Response to Referee Comment #1 (RC1)

Manuscript:

Long-term hydro-sediment 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.

30 March 2026

We thank Referee #1 for its rigorous and detailed review of our manuscript. The comments directly target the core methodological challenges of the study, the separation between calibration and validation, the treatment of equifinality, and the quantification of uncertainty in the reported floodplain budgets, and have led to substantial improvements in the manuscript. We have revised the manuscript accordingly and address each comment in detail below.

The referee comments are reproduced in italic, followed by our responses.

Sincerely,
W. Santini et al.

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	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 S1) 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.

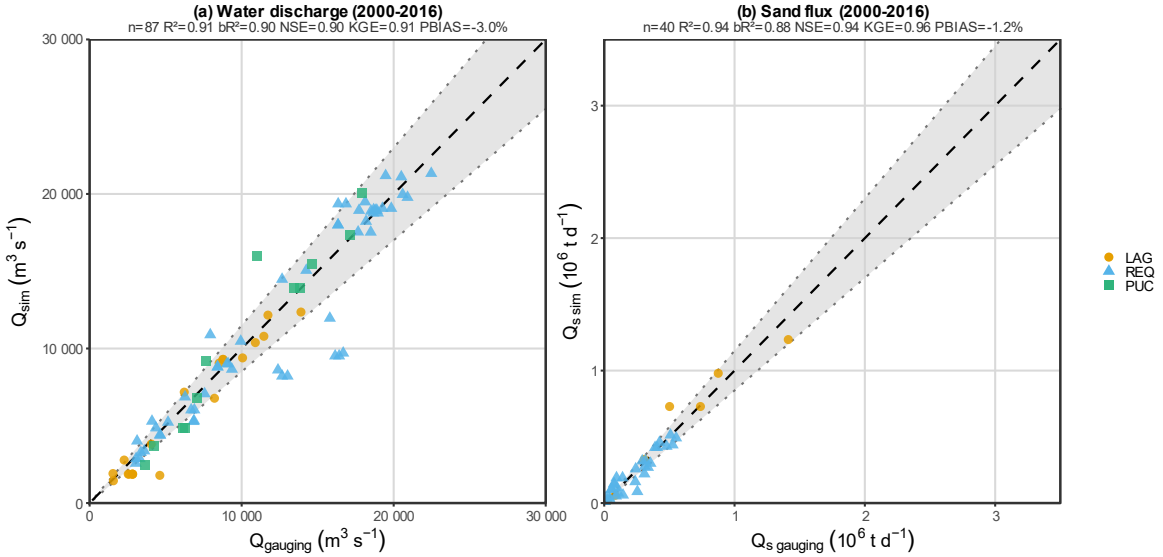

Figure RC1.1: Comparison of simulated and (a) measured river discharge and (b) suspended sediment flux at gauging stations in the Ucayali basin. Simulated values ($\boldsymbol{Q_{sim}}$, $\boldsymbol{Q_{s\ sim}}$) correspond to daily outputs from the SWAT-Amazon model extracted at the dates of field 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 Supplementary Material (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 Supplementary Material 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 (Supplementary Material, 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-section. 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. While such datasets would provide more direct constraints on internal floodplain processes, 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 6.3.1. A fully independent validation of internal floodplain processes therefore remains an important direction for future work.

Indeed, 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 hydro-system, namely discharge, water levels, velocities, rating curves, and sediment fluxes. As clarified in the revised manuscript, model performance is evaluated over independent time periods for discharge (Table RC1.1) and sand fluxes (Table RC1.2) and against extensive field measurements of water discharge and sand fluxes (Fig. RC1.1), providing strong constraints on the integrated behaviour of the river–floodplain system.

Furthermore, in large lowland Amazon rivers, floodplain water levels are strongly controlled by main-channel dynamics through hydraulic connectivity during overbank flooding (e.g. Rudorff et al., 2014a). As a result, floodplain stages and storage are not independent from main-channel conditions. In the modelling framework, floodplain water level dynamics emerge from exchanges with the main channel and the overall water balance. Therefore, inconsistencies in floodplain storage or exchanges would necessarily propagate to main-channel hydraulics through mass conservation, affecting discharge, water levels, velocities, and rating curves, which are explicitly evaluated here. Most importantly, floodplain water levels simulated with a closely related SWAT-Amazon configuration were previously evaluated against satellite altimetry observations across the Amazon Basin (Guilhen et al., 2022), showing good agreement at large spatial scales. Although not strictly identical to the present configuration, this constitutes the strongest available independent evidence that the modelling framework correctly reproduces large-scale floodplain storage dynamics.

