Larhyss Journal, ISSN 1112-3680, n°66, Jun 2026, pp. 31-75
© 2026 All rights reserved, Legal Deposit 1266-2002

A PROBABILISTIC TWO-REGIME GOVERNING MODEL FOR PREDICTING INFLOW SUSPENDED SEDIMENT CONCENTRATION IN DAM RESERVOIRS

ACHOUR B.[1*]***, AMARA L.***[2]***, MEHTA D.***[3]

[1] Professor, Research Laboratory in Subterranean and Surface Hydraulics (LARHYSS), University of Biskra, PO Box 145 RP, Biskra, Algeria
[2] Professor, Department of Civil Engineering and Hydraulics, Faculty of Science and Technology, University of Jijel, PO Box 98, Ouled Aissa, Jijel, Algeria
[3] Assistant Professor, Civil Engineering Department, Dr. S. & S. S. Ghandhy Government Engineering College, Surat, Gujarat, 395001, India

(*) bachir.achour@larhyss.net

Research Article – Available at http://larhyss.net/ojs/index.php/larhyss/index
Received June 12, 2025, Received in revised form April 21, 2026, Accepted April 24, 2026

ABSTRACT

Accurate prediction of inflow suspended sediment concentration is a critical prerequisite for the effective management of sedimentation in dam reservoirs. However, most existing approaches rely on classical formulations that are not predictive by construction and suffer from fundamental conceptual and numerical limitations. In particular, widely used log-scale models introduce an arbitrary offset parameter to circumvent the indetermination of the logarithm at zero concentration, implicitly merge zero-transport and active-transport conditions into a single regime, and induce bias when back-transforming predictions to the physical concentration scale. Likewise, traditional sediment rating curves of the form $C = aQ^b$ lack dimensional homogeneity, provide no probabilistic interpretation, and are unable to represent the intermittent, event-driven nature of sediment transport.

This study develops a new probabilistic two-regime governing model for predicting inflow suspended sediment concentration that explicitly addresses these shortcomings. The proposed framework is derived from first principles by recognizing that sediment time series exhibit two physically distinct regimes: (1) a zero or negligible transport regime, and (2) an active transport regime characterized by strictly positive concentrations. The occurrence of zero sediment is modeled probabilistically through a Bernoulli process, whose probability is linked to hydrometeorological predictors via a logistic formulation. Conditional on active transport, sediment magnitude is modeled using a positive-regime formulation based on an inverse hyperbolic sine transformation, which behaves linearly at low concentrations and logarithmically at high concentrations while remaining well-defined at zero without introducing any arbitrary correction parameter.

The resulting predictive model combines the probability of occurrence and the conditional sediment magnitude into a single closed-form forecast expressed as a probability-weighted expectation. All model components are derived explicitly, parameters are estimated through well-defined likelihood-based or least-squares procedures, and the scale parameter governing the transformation is shown to have a clear physical interpretation. A fully worked numerical example demonstrates the transparency, internal consistency, and practical implementation of the proposed approach.

Compared to classical log-scale and rating-curve methods, the new two-regime model offers several decisive advantages: it removes the need for ad hoc numerical fixes, preserves dimensional consistency, separates occurrence and intensity mechanisms, provides a probabilistic interpretation of predictions, and directly targets future sediment concentration at a specified lead time. The proposed framework therefore constitutes a robust and physically consistent alternative for predictive sediment modeling and reservoir sediment management.

Keywords: Inflow suspended sediment concentration, Reservoir sedimentation, Two-regime predictive model, Probabilistic sediment modeling, Sediment occurrence probability, Bernoulli-logistic model.

