:

Section 4.1.3:

Although SWAT simulations are reported at a daily time step, river routing is internally solved at a sub-daily time step dynamically determined by the Courant–Friedrichs–Lewy (CFL) stability condition. The routing module automatically adjusts the internal time step to satisfy the CFL criterion and ensure numerical stability (Bates et al., 2010). Daily outputs therefore correspond to the model state at the end of each simulation day.

Section 4.2.5:

As for water routing, sand routing is internally solved at a sub-daily time step to satisfy the CFL stability condition.

13- ***Section 4.2****: The approach relies on an effective bed-material sand diameter; please justify its spatial representativeness (Amazon has strong upstream–downstream contrasts) and provide a sensitivity test to plausible diameter ranges (or consider a simple multi-class sand representation).*

In the Upper Amazon, suspended sand transported by the mainstem largely originates from Andean sources (e.g. Baby et al., 2009). Downstream of the Sub-Andean zone, strong hydraulic sorting and long transport distances tend to produce relatively homogeneous sand grain-size characteristics along the lowland reaches. Available particle size measurements from the HyBAm observatory indicate a consistent sand mode along the Ucayali mainstem, with mean diameters of approximately 260 µm at Lagarto, 243 µm at Pucallpa, and 228 µm at Requena.

This relatively limited variability in the Ucayali is consistent with previous observations in Amazonian rivers, where riverbed and suspended sand fractions tend to remain relatively uniform along long lowland reaches due to hydraulic sorting and mixing processes (e.g. Guyot et al., 1999). More pronounced grain-size contrasts may occur further downstream in the central and lower Amazon due to inputs from tributaries draining distinct geological units, such as the Brazilian and Guiana shields (e.g. the Madeira or Negro rivers; Nordin et al., 1980), or locally in systems influenced by major geomorphological structures such as the Pastaza megafans in the Marañón basin. However, such contrasts remain limited within the Upper Amazon basin considered here.

We agree that the sensitivity to the representative sand diameter should be assessed. In the present study, this aspect is already addressed through the global Sobol sensitivity analysis presented in Section 6.1.2. The parameter d_b is identified as one of the most influential parameters controlling sand routing. To ensure a robust exploration of model sensitivity, a

deliberately wide parameter range (180–300 µm) was explored, substantially exceeding the variability indicated by the available field measurements. Interestingly, the calibrated values converge toward the narrower range suggested by observations along the Ucayali mainstem (Lagarto, Pucallpa, and Requena), which provides additional support for the physical consistency of the adopted parameterization.

Future developments of the framework could incorporate a multi-class sand representation if required. However, given the relatively narrow range of sand diameters observed in the Ucayali system and the scale of the present modelling framework, the use of a representative diameter provides a parsimonious and physically consistent representation of sand transport.

Finally, the transport formulation used in the model (Camenen and Larson, 2005, 2008) relies on a representative grain diameter for the sand fraction, which supports the use of a single bed-material sand diameter in the present framework. The formulation could also be applied to several sand classes if required, but given the limited variability observed in the Ucayali system, such an extension was not considered necessary in the present study.

Results and discussion

14- I recommend strengthening the separation of calibration and validation (including timing metrics), making 'not shown' key comparisons available in Supplement, and adding a clearer treatment of uncertainty.

This is an important point that deserves clarification. We acknowledge that the wording in the original manuscript was ambiguous regarding the distinction between calibration and the reported performance periods, which may have given the impression that calibration and validation fully overlap. We also acknowledge that the available monitoring dataset limits the definition of a fully independent validation period for sand routing in the conventional sense. Nevertheless, model robustness is supported by: (i) consistent performance metrics over periods outside the calibration window (Table CC1.1 and CC1.2), with reduced performance at Lagarto attributable to increased forcing and observational uncertainties rather than model degradation; and (ii) the comparison with direct gauging measurements (Fig. CC1.3), which constitute a quasi-independent validation as these observations are not explicitly used in calibration. Residual uncertainty is further addressed through the GLUE analysis presented in Section 6. We address a complete response below:

For water routing (Section 5.1), the 2000–2016 period does not correspond to a calibration period, but rather to a common observation window used to compute performance metrics across stations, depending on data availability. Model calibration for water discharge (Table CC1-1) was conducted over the 2010–2015 period (except for Puerto Inca, for which the period was 2012–2014), selected based on data quality and availability. Importantly, calibration does not rely on the optimization of a single time series, but on multiple hydraulic diagnostics, including water levels (Fig. 4a), velocities (Fig. 4b), and stage–discharge and stage–velocity relationships (Fig. 5). Moreover, independent hold-out periods were used to further evaluate model performance (Table CC1-1), showing consistent performance metrics across all periods.

