Supplement of Hydrol. Earth Syst. Sci., 30, 3367–3397, 2026
https://doi.org/10.5194/hess-30-3367-2026-supplement

Hydrology and Earth System Sciences
Open Access
EGU

Supplement of

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

William Santini et al.

Correspondence to: William Santini (william.santini@ird.fr)

Supplementary material

S1: Particle-size distributions (PSD) and vertical grain-size profiles

Particle-size distribution (PSD) measurements available for the Ucayali River and other white-water rivers of the Amazon Basin consistently show a predominantly bimodal structure, with a fine mode composed mainly of silts and clay aggregates and a sand mode centred in the very fine sand range (Fig. S1a). Vertical grain-size profiles (Fig. S1b) further illustrate the distribution of these fractions within the water column, with fine sediments remaining well distributed due to their low settling velocities, while sand concentrations increase towards the near-bed region. These observations support the use of a simplified two-fraction representation of suspended sediment dynamics at the basin scale, as adopted in the present study.

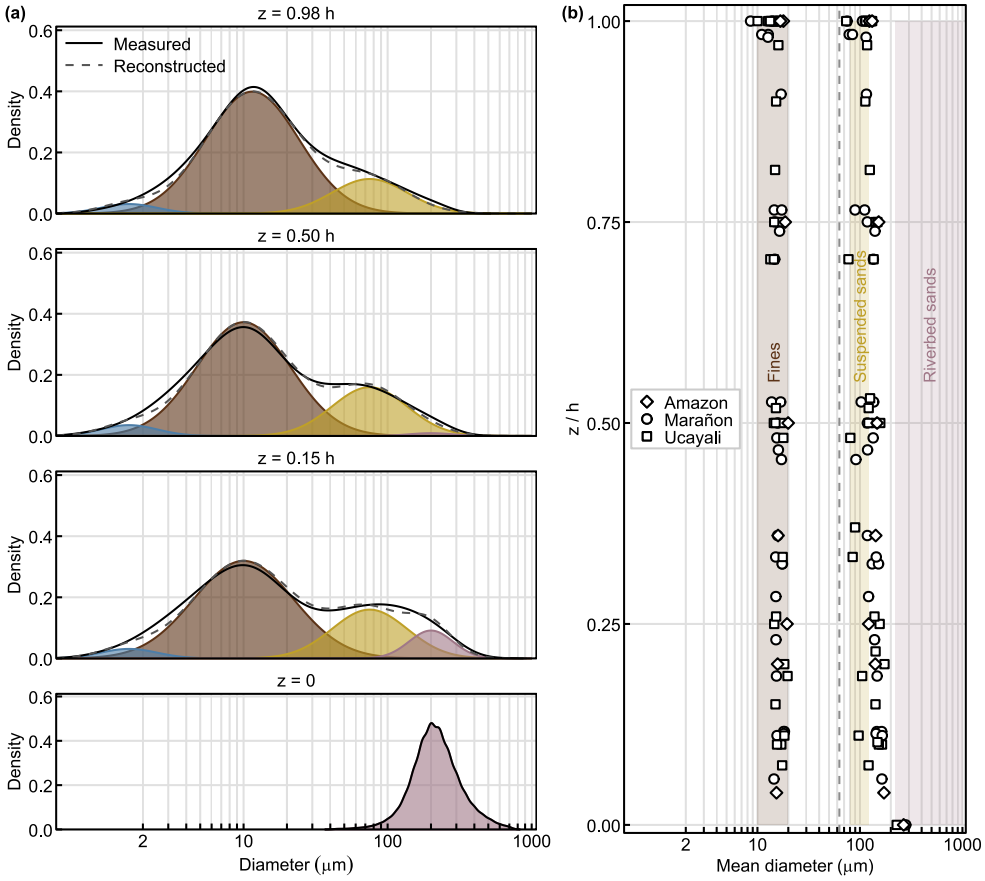

Figure S1: Typical PSD of large white-water rivers in the Amazon Basin (adapted from Santini et al., 2019). (a) Particle-size distributions (PSD) measured in suspension at different relative depths and in the riverbed (Requena station, Ucayali River, 16 March 2015). The grey line shows the measured distribution, 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.

S2: Measurement protocols (HyBAm stations)

The hydrological dataset used in this study was collected within the framework of the Critical Zone Observatory HyBAm (Hydrogeochemistry of the Amazon Basin) (https://hybam.obs-mip.fr), an international long-term monitoring program of water and material fluxes in the Amazon, Orinoco, and Congo basins, as well as in two additional basins in French Guiana.