Overall, model evaluation combines (i) evaluation of discharge and sediment fluxes over independent time periods, (ii) comparison with an extensive *in situ* dataset of direct measurements of water discharge and sediment fluxes, and (iii) multiple hydraulic diagnostics to constrain 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 – validation campaigns. The evaluation conducted here primarily focuses on variables that directly constrain main-channel hydrology and sediment transport. Although the present study does not provide a dedicated independent validation of internal floodplain dynamics, previous work using an earlier version of the SWAT-Amazon model showed good agreement between simulated floodplain stages and satellite altimetry observations (Guilhen et al., 2022).

These results demonstrate the relevance of the proposed framework for integrated monitoring of hydro-sedimentary fluxes. This is further supported by the increasing availability of satellite data over time, as illustrated in Fig. 8b, where missing data are mainly observed 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. The floodplain activation threshold h_f and the riverbed sand diameter d_b display relatively well-defined optima close to the calibrated values, indicating that they are effectively constrained by the available observations. By contrast, the parameters controlling floodplain geometry and flow resistance (C_{nfp}, k_{fp}, n) show reach-dependent sensitivity, ranging from weakly identifiable to moderately constrained depending on reach hydraulic conditions. However, since h_f is well identified across all reaches, the hydro-sedimentary budgets remain robustly constrained despite the equifinality of secondary floodplain parameters.

For d_b, the dotty plots also reveal secondary optimum(s) at unrealistically large values, which precludes the suspension onset. This alternative solution(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. This result highlights an important limitation of purely statistical calibration: without observational constraints, acceptable goodness-of-fit can be obtained for parameter values that are not physically meaningful.

Overall, the equifinality analysis confirms that the key parameters driving the sediment budget (h_f, d_b) are well constrained by the available observations, while secondary floodplain parameters show wider behavioural ranges consistent with their low Sobol sensitivity indices. 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.