Table CC1-1. Discharge model performance over calibration and independent validation periods. Calibration was performed over 2010–2015 (except for Puerto Inca, for which the period was 2012–2014), while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed daily discharge records.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Independent validation	01/2009–12/2009	287	0.9	0.94	-1.3
Lagarto	Calibration	01/2010–12/2015	2191	0.89	0.94	-1.7
Lagarto	Independent validation	01/2016–12/2019	1162	0.84	0.90	-1.7
Puerto Inca	Calibration	09/2012–08/2014	730	0.73	0.86	-1.1
Puerto Inca	Independent validation	09/2015–08/2016	731	0.69	0.83	-7.1*
Pucallpa	Independent validation	01/2000–09/2009	3653	0.94	0.94	-4.8
Pucallpa	Calibration	01/2010–12/2015	2161	0.92	0.89	1.1
Pucallpa	Independent validation	01/2016–09/2019	1389	0.91	0.94	-1.2
Requena	Independent validation	01/2000–09/2009	3653	0.92	0.93	-3.8
Requena	Calibration	01/2010–12/2015	2191	0.9	0.95	0.1
Requena	Independent validation	01/2016–12/2019	1461	0.85*	0.83*	-9.5*

** Lower performance during these periods is associated with reduced quality of observed discharge data and/or precipitation inputs.*

For sand routing (Section 5.3), the 09/2009–08/2015 period shown in Fig. 6 corresponds to the interval with the most reliable and dense observations, as explained in the manuscript, although data quality remains heterogeneous depending on the period and sampling protocols. This period was therefore selected for model calibration. However, outside this interval, uncertainties in sand flux observations (derived from surface sediment concentration monitoring and sediment gauging; see Supplementary Material S3) increase, which precludes the definition of a robust and fully independent validation period. In particular, at Lagarto, the NSE for the validation period decreases to 0.45 (Table CC1-2). When excluding flood events, the NSE remains high (0.80), indicating robust model performance under low- and intermediate-flow conditions. The lower performance is therefore mainly associated with rapid Andean flood events and related uncertainties in rainfall forcing.

Table CC1-2. Sand routing model performance over calibration and independent validation periods. Calibration was performed over 2009–2015, while independent validation periods correspond to time windows outside the calibration interval, selected based on data availability at each station. Metrics are computed using observed sand flux at surface concentration sampling time step.

Station	Period type	Period	N (days)	NSE	KGE	PBIAS (%)
Lagarto	Calibration	09/2009–08/2015	323	0.80	0.87	7.9
Lagarto	Independent validation	09/2015–08/2018	181	0.45	0.60	27.7
Requena	Calibration	09/2009–08/2015	404	0.86	0.92	-2.3
Requena	Independent validation	09/2015–08/2018	77	0.80	0.70	-24.0*

Overall, reduced performance after the calibration period reflects both increased observational and forcing uncertainties rather than a degradation of model performance.

Therefore, to further evaluate model performance, we performed an additional comparison between daily simulations (discharge and suspended sand fluxes) and direct gauging measurements over the 09/2009–08/2016 period (Fig. CC1-3). This period starts with the implementation of improved sediment sampling protocols during field measurements, while

simulations after 2016 were not considered due to increased uncertainties in rainfall inputs at Requena (Fig. 3d). This comparison bypasses rating curves and provides a more direct evaluation of model performance. Although not fully independent, it offers a complementary and partially independent validation, as direct gauging measurements are not explicitly used in the calibration, even though they contribute to the construction of rating curves over the full observation period. The agreement between simulations and gauging measurements supports model robustness, with bias-corrected R^2 (bR^2) values of 0.90 for discharge and 0.88 for sand fluxes (Fig. CC1-3).