At each gauging station, water levels and surface index concentration samples are collected according to a calendar-based sampling scheme. In the stations considered here, one sample was collected every ten days during the period 2003–2013. Sediment monitoring protocols were subsequently enhanced, including higher sampling frequencies at Requena between November 2012 and June 2013 (one sample every two days, plus three additional repetitions every ten days), and five-day sampling intervals during the wet season between July 2013 and September 2015. Surface samples are collected 20-50 cm below the water surface at a fixed location in the main channel.

For establishing stage-discharge relationships and relating surface index concentrations to cross-sectional average concentrations for sediment flux calculations, 82 field campaigns were conducted between 2003 and 2019 at the stations considered in this study. These campaigns included numerous discharge measurements and sediment sampling across river cross-sections and covered highly contrasted hydrological regimes, including extreme droughts (2005, 2010) and major floods (2012, 2015) (Espinoza et al., 2012, 2013; Marengo et al., 2016). Water discharge was measured using a 600 kHz Teledyne RDI Workhorse Rio Grande ADCP, coupled with a GPS sensor to correct for potential moving-bed bias. A detailed description of the measurement protocols is provided in Santini et al. (2019) and Santini (2020).

To estimate cross-sectional average suspended sediment concentration, vertical concentration profiles were obtained using a point-sampling method (Santini et al., 2019). Sampling was performed from a boat drifting along a streamline using a 650 ml sampler. (i) Before June 2010, sampling was conducted along three verticals dividing the cross-section into four subsections, typically with three samples per vertical (near-bed, mid-depth, and near-surface). Due to the difficulty of sampling near sandy riverbeds and the very low sand concentrations close to the surface, only the mid-depth samples were considered reliable for estimating suspended sand concentrations. (ii) After June 2010, the sampling protocol was improved by increasing the number of sampling levels along the vertical. Samples were typically collected at 0.98 h, 0.75 h, 0.5 h, 0.25 h, and occasionally 0.15 h and 0.1 h above the riverbed along three verticals across the channel. With this revised protocol, between 9 and 15 samples per cross-section were generally usable to estimate total, fine and sand concentrations.

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

S3: Time series of sand and fine fluxes inferred from index concentration and sediment discharge measurements

The point sampling method allows for accurate assessment of mean sand concentration when a sufficient number of samples are collected and the river cross-section is thoroughly explored using a well-designed grid of sampling points. However, these measurements are typically performed on an episodic basis, primarily due to operational constraints and cost considerations. Furthermore, as demonstrated in this study, there is no straightforward relationship between discharge and sand concentration in the Ucayali floodplain. As a result, only the surface-based monitoring of sand concentration, carried out by field observers under the CZO HyBAm framework, enables the construction of continuous time series of suspended sand fluxes.

At lowland stations, however, concentrations derived from this monitoring approach are subject to substantial uncertainty. This is because samples collected at the free surface, where turbulent mixing is significantly reduced, tend to exhibit low sand mass fractions. Although it is theoretically possible to determine a site-specific ratio α_s between these surface concentrations and the depth-averaged mean sand concentration using the model of Santini et al. (2019), the associated uncertainty remains high. This is particularly true for the range of Rouse numbers observed at the study sites ($0.3 < P_s < 0.5$), as also discussed by Santini et al. (2019).

In this study, we aim to reconcile the available data by combining the time series of cross-sectionally averaged total suspended sediment concentration, $C = \alpha\, C_\chi$ (mg l^{-1}) derived from surface index concentration monitoring $C_\chi(z \cong h)$ (mg l^{-1}), with the cross-sectional concentrations obtained from *in situ* point sampling. A time series of cross-sectionally averaged sand concentration C_s is derived from the monitored index concentration using an empirical relationship between C and C_s calibrated from gauging data (Fig. S2). The α ratios used were those computed by Santini et al. (2019).

It is important to note that the empirical relationship between total suspended sediment and sand concentrations was observed in only a few rivers of the Amazon Basin, and therefore cannot be generalized. This limitation highlights the need to explicitly model sand transport dynamics rather than rely solely on empirical extrapolation.

The fine sediment fluxes are calculated following the same method (Fig. S2).