INTRODUCTION

Sedimentation remains one of the most critical long-term challenges affecting the sustainability, safety, and operational efficiency of dam reservoirs worldwide, especially in arid and semi-arid regions (Remini and Bensafia, 2016; Achour et al., 2024; Chadee et al., 2024). Once sediment enters a reservoir, it progressively reduces storage capacity (Benfetta et al., 2016), alters hydraulic performance, compromises flood control (Abd Rahman et al., 2023; Zegait and Pizzo, 2023; Baudhanwala et al., 2023; Ben Said et al., 2024; Ezz, 2025; Do et al., 2025; Athmani et al., 2025), and shortens the effective lifespan of hydraulic infrastructure (Mezenner et al.; 2022; Mehta et al., 2023; Verma et al., 2023; Panchal and Suryanarayana, 2025). Accurate prediction of inflow suspended sediment concentration therefore constitutes a central requirement for reservoir sediment management, sediment routing operations, and strategic planning (Remini, 2010; Remini, 2017; Bougamouza et al., 2020; Remini, 2022).

Early foundational work established sediment continuity principles linking sediment inflow, deposition, and storage loss in reservoirs (Remini, 2003; Remini and Toumi, 2017; Remini et al., 2019; Toumi an Remini, 2020). Classical formulations express sediment mass flux as the product of discharge and suspended sediment concentration, a standard representation in sediment monitoring and hydrosedimentology (Walling, 1977; USGS, 2006). These concepts were later embedded within sediment budget frameworks that account for inflow, outflow, and deposition processes (Graf, 1984; Morris and Fan, 1998). Relationships between sediment mass, bulk density, porosity, and deposited volume were formalized through soil and sediment mechanics (Terzaghi et al., 1996), enabling the conversion between mass-based and volume-based sediment estimates (Bear, 1972; Morris and Fan, 1998; Chadee et al., 2023b).

Subsequent studies incorporated these relationships into operational sedimentation forecasting and reservoir capacity loss estimation (ICOLD, 1989; Palmieri et al., 2003; Annandale, 2013). These approaches are now standard in long-term reservoir sediment management and planning studies, where sediment inflow constitutes a key driving variable.

At the river and catchment scale, suspended sediment concentration has traditionally been related empirically to hydrological variables. Classical sediment rating curves express sediment load or concentration as a power-law function of discharge, typically written in the form $C = \mathrm{a}Q^{\mathrm{b}}$ (Walling, 1977; Asselman, 2000; 2024; Mehta et al., 2021). Similar power-law relationships have been developed between turbidity and suspended sediment concentration (Lewis, 1996; Sutherland et al., 2000). For calibration purposes, these nonlinear relationships are commonly linearized using logarithmic transformations, allowing parameter estimation through ordinary least squares or related regression techniques (Helsel and Hirsch, 2002; Montgomery et al., 2012).

Although widely applied, these formulations suffer from important conceptual limitations. Power-law rating curves are not dimensionally homogeneous, implicitly assume stationarity, and are unable to represent the highly intermittent and event-driven nature of sediment transport (Mehta et al., 2024; Mehta et al., 2017). Moreover, logarithmic linearization requires ad hoc numerical adjustments to handle zero or near-zero concentrations, an issue that has persisted for decades in sediment modeling practice.

The logarithmic transformation of suspended sediment concentration has been extensively adopted to stabilize variance, reduce skewness, and facilitate Gaussian error assumptions in regression and forecasting models. In this classical framework, sediment concentration is never modeled directly; instead, a log-transformed variable is introduced, calibrated, and subsequently back-transformed to recover concentration values.

While effective in certain contexts, this approach introduces several well-documented issues: (1) the need to introduce an arbitrary offset parameter to avoid the indetermination of the logarithm at zero concentration, (2) the implicit and uncontrolled treatment of zero sediment events, (3) systematic bias introduced during back-transformation, particularly at low concentrations, and (4) the assumption that zero transport and active transport belong to a single continuous regime. These limitations are not merely numerical inconveniences; they reflect a deeper conceptual mismatch between model structure and sediment transport physics.