The corresponding results have been added in the Supplementary Material (Fig. S9) and are briefly discussed in the revised manuscript. We have also revised the text to explicitly distinguish between calibration, performance assessment, and the different sources of observational constraints.

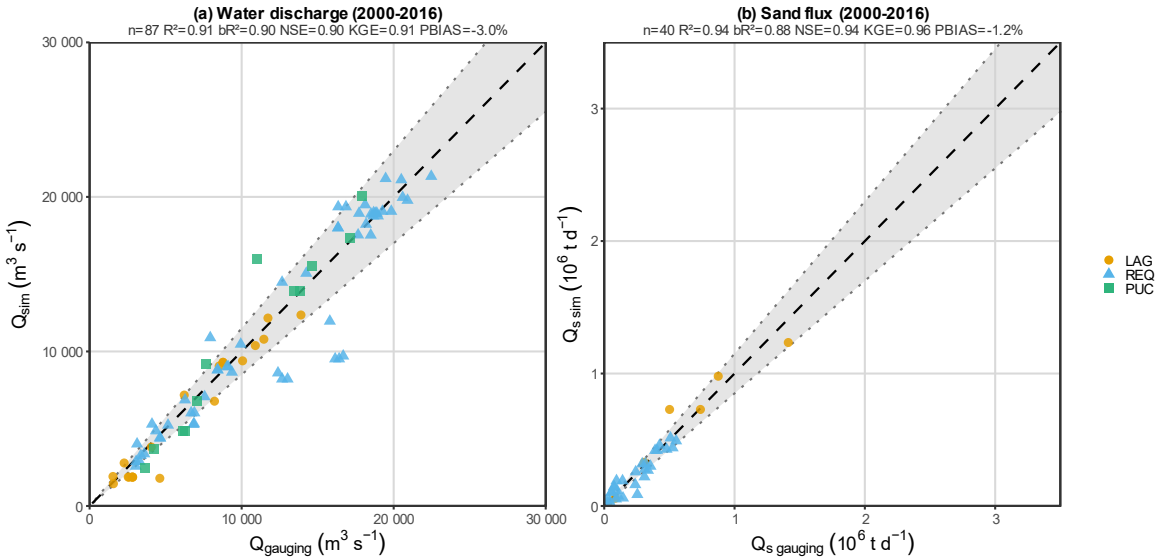

Figure CC1-3: Comparison of simulated and measured river discharge (a) and suspended sediment flux (b) at gauging stations in the Ucayali basin. Simulated values (Q_{sim}, $Q_{s,sim}$) correspond to daily outputs from the SWAT-Amazon model extracted at the dates of field measurements. Observed values ($Q_{gauging}$, $Q_{s,gauging}$) are direct field measurements, independent of rating curve derivation. The dashed line indicates the 1:1 relationship, and the shaded band represents the ±15% envelope. The Puerto Inca station was excluded, as its sub-daily flood pulse dynamics are not adequately captured at the model's daily time step. Axis ranges reflect the observed variability of Q and Q_s across all stations. Statistical indicators are computed over all stations pooled: n = number of paired observations; R^2 = coefficient of determination; bR^2 = bias-corrected coefficient of determination; NSE = Nash–Sutcliffe efficiency; KGE = Kling–Gupta efficiency; PBIAS = percent bias.

The manuscript was revised as follows:

Section 5.1

For water discharge, model calibration was performed over the 2010–2015 period based on multiple hydraulic diagnostics (water levels, velocities, and stage–discharge relationships). Model performance was further evaluated using independent hold-out periods and direct comparisons with gauging measurements that bypass rating curves, as detailed in the Supplement S9. These complementary evaluations show consistent performance across periods and support the temporal robustness of the model.

At a daily time step, SWAT-Amazon simulations at the watershed outlet show excellent performance (Fig. 3d): NSE (Nash-Sutcliffe Efficiency) = 0.92, KGE (Kling–Gupta efficiency) = 0.95, PBIAS (Percent Bias) = -1.8%, LogNSE (NSE on the logarithms of the series) = 0.92 over the 2000–2016 observation period used here for inter-station evaluation (see Moriasi et al. (2007) for details on these metrics).