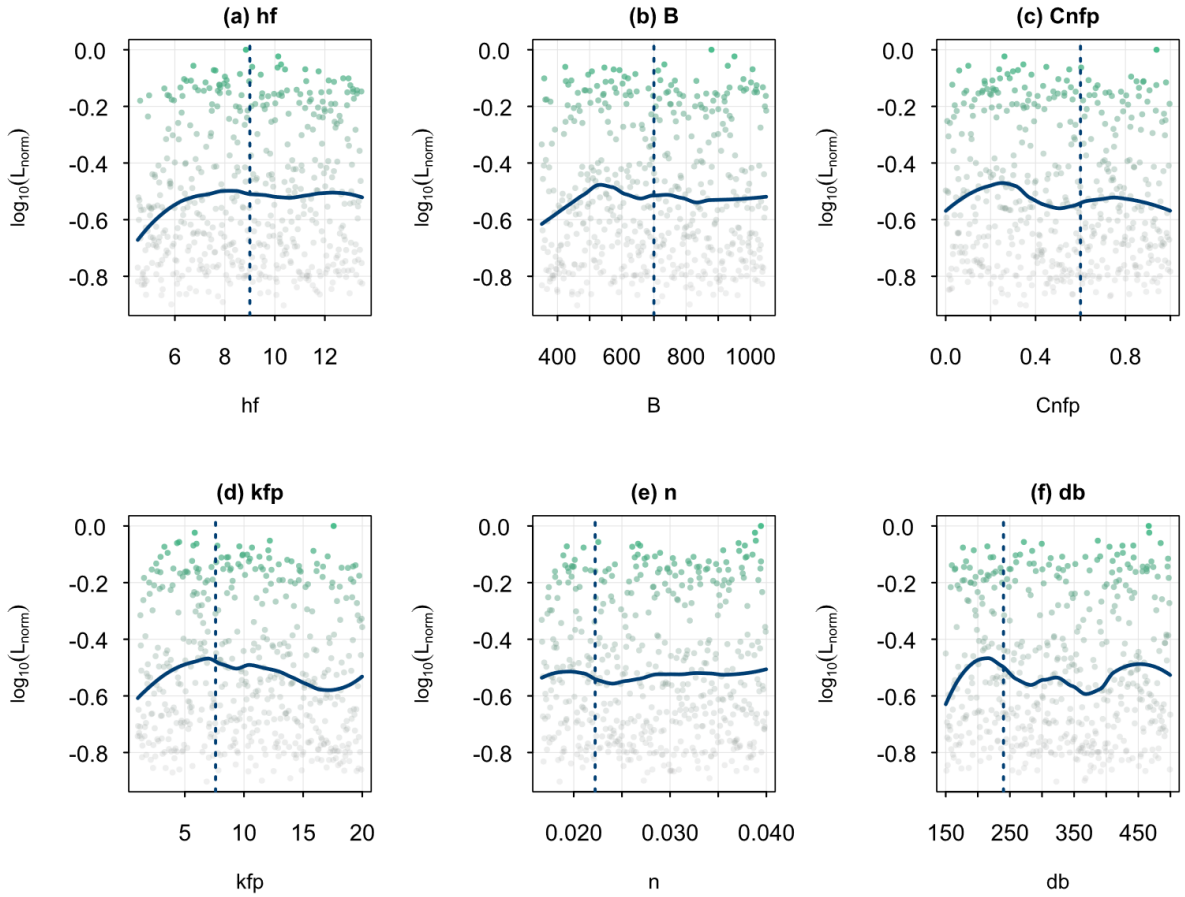

Figure RC1.3: Example of dotty plots for sub-basins 5, showing normalised GLUE likelihood versus parameter value for the Sobol-influential parameters. Green shading indicates higher likelihood. The blue curve is a loess smoother. Dashed blue vertical lines mark the calibrated values. Note the bimodal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

The manuscript was revised as follows:

6.1.3 Equifinality analysis

GLUE-based dotty plots (Beven and Binley, 1992) were produced by sampling the Sobol-influential parameters over wide ranges (Supplementary Material S11). Results confirm the Sobol hierarchy: h_f, B *and* d_b *are well identified with optima consistent with calibrated values, whereas* C_{nfp} *and* k_{fp} *and* n *exhibit reach-dependent sensitivity, ranging from weakly constrained to moderately identifiable depending on local hydraulic and morphological conditions. 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, the key parameters driving the sediment budget (*h_f, d_b*) are well constrained by the available observations, while secondary floodplain parameters show wider behavioural ranges consistent with their low Sobol sensitivity indices. These findings justify the use of expert measurement-informed priors for uncertainty propagation (Supplementary Material S12).*

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 the coarser 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}, h_f, 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 not sampling a too large parameters set across, 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).