More recently, data-driven and machine-learning approaches have been introduced to predict sediment concentration and sediment load, often using nonlinear regressors, neural networks, or other flexible function approximators (Karpatne et al., 2017). These models can capture complex nonlinear relationships but frequently lack physical interpretability and may violate fundamental sediment continuity principles if unconstrained.

To address this, hybrid approaches combining data-driven components with physical constraints have been proposed, including physics-informed loss functions and hybrid sediment transport closures (Raissi et al., 2019; Reichstein et al., 2019). In parallel, morphodynamic theory continues to rely on sediment continuity equations, such as the Exner equation and its simplified forms, to describe bed evolution and sediment transport processes (Parker, 2004; Wu, 2007). While powerful, these formulations typically require sediment concentration as an input rather than predicting it explicitly.

Despite the breadth of existing work, a fundamental gap remains: most available approaches are not predictive by construction with respect to inflow suspended sediment concentration. Classical log-scale models are transformations rather than governing equations; rating curves lack probabilistic meaning and dimensional consistency; and many machine-learning models focus on fitting rather than explicitly addressing the dual nature of sediment occurrence and magnitude.

Measured sediment time series exhibit two qualitatively distinct regimes: (1) periods of zero or negligible transport, corresponding to low hydrological forcing, and (2) periods of active transport characterized by strictly positive concentrations driven by rainfall-runoff-erosion processes. Treating these regimes within a single continuous model inevitably leads to conceptual ambiguity and numerical artifacts.

In response to these limitations, the present study introduces a new probabilistic two-regime governing model for predicting inflow suspended sediment concentration to dam reservoirs. The proposed framework explicitly distinguishes between sediment occurrence and sediment magnitude. The zero-transport regime is modeled probabilistically using a Bernoulli formulation, while the positive-transport regime is described through a physically interpretable transformation that remains well-defined at zero without introducing any arbitrary numerical correction.

The model is derived systematically from first principles, leading to a closed-form predictive expression that combines the probability of occurrence and the conditional sediment magnitude into a single forecast at a specified lead time. Unlike classical approaches, the proposed model directly targets inflow suspended sediment concentration, preserves dimensional consistency, avoids ad hoc parameters, and provides a clear probabilistic interpretation.

The objective of this work is therefore to develop, derive, and demonstrate a robust and physically consistent predictive framework that overcomes the conceptual and numerical limitations of existing methods and offers a sound basis for reservoir sediment management and operational decision-making.

Classical Log-scale model

The present work develops an advanced predictive model that overcomes the limitations of classical approaches traditionally employed to represent inflow suspended sediment concentrations to dam reservoirs. The "ancient" (classical) model in the literature expressed as follows:

$$y(t) = \ln\left[C_{\text{in}}(t) + \varepsilon\right] \tag{1}$$

is not a predictive model by itself. It is a variable transformation that has been extensively used in the literature on: (1) sediment transport modelling, (2) hydrological time-series analysis, (3) environmental regression and forecasting, (4) geophysical and geochemical data analysis. This transformation converts the measured inflow suspended sediment concentration $C\text{in}(t)$ into a new variable $y(t)$ defined on the logarithmic scale.

This formulation has been used for more than four decades because: (1) sediment concentrations are strictly nonnegative, while many statistical models assume real-valued variables, (2) sediment data often exhibit: strong right skewness, high variance, and multiplicative variability, (3) the logarithmic transformation stabilizes variance, reduces skewness, and makes linear regression and Gaussian assumptions more plausible.

It is worth noting that the constant $\varepsilon > 0$ is introduced in Eq. (1) solely to avoid the following mathematical indetermination that occurs in the case where $C\text{in}(t) = 0$:

$\ln(0)$

Is undefined.

This approach appears in classical sediment rating curves, log-linear regression models, and early AI/ML sediment predictors.