Section 5.3

Calibration focused on the 09/2009–08/2015 period, when sediment monitoring protocols were enhanced, including higher sampling frequency at Requena between 11/2012 and 06/2013, where one sample was collected each two days plus three sampling repetitions each ten days. Beyond, sampling was conducted at five-day intervals during the wet period between July 2013 and September 2015. Additionally, the concentration gaugings were performed in all sites with a higher number of samples collected throughout the cross-section, particularly in the first half of the water column, to ensure more accurate sand concentration calculations. Outside this interval, uncertainties in sand flux observations increase, which complicates the definition of a robust and fully independent validation period (see Supplement S9 for a summary of model performance across periods). The lower performance outside the calibration period primarily reflects uncertainties in rainfall forcing and observations, particularly during rapid Andean flood events, rather than a degradation of model performance. To further evaluate model performance, direct comparisons between simulations and gauging measurements were performed (Fig. S9). These complementary evaluations support the temporal robustness of the model despite observational limitations and data heterogeneity.

Supporting analyses and details are now provided in the Supplementary Material (S9).

15- The Sobol analysis is performed on 1 sub-basin. Would the authors clarify how this is representative and ideally include at least one additional contrasting sub-basin.

The Sobol sensitivity analysis was conducted on the most downstream sub-basin of the model domain, which is representative of the lowland floodplain system analyzed in this study. This reach is characteristic of large Amazonian rivers, with a very wide floodplain, a predominantly unimodal annual flood pulse, and frequent overbank flooding except during extreme drought years. Because the same routing equations and similar hydraulic conditions govern water and sediment transport along the lowland mainstem, comparable parameter sensitivities would be expected in other lowland sub-basins. Extending the analysis to all sub-basins would therefore not modify the physical interpretation of parameter sensitivities, given the hydraulic homogeneity of the lowland reaches considered here.

16- In sand routing sensitivity, I would like to clarify whether the PSD is stationary or seasonally varying.

In the present modelling framework, the particle size distribution (PSD) of the sand PSD is assumed to remain stationary over the seasonal cycle.

The manuscript was clarified accordingly in section 6.1.2:

The sensitivity analysis, using the parameter set ($\beta_s, C_{bk}, d_b, d_s, \eta, k_{bed}$), shows that d_b is the most critical calibration parameter (Fig. 9d, f), as previously highlighted by Fagundes et al. (2023). In the present framework, d_b and d_s are assumed to remain constant over the seasonal cycle.

17- They state discharge calibration is insensitive to and under some conditions. kindly specify it is sensitive to which regimes (low flow vs floodplain-active) and show quantitatively.

The comment likely refers to Section 6.2. We agree that the original wording was not sufficiently explicit and may have suggested a regime-dependent sensitivity. This was not our intention.

In the hydraulic routing method (Section 2.4), the influence of channel width B and Manning roughness n on simulated discharge Q is strongly limited by compensating hydraulic effects. As explained in the manuscript, reducing n increases flow velocity u and discharge Q (Eq. 5), which reduces the water volume stored in the reach and therefore water depth h (Eq. 1). The resulting decrease in cross-sectional area leads to a proportional reduction of u and Q. This compensation mechanism limits the influence of B and n on routed discharge, while these parameters remain important for reproducing water depth and flow velocity.

The manuscript has been revised (section 6.2) to clarify this point and avoid any ambiguity:

In the hydraulic routing method (Section 4.1), compensating hydraulic effects limit the influence of B *and* n *on routed discharge. Reducing* n *increases flow velocity* u *and discharge* Q *(Eq. 5), which decreases the water volume stored in the reach and therefore* h *(Eq. 1). The resulting decrease in cross-sectional area leads to a proportional reduction in* u *and* Q*.*

18- The authors argue that sand resuspension is driven by floodplain recession and banks/bars erosion and that bed erosion is negligible because allowing bed erosion does not fit peaks as well. Isn't it a strong process statement to make, just based on calibration.