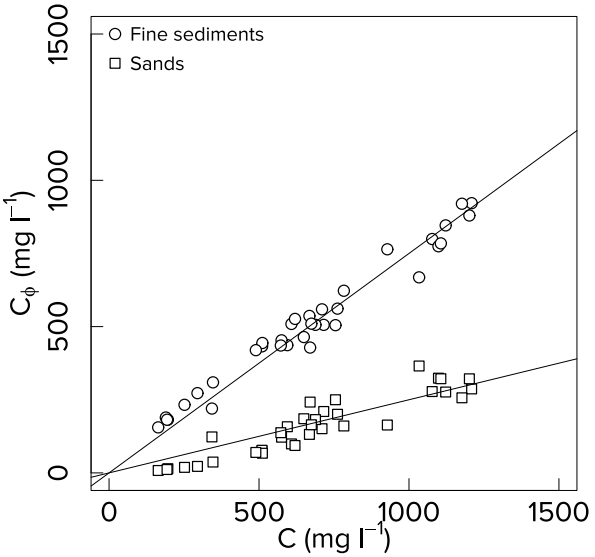

Figure S2: Empirical relationships at Requena between the total cross-sectionally averaged suspended sediment concentration C, and the corresponding cross-sectional concentrations C_ϕ of each sediment fraction, where $\phi = f$ denotes fine sediments $\phi = s$ denotes sands.

S4: Triangular floodplain cross-section

When the floodplain cross-section is assumed to be triangular, the water height h in the main channel can be obtained by solving the following second-order polynomial equation (the physically meaningful solution is the greater of the two roots), derived from basic geometric considerations:

$$\left(\frac{\Delta x_i}{2\tan(\theta_{fp,i})}\right)(h_i^t)^2 + \Delta x_i\left(B_i - \frac{h_{f,i}}{\tan(\theta_{fp,i})}\right)h_i^t + \Delta x_i\frac{h_{f,i}^2}{\tan(\theta_{fp,i})} = V_i^t\,, \quad \text{(S1)}$$

where θ_{fp} (rad) is a calibrated angle representing the riverward slope of the floodplain, which determines the floodplain width W_{fp}.

S5: Additional flow resistances

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. S3)**.** 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).

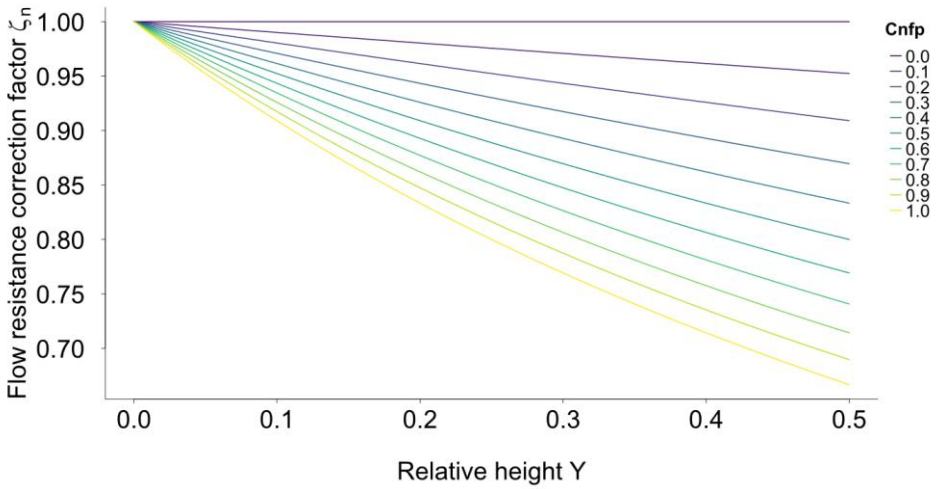

Figure S3: Effect of the coefficient C_{nfp} on the Manning coefficient when the floodplain is active.

S6: Additional flow resistances: impact on the velocity and transport capacity when the floodplain is active

In a 2d context, the transverse profiles of $\theta'$and C_b^* could be more affected than the lateral profile of depth-averaged velocity. The velocity declines induced by the floodplain drag, at least near the banks and on the bars, should lead to the sedimentation of finer materials (such as very fine sands, silts, and even clay aggregates) compared to a scenario without this additional flow resistance. According to the Yalin-Shields curve, this would result in an increase in the critical shear stress, which in turn reduces the bottom reference concentration C_b^* (Eq. 12). Furthermore, the Darcy-Weisbach factor could also decrease, leading to a reduction of θ'. The geometry of bedforms could also be impacted. Thus, the repartition of total shear stress θ between grain-related $\theta'$and bedform-related θ'' dimensionless shear stresses could also change. Indeed, according to the momentum concept, $\theta = \theta' + \theta''$, any increase or decrease in θ'' would impact θ', then C_b^* and consequently the transport capacity Q_s^*. Among the poorly understood other possible effects, the large and mesoscale secondary currents induced by the floodplain activation could weaken the turbulent velocity profiles in the upper part of the flow (e.g. Guo and Julien, 2008), resulting in a decline in velocities above a dipping point. This would lead to a reduction in the diffusivity ratio β, and thus an attenuation of the sand suspension (hence of Q_s^*) in this flow region.