In addition, importantly, ε has no physical meaning in sediment transport theory: It is not a threshold concentration, it is not related to erosion mechanics, grain size, or hydraulics, and it does not correspond to measurement resolution in a rigorous way. Its value is usually chosen arbitrarily, e.g., $10^{-6} \leq \varepsilon \leq 10^{-2}$, or a fraction of the minimum observed concentration, and different choices of ε lead to different predictions, especially at low concentrations.

Eq. (1) should be read as follows: "The sediment concentration is modeled indirectly by first transforming it into a logarithmic variable $y(t)$, which is then assumed to behave more regularly and be easier to predict."

In addition, in this classical framework, the actual modeling is done on $y(t)$, and the predicted concentration is recovered by exponentiation as follows:

$$\hat{C}\text{in}\,(t) = \exp[\hat{y}(t)] - \varepsilon \tag{2}$$

The symbol "^" (called a *hat*) has a precise and standard meaning in statistics. A quantity written with a hat denotes an estimate or prediction produced by a model, as opposed to a directly measured physical quantity. Thus, the following can be written:

$\hat{y}(t)$

is the model-predicted value of the log-transformed sediment variable $y(t)$, and

$\hat{C}\text{in}\,(t)$

is the estimated (or forecasted) inflow suspended sediment concentration, obtained after inverse transformation.

The hat emphasizes that these quantities are model outputs, subject to uncertainty, and not direct observations.

Accordingly,

$$\hat{y}(t)$$

represents the predicted value of a transformed variable, not a physical sediment concentration. It has: (1) no physical units, (2) no direct physical interpretation in sediment mechanics, (3) meaning only through its inverse transformation back to concentration. In practice, it is obtained from a calibrated regression, time-series model, or machine-learning predictor operating on the log scale.

The variable $y(t)$, expressed by Eq. (1), has several standard names, depending on discipline and context. In high-level scientific writing, $y(t)$ is called Log-transformed suspended sediment concentration, Logarithmic sediment concentration, Log-transformed response variable, or Log-scale sediment concentration. These are the most accurate and widely accepted terms. It is crucial to emphasize that $y(t)$ is not a physical quantity measured in nature; it is a mathematical construct introduced to facilitate modeling; Its meaning is "A transformed representation of sediment concentration intended to linearize relationships, stabilize variance, and enable classical statistical modeling."

In the literature, $y(t)$ is assumed to satisfy a model of the following general form:

$$y(t) = f\,[\mathbf{x}(t);\, \theta] + \varepsilon_y(t) \tag{3}$$

where:

$y(t)$ is the Log-transformed suspended sediment concentration expressed by Eq. (1).

$\mathbf{x}(t)$ is the vector of known predictors at time t, e.g. precipitation, discharge, antecedent rainfall.

$f(\cdot)$ is a chosen model structure.

θ is a vector of unknown model parameters.

$\varepsilon_y(t)$ is a random error term.

The parameters θ are identified by calibrating the predictive model against historical sediment observations, using an optimization criterion that adjusts the parameters so as to achieve the best agreement between predicted and observed values. In the literature, the formula used to compute θ is called: an estimator, more specifically a parameter estimator, an estimation rule, or an estimation rule. When the classical approach is used, it is most often referred to as a least-squares estimator, or a maximum likelihood estimator, under Gaussian error assumptions.

Although well-established, the model expressed by Eq. (1) has known limitations: (1) Arbitrary choice of ε, so different values of ε can lead to different predictions, (2) Implicit handling of zero sediment events, so that zeros are not modeled explicitly; they are "patched" numerically, (3) Bias at low concentrations, so that the back-transformation [exp (·) – ε] introduces bias, (4) Single-regime assumption as the model assumes one continuous process, while sediment dynamics are often multi-regime.

Thus, one may write that, in the classical framework, *C*in(*t*) is never predicted directly. Instead, the procedure is as follows: (1) Transform the observed concentration using Eq. (1), (2) Build and calibrate a predictive model for *y*(*t*), not for *C*in(*t*), (3) Compute a predicted value y^(t), (4) Recover the concentration by inverse transformation, using Eq. (2).