We agree that such an interpretation should not rely solely on calibration results. The manuscript has therefore been revised to moderate the wording and clarify that the calibration outcomes only provide indirect support for the dominant role of C_{bk} over K_{bed}.

In the modelling framework, K_{bed} controls the susceptibility of the riverbed to erosion when the simulated sand transport capacity exceeds the available sand load in the reach. As clarified in Section 4.2.4, this parameter was calibrated together with the other sediment routing parameters and therefore allows the potential contribution of bed erosion to be tested within the model structure. The calibration experiments indicate that when bed erosion is activated ($K_{bed} > 0$), the simulated sand transport capacity can rapidly entrain bed material once the transport threshold is exceeded, producing sharp and abrupt peaks in simulated sand flux. These peaks occur earlier and appear more intense than the resuspension signal observed in the measurements. In contrast, adjusting the parameter C_{bk}, which represents sediment inputs associated with bank and bar erosion during floodplain recession, produces a smoother and more progressive increase in sand concentration that better reproduces the secondary peaks observed during the falling limb of the hydrograph.

In addition, the calibrated values of K_{bed} remain very small ($K_{bed} \ll 1$), suggesting that the contribution of riverbed erosion to the simulated suspended sand flux in the main stem is limited under the conditions represented in the model. These results therefore suggest a dominant contribution of floodplain-related recycling processes to the observed resuspension signal, although a minor contribution from riverbed erosion cannot be fully excluded.

Revised text (Section 6.3):

Lastly, the distinction between E_{bk} *and* E_{bed} *is partly supported by the calibration experiments. When bed erosion is activated (*$k_{bed} > 0$*), the model tends to generate rapid and abrupt peaks in simulated sand flux once the transport capacity exceeds the available sand load, whereas adjusting*

C_{bk} produces a smoother and more progressive increase in sand concentration that better reproduces the secondary peaks observed during the recession phase. The calibrated values of K_{bed} remain very small ($K_{bed} \ll 1$), suggesting that riverbed erosion contributes only weakly to the simulated suspended sand flux in the main stem, although a minor contribution cannot be fully excluded.

Additional references

Armijos, E., Crave, A., Espinoza, R., Fraizy, P., Dos Santos, A. L. M. R., Sampaio, F., De Oliveira, E., Santini, W., Martinez, J. M., Autin, P., Pantoja, N., and Oliveira, M.: Measuring and modeling vertical gradients in suspended sediments in the Solimões/Amazon River, Hydrol. Process., 31, 654–667, https://doi.org/10.1002/hyp.11059, 2016.

Bouchez, J., Métivier, F., Lupker, M., Maurice-Bourgoin, L., Perez, M., Gaillardet, J., and France-Lanord, C.: Prediction of depth-integrated fluxes of suspended sediment in the Amazon River: particle aggregation as a complicating factor, Hydrol. Process., 25, 778–794, https://doi.org/10.1002/hyp.7868, 2011.

Guyot, J. L., Jouanneau, J. M., and Wasson, J. G.: Characterisation of river bed and suspended sediments in the Rio Madeira drainage basin (Bolivian Amazonia), Journal of South American Earth Sciences, 12, 401–410, https://doi.org/10.1016/S0895-9811(99)00020-8, 1999.

Nordin, C. F., Meade, R. H., Curtis, W. F., Bosio, N. J., and Landim, P. M. B.: Size distribution of Amazon River bed sediment, Nature, 286, 52–53, https://doi.org/10.1038/286052a0, 1980.

Park, E. and Latrubesse, E. M.: Modeling suspended sediment distribution patterns of the Amazon River using MODIS data, Remote Sensing of Environment, 147, 232–242, https://doi.org/10.1016/j.rse.2014.03.013, 2014.

Vanoni, V. A.: Sediment studies in the Brazilian Amazon River basin, World Meteorological Organization, Geneva, Switzerland, 1979.

Vanoni, V. A.: Sediment studies in the Brazilian Amazon River basin, Report KH-P-168, W.M. Keck Laboratory of Hydraulic and Water Resources, California Institute of Technology, Pasadena, California, USA, 1980.