To account for these complex effects on transport capacity Q_s^*, another shape factor γ (–) has been introduced to adjust the transport capacity $Q_{s_{cf0}}^*$ which was initially calculated using the corrections for $\theta'$and u_* corrections in Eqs. 29 and 30.

$$Q_{s_{cf}}^* = \gamma \; Q_{s_{cf0}}^* \, . \tag{S2}$$

In a 2d context, ζ_n can be also interpreted as a shape coefficient between the depth integrated velocity profiles $\overline{u_c}(y)$ and $\overline{u_{cf}}(y)$ in the transverse direction y:

$$\zeta_n = \frac{\frac{1}{B}\int_0^B \overline{u_{cf}}(y)\,dy}{\frac{1}{B}\int_0^B \overline{u_c}(y)\,dy}, \tag{S3}$$

and the shape correction factor γ relates to the following ratio:

$$\gamma = \frac{\frac{1}{B}\int_0^B Q^*_{s\,cf}(y)\,dy}{\frac{1}{B}\int_0^B Q^*_{s\,cf0}(y)\,dy}. \tag{S4}$$

Here, it was assumed that γ should be a function of ζ_n:

$$\gamma = \lambda\,\zeta_n^{\xi}, \tag{S5}$$

with $\lambda \leq 1$ a coefficient and ξ (–) an exponent to calibrate. As ζ_n is between 0 and 1, λ is not useful and this parameter was thus eliminated in order to simplify the model calibration. Thus, when the floodplain is active (*i.e.* when $h > h_f$), the attenuation of the transport capacity is calculated as follows:

$$Q^*_{s\,cf}{}_i^{\,t} = \zeta_{n\,i}^{\eta\,t}\,Q^*_{s\,cf0}{}_i^{\,t}, \tag{S6}$$

where the exponent η includes the term λ.

S7: Additional flow resistances: typical example of bed roughness influence in the Amazonian Basin

The Chazuta gauging station (−6.5704°, −76.1193°) on the Huallaga River (a tributary of the Marañón River), also monitored by the CZO HyBam, is a typical example of the influence of bed roughness on flow resistance, as well as an illustration of the increased flow resistance when the floodplain becomes active (Fig. S4).

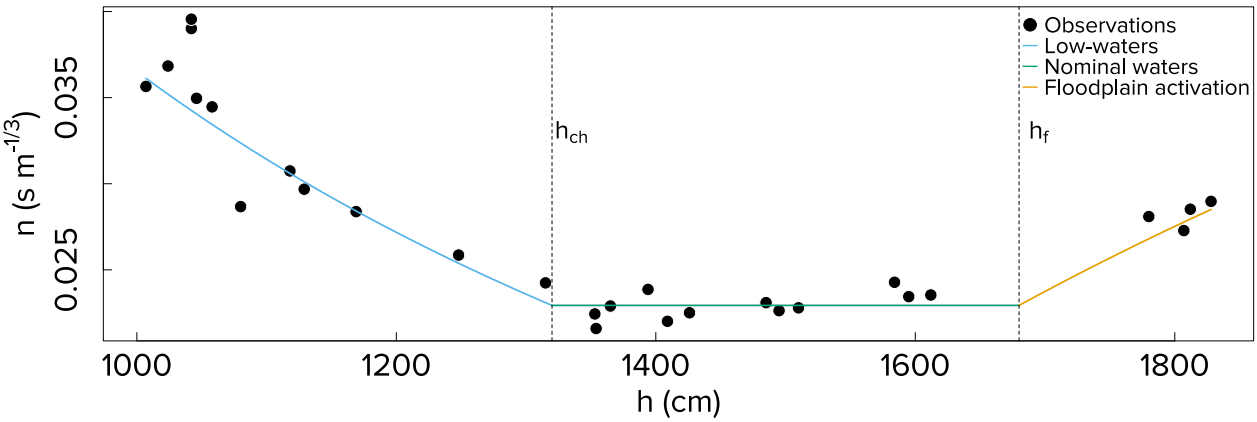

Figure S4: Typical evolution of the Manning coefficient with water level at the Chazuta station on the Huallaga River. Black dots indicate Manning values derived from ADCP gaugings, the blue line shows Manning modelled according to Eq. 32, and the orange line shows Manning modelled according to Eq. 28. Dashed vertical lines represent the water levels $\boldsymbol{h_{ch}}$ and $\boldsymbol{h_f}$.