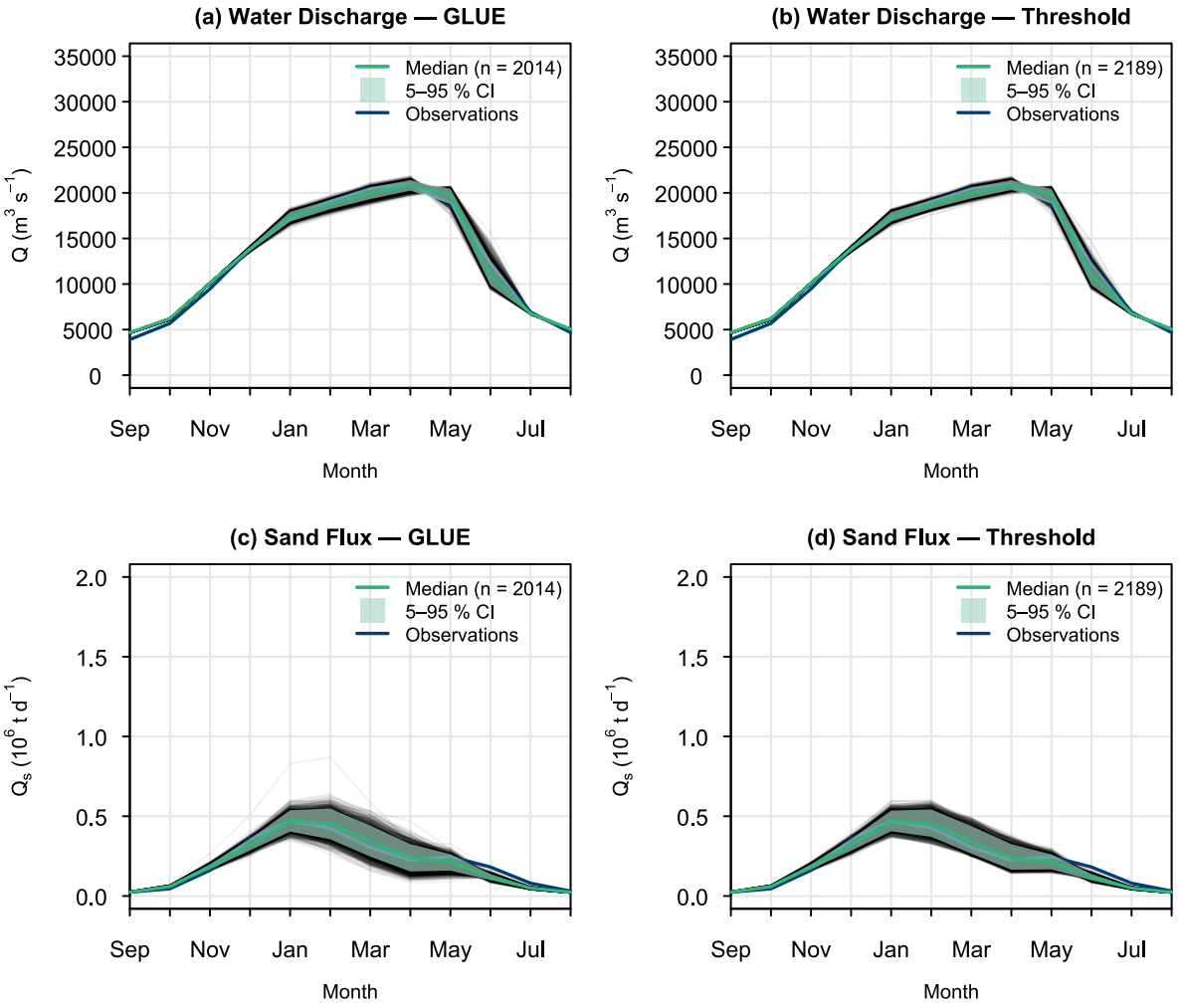

Figure RC1.3: Interannual cycles of water discharge and sand flux at the Ucayali outlet (subbasin 21, Requena station), comparing GLUE and threshold-based uncertainty propagation methods. (a, b) Water discharge Q ($m^3\ s^{-1}$) and (c, d) suspended sand flux Qs ($10^6\ t\ d^{-1}$), computed over the 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.