Thus, determining *y*(*t*) means estimating it through a statistical or data-driven model, typically as a function of hydrological predictors such as precipitation, discharge, or antecedent conditions. In addition, in the literature, *y*(*t*) is usually determined by assuming a functional relationship of the following form applied on the log-transformed scale: linear regression, log-linear rating curve, log-linear rating curve, autoregressive model, or, more recently, a machine-learning regressor. In other words, *y*(*t*) is treated as a response variable whose value is inferred from explanatory variables through a calibrated model. The calibration is performed using historical data by minimizing a loss function, most often least squares, implicitly assuming that errors on the *y*-scale are approximately Gaussian.

In Eq. (2), ε reappears explicitly in the subtraction which has several consequences:

(1) Bias at low concentrations
The back-transformation is nonlinear, and subtracting ε induces systematic bias, particularly when y^(*t*) is small.

(2) Artificial negative values
For small predicted

$\hat{y}(t)$

the term $\exp[\hat{y}(t)] - \varepsilon$

may become negative, requiring further ad hoc truncation.

(3) Implicit treatment of zero sediment
Zero sediment concentrations are not modeled explicitly; they are reconstructed indirectly through numerical manipulation.

Thus, Eq. (2) should therefore be interpreted as follows:

The predicted inflow suspended sediment concentration is obtained indirectly by exponentiating a predicted log-transformed variable and subtracting an arbitrary numerical constant introduced to bypass the logarithmic indetermination at zero.

This confirms that Eq. (2) is not a governing physical equation, but a numerical reconstruction formula.

Authors' new two-regime "zero-or-positive" governing model

This is the stronger model because it handles a key reality: in measured sediment series, zeros, or effectively-zero, often occur and are qualitatively different from positive sediment events.

Suspended sediment concentration time series typically have two fundamentally different behaviors:

(a) No transport / negligible transport; concentration is exactly zero (or reported as zero), and physically corresponds to low rainfall, low runoff, no erosion.
(b) Active transport; concentration is strictly positive, and physically corresponds to rainfall-runoff-erosion processes.

These two behaviors are not just different magnitudes of the same process; they are different physical regimes. If one tries to describe both with one single regression equation, we force the model to: average "no transport" and "active transport", distort low values, and rely on artificial fixes such as adding ε.

Target and time indexing

Definition 1 (measured target)

$$C_{\text{in}}(t)$$

Definition: Inflow suspended sediment concentration.
Units: [kg/m^3].
Meaning: Mass of suspended sediment per unit water volume entering the reservoir. The inflow suspended sediment concentration $C\text{in}(t)$ governs the magnitude and timing of sediment inflow; its strong temporal variability and event-driven behavior make it the primary focus in sediment management. The effectiveness of sediment routing operations depends primarily on anticipating peaks in $C\text{in}(t)$ rather than peaks in discharge alone.
Practical determination: Turbidity sensors calibrated with bottle samples; automatic samplers during floods; sometimes satellite turbidity proxy with calibration.

Definition 2 (forecast target)

$$\hat{C}\text{in}\left(t+\tau\right)$$

Definition: it is the predicted inflow suspended sediment concentration at time $(t+\tau)$, using information available up to time t.
Units: [kg/m^3].
Meaning: Forecast used to trigger and schedule sediment-routing operations.

Definition 3 (lead time)

$$\tau = 1\ \text{day}$$

Meaning: information available at day t is used to predict the value at day $(t+1)$; so "next-day concentration" means

$$C_{\text{in}}(t+1)$$

is the concentration one day after time *t*. Example: if *t* = 3, then the "next day" is (*t*+1)=4.

Predictors and feature vector

We choose predictors available at time t.

Predictor variables

Definition 4 (precipitation)

$$P(t)$$

Definition: Precipitation intensity (or flow depth per time step) over the catchment at time *t*.