S8: Negligible optical contribution of sand to surface reflectance

No significant relationship was found between surface sand concentration and the reflectance ratio measured during field campaigns (Fig. S5). This confirms that the backscattering contribution of sand particles is negligible compared to that of fine sediments, as predicted by Mie theory (Pinet, 2017). Furthermore, the relationship between reflectance ratio and sediment concentration improves when sand is excluded from the concentration estimates.

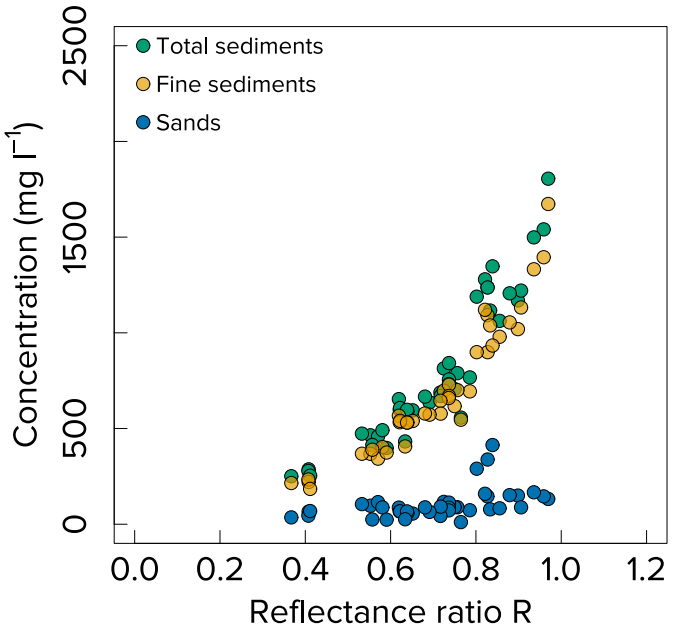

Figure S5: Concentrations of total suspended sediments (green dots), fine sediments (orange dots) and sand fractions (blue dots) at the water surface plotted against the reflectance ratio (NIR/Red) measured during calibration–validation campaigns using hyperspectral radiometers.

S9: Calibration – Validation

Model calibration for water discharge (Table S1) 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 S1), showing consistent performance metrics across all periods.

Table S1. 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, 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 S2). 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 S2. 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 the sampling time step of surface concentration measurements.

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. S6). 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. S6).