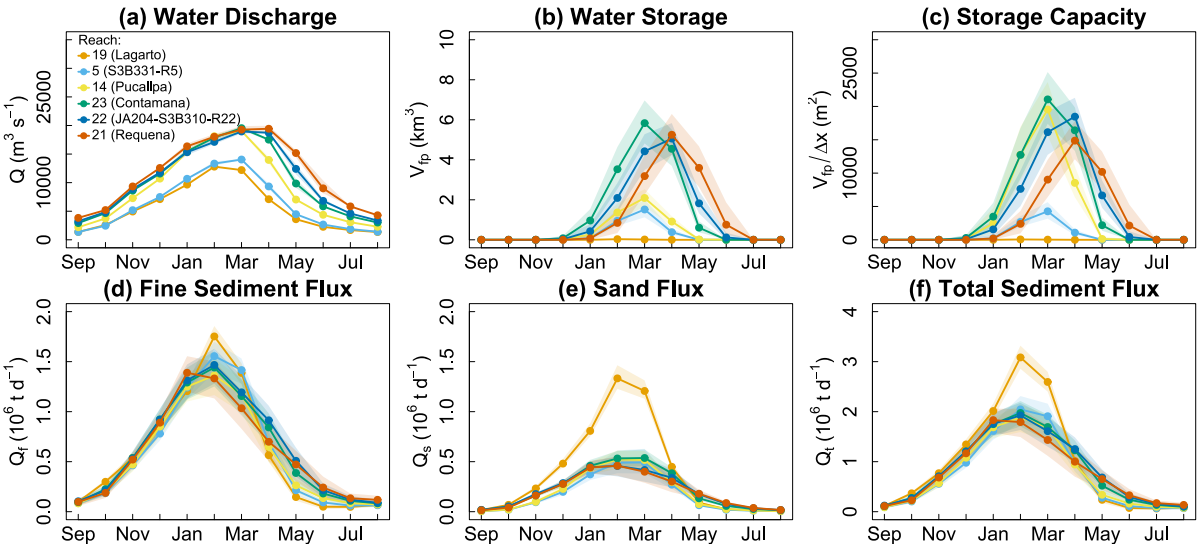

Figure RC1.4: Interannual Figure 11: Interannual water (1983¬2019) and sediment (2000¬2019) balances along the Ucayali River. Lines show interannual monthly means and shaded bands represent the 5th–95th percentile envelope of the behavioural GLUE ensemble (2500 runs). (a) Water Discharge (b) Floodplain water storage (V_{fp}) in km3. (c) Normalization of V_{fp} by the reach's length (Δx) 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 (in %) 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 (Supplementary Material, Section 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 hf, emphasizing the need for accurate stage–discharge rating curves to constrain this parameter reliably (Sect. 6.2.3).

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.

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

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:

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.

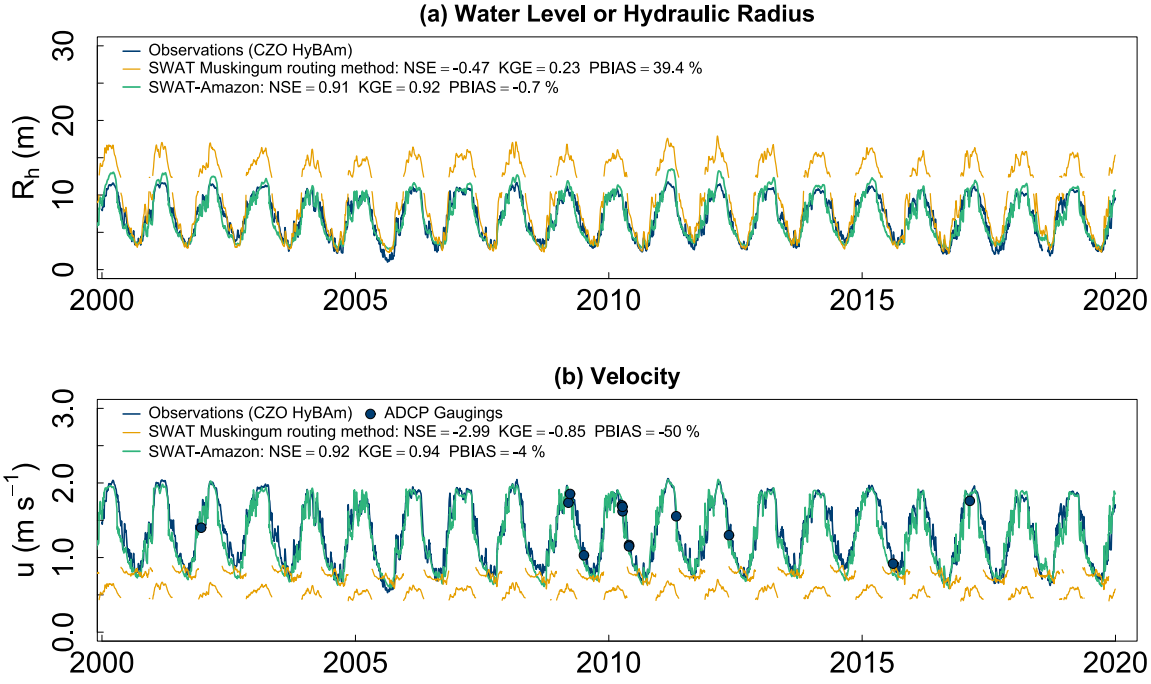

Figure RC1.5: Figure 4, modified with the new color palette.

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.