Units: commonly [mm] per Δt, [mm/h], or [mm/day].

Meaning: Primary storm forcing driving runoff and erosion.

Practical determination: Rain gauges, weather radar, satellite precipitation (e.g., Global Precipitation Measurement, GPM), or meteorological forecasts.

We consider the dimensionless index ***k*** as an integer index used for lagged/antecedent variables. It counts discrete time steps backward in time, e.g., "*k* = 1" means one time step ago. It is used to build antecedent rainfall or discharge indices such as what follows:

$$\sum_{k}^{K} P(t - k\Delta t) \tag{4}$$

K is the dimensionless maximum number of lags included in an antecedent index. It defines the "memory length" of catchment wetness used as a predictor for sediment. It is selected by hydrological reasoning (typical soil moisture memory) and validated by model performance (cross-validation).

Definition 5 (antecedent precipitation index)

To represent "antecedent wetness" in a single scalar, we define an antecedent precipitation index *P*ant(*t*) as a weighted sum of recent rainfall. Because *k* is the "lag index" and *K* is the "maximum number of lags", we must define how past rain enters. A standard, physically motivated choice is a weighted sum of past rainfall, where recent rainfall counts more:

$$P_{\text{ant}}(t) = P(t-1) + \frac{1}{2}P(t-2) \tag{5}$$

This is a specific "two-lag" construction, so *K* = 2; the weight 1/2 means "rain two days ago contributes, but less than yesterday's rain"; this is a *definition*, not a theorem: we define it to operationalize the “antecedent” concept implied by lag notation. it specifies exactly how past rainfall is summarized.

Feature vector

Definition 6 (feature vector)

$$\mathbf{x}(t) = \begin{bmatrix} 1 \\ P(t) \\ P_{\text{ant}}(t) \end{bmatrix} \tag{6}$$

The first entry is the constant **1** so the model can learn a baseline (intercept). The remaining entries are the predictors we selected.

In the numerical example reported later, the authors will use the simpler feature vector to keep computations completely transparent; it reads as follows:

$$\mathbf{x}(t) = \left[1,\ P(t)\right]^T \tag{7}$$

The full three-feature case, expressed by Eq. (6), is identical in method, only longer in arithmetic.

The earlier Eq. (1) form is widely used, but it introduces a tuning constant ε whose value can bias small concentrations. The authors’ “strong” alternative avoids any ε by explicitly modeling two physical regimes: (1) zero sediment, or zero reported by measurement, (2) positive sediment, or real transport events.

The strong two-regime governing model

Zero indicator variable

Definition 7 (zero indicator)

$$Z(t) = \begin{cases} 1, & C_{\text{in}}(t) = 0 \\ 0, & C_{\text{in}}(t) > 0 \end{cases} \tag{8}$$

This definition introduces a binary indicator variable that classifies each observed sediment concentration at time *t* into one of two physically distinct regimes. *Z*(*t*) is, by construction, a Bernoulli random variable. More precisely, *Z*(*t*) is as follows:

$$Z(t) \sim \text{Bernoulli}\left[\pi(t)\right] \tag{8a}$$

where

$$\pi(t) = \mathbb{P}\left[C_{\text{in}}(t) = 0 \mid \mathbf{x}(t)\right] \tag{8b}$$

is the probability that the inflow suspended sediment concentration is zero at time t, conditional on the predictors available at the same time. Accordingly, the following can be written:

$$\mathbb{P}\left[Z(t)=1\right]=\pi(t),\text{ and }\mathbb{P}\left[Z(t)=0\right]=1-\pi(t) \tag{8c}$$

Thus, $Z(t)$ is a binary stochastic variable taking values in $\{0, 1\}$, fully characterized by the single probability parameter $\pi(t)$. The zero indicator $Z(t)$ encodes a physical regime classification such as writing the following: (1) $Z(t) = 1$: zero-transport (inactive) regime, corresponding to hydrological conditions under which sediment is not mobilized and no suspended sediment enters the reservoir, (2) $Z(t) = 0$: positive-transport (active) regime, corresponding to rainfall-runoff-erosion conditions that generate sediment transport.