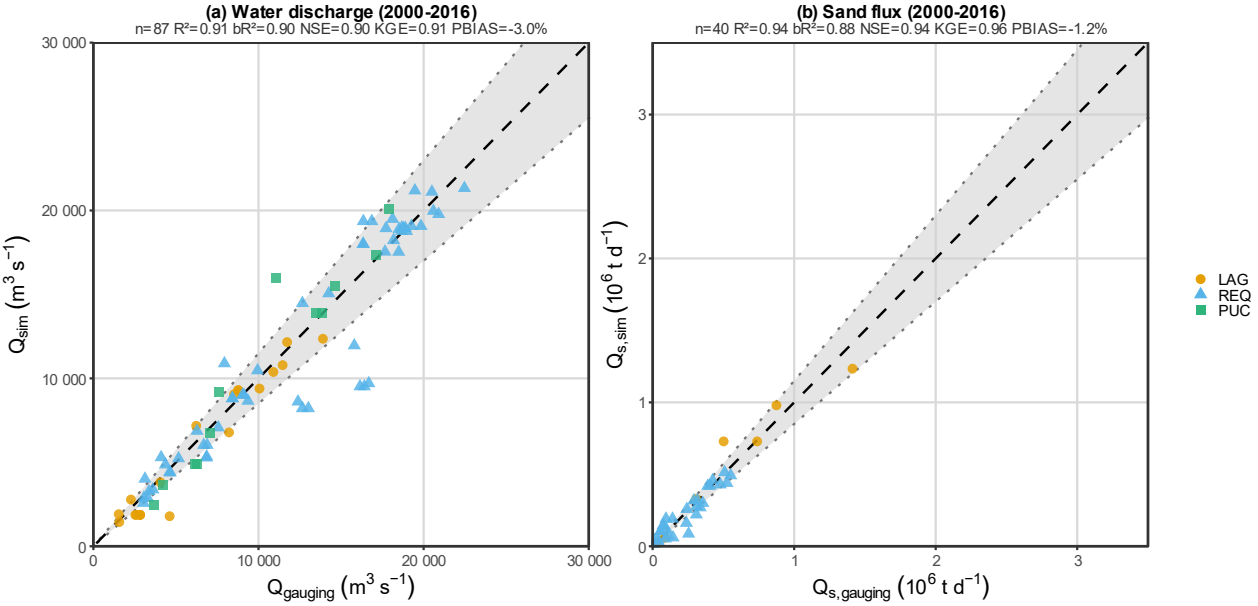

Figure S6: Comparison of simulated and measured river discharge (a) and suspended sediment flux (b) at gauging stations in the Ucayali basin. Simulated values (Q_{sim}, $Q_{s,sim}$) correspond to daily outputs from the SWAT-Amazon model extracted at the dates of field measurements. Observed values ($Q_{gauging}$, $Q_{s,gauging}$) 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.

S10: Sobol sensitivity analysis

The Sobol analysis was performed over the 2009-2015 period, with wide range parameters

Table S3. Parameter range adopted for the Sobol sensitivity analysis

Water Routing Parameter	Unit	Range	Sand routing Parameter	Unit	Range
h_f	(m)	[10; 20]	d_s	(µm)	[60; 120]
B	(m)	[500; 900]	d_b	(µm)	[180; 300]
k_{fp}	(–)	[3; 10]	β_s	(–)	[0.2; 1]
n	(s $m^{-1/3}$)	[0.02; 0.04]	K_{bed}	(–)	[0; 0.5]
C_{nfp}	(–)	[0; 2]	C_{bk}	(t m^{-3})	[100; 500]
			η	(–)	[1.5; 3.5]

S11: Equifinality

A dedicated equifinality analysis using GLUE-based dotty plots was performed (Fig. S7, S8 and S9) by sampling the main routing and sand transport parameters identified as influential in the Sobol analysis (h_f, B, C_{nfp}, k_{fp}, n, d_b) 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, 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.

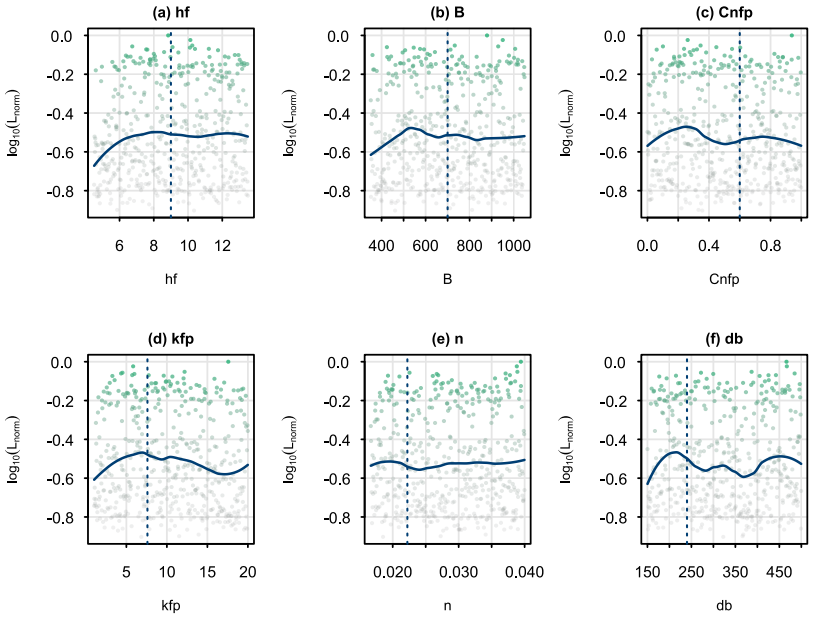

Figure S7: Example of dotty plots for sub-basin 5, showing normalised GLUE likelihood versus parameter value for the Sobol-influential parameters. Green shading indicates higher likelihood. The blue curve is a loess smoother. Dashed blue vertical lines mark the calibrated values.

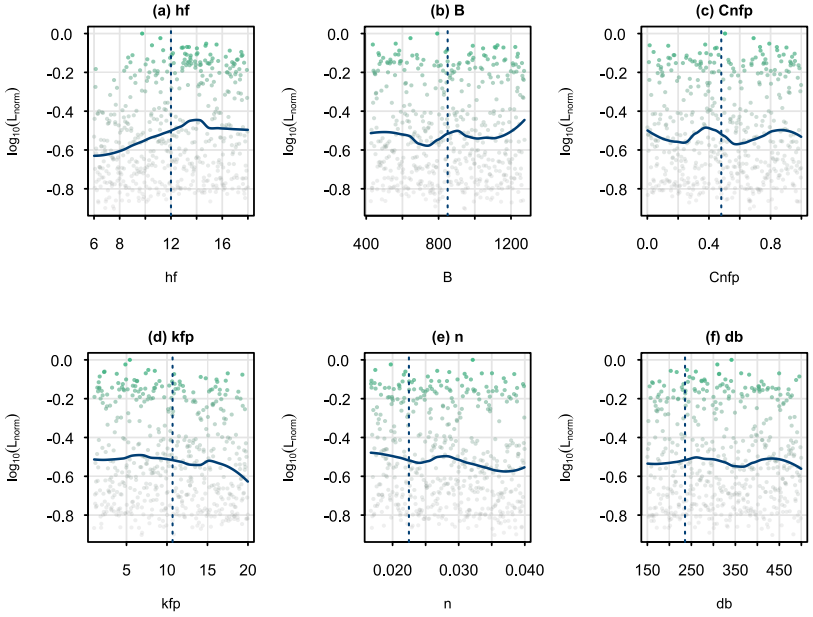

Figure S8: Example of dotty plots for sub-basin 22, showing normalised GLUE likelihood versus parameter value for the Sobol-influential parameters. Green shading indicates higher likelihood. The blue curve is a loess smoother. Dashed blue vertical lines mark the calibrated values.

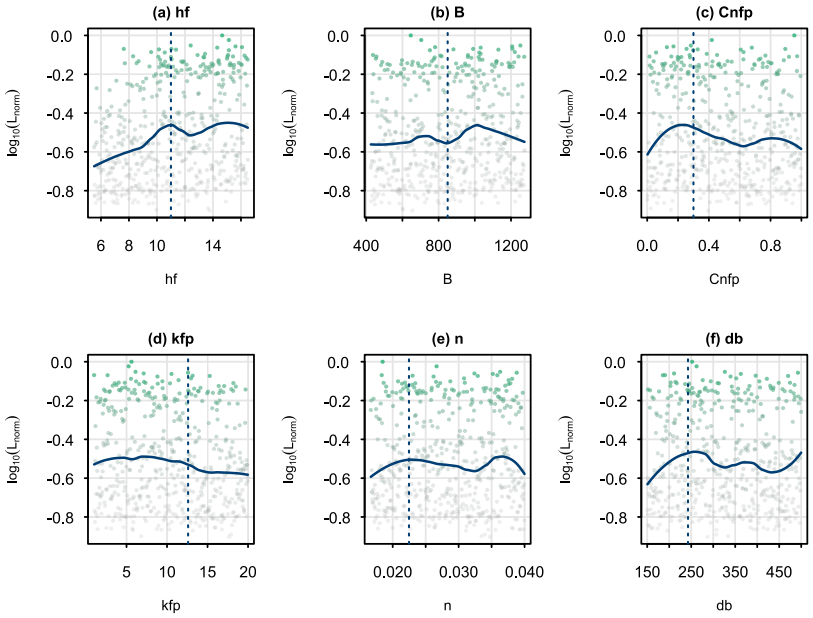

Figure S9: Example of dotty plots for sub-basin 23, showing normalised GLUE likelihood versus parameter value for the Sobol-influential parameters. Green shading indicates higher likelihood. The blue curve is a loess smoother. Dashed blue vertical lines mark the calibrated values.

S12: Uncertainty propagation

S12.1 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.

S12.2 Parameter sampling strategy

Parameter uncertainty was represented using parameter-specific prior distributions (Table S4). 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 S4. Prior distributions used for uncertainty propagation of SWAT-Amazon routing parameters.

Parameter	Unit	Distribution	Sampling	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.

S12.3 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 S5). 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.

Table S5. 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

S12.4 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

S12.5 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. Observational uncertainty is implicitly accounted for within the GLUE framework, as the informal 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.

Finally, a detailed synthesis of hydro-sedimentary budget indicators, including GLUE-based uncertainty bounds, is provided in Table S6, allowing a quantitative assessment of flux partitioning, storage, and associated uncertainties along the Ucayali system.

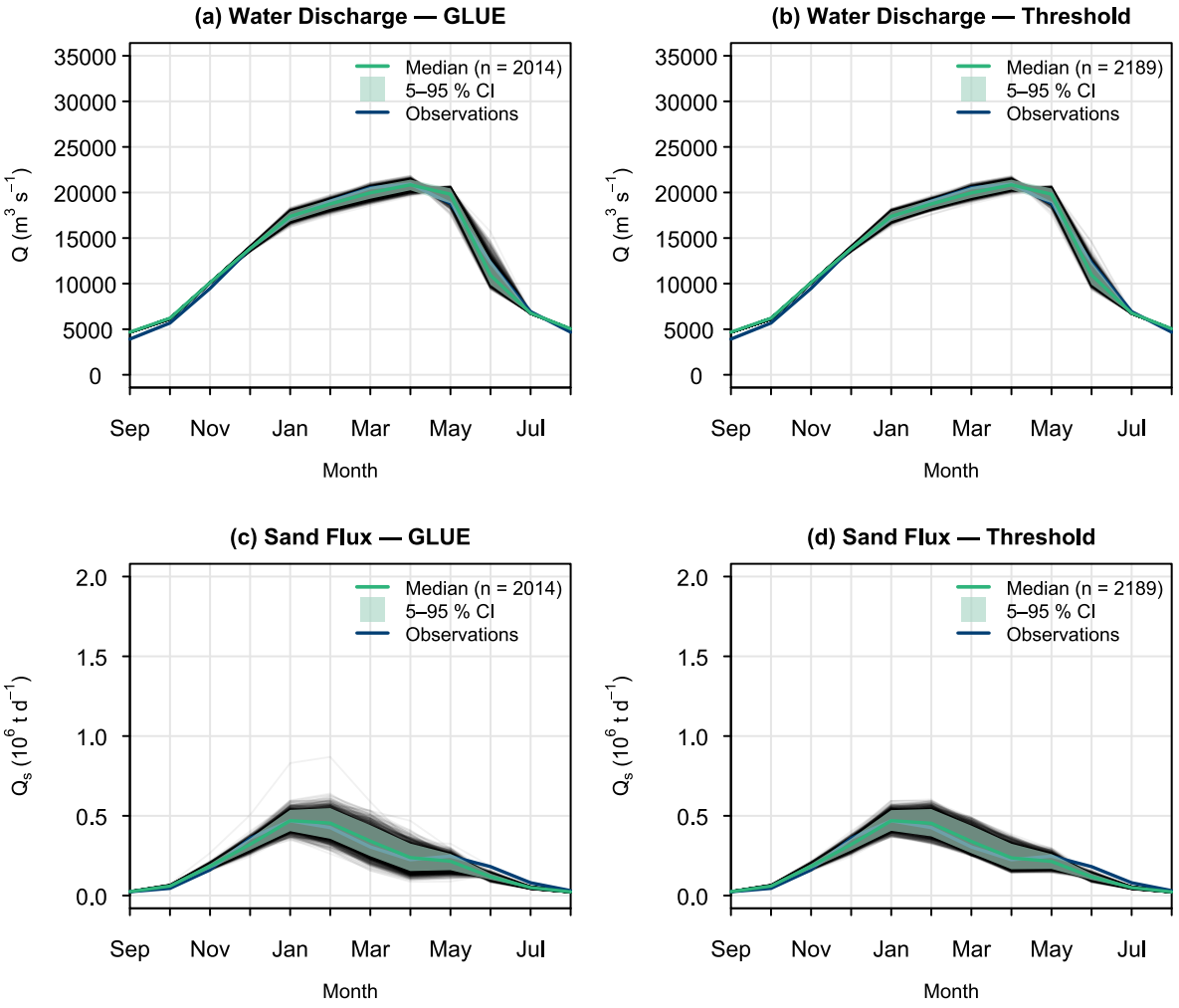

Figure S10: Interannual cycles of water discharge and sand flux at the Ucayali outlet (subbasin 21, Requena station), comparing GLUE and threshold-based uncertainty propagation methods. (a, b) Water discharge Q ($m^3 s^{-1}$) and (c, d) suspended sand flux Qs (10^6 t d^{-1}), computed over the September–August 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.

Table S6. 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	—