Importantly, these two cases are not simply low and high values of the same process; they represent qualitatively different sediment dynamics.

In short, the binary variable $Z(t)$, defined in Eq. (8), is a Bernoulli indicator that identifies whether inflow suspended sediment concentration is zero or positive at time t, thereby enabling an explicit separation between sediment occurrence and sediment magnitude.

Thus, from Eq. (8c), the following can be written:

$Z = 1$ means the concentration is exactly zero, and $Z = 0$ means it is strictly positive.

Probability of zero sediment

Definition 8 (probability of zero at forecast time)

$$\pi(t+\tau)=\mathrm{P_r}[Z(t+\tau)=1\,|\,\mathbf{x}(t)] \tag{9}$$

This is the probability that the future concentration equals zero, conditioned on predictors at time t. It must lie in the interval [0, 1], and it represents the likelihood that sediment concentration is zero at time $(t + \tau)$. At the same time, the predictors

$$\mathbf{x}(t)$$

(rainfall, discharge, antecedent conditions, etc.) combine additively and continuously in physical processes. Definition 9 (probability of positive sediment)

$$1-\pi(t+\tau)=\mathrm{P_r}[Z(t+\tau)=0\,|\,\mathbf{x}(t)] \tag{10}$$

This is the probability that the future concentration is positive.

Logistic model for π (t + τ)

The authors introduce a real-valued score as follows:

$$\eta(t+\tau)$$

and map it to a probability with the logistic function. It is a latent activation variable that aggregates the effect of predictors on the probability of zero sediment.

Definition 10 (linear score for zero probability)

$$\eta\left(t+\tau\right)=\mathrm{a}^{T}\mathbf{x}(t) \tag{11}$$

where a^{T} is the transpose of vector **a** which can be written as follows:

$$\mathrm{a}=\begin{bmatrix} a_0 \\ a_1 \\ . \\ . \\ . \end{bmatrix} \tag{12}$$

is a parameter vector. Thus, one may write the following:

$$\mathrm{a}^{T}=\left[a_1\, a_2 \ldots a_n\right] \tag{12a}$$

Definition 11 (logistic function)

The logistic function is defined as follows:

$$\sigma\left(u\right)=\frac{1}{1+e^{-u}} \tag{13}$$

u is a latent activation variable (or linear predictor) that aggregates the influence of the physical predictors on the likelihood of an event, in our case, the occurrence of sediment transport. In your framework, the following can be written:

$$u=\mathrm{a}^{T}\mathbf{x}(t) \tag{13a}$$

It is important to be explicit: u is not a measurable physical variable, u is not a probability, u is not a sediment concentration. Instead, it is an intermediate latent quantity introduced to connect physical forcing to probabilistic behavior. Although u is not directly observable, it has a clear physical interpretation: it represents the net “propensity” or “driving potential” of the system to activate sediment transport under the prevailing conditions. More concretely: (1) Large negative u means physical conditions are insufficient to mobilize sediment, i.e., sediment transport is very unlikely, (2) $u \approx 0$ means that conditions are close to a threshold, that is to say sediment transport is possible but uncertain, and (3) Large positive u indicates conditions strongly favor erosion and transport, i.e., sediment transport is very likely.

Governing equation 1 for zero probability model

$$\pi\left(t+\tau\right)=\sigma\left[\eta\left(t+\tau\right)\right]=\frac{1}{1+\exp\left[-\mathrm{a}^{T}\mathbf{x}(t)\right]} \tag{14}$$

The above equality should be read as follows: The probability of zero sediment at time ($t + \tau$) is a monotonic function of a latent activation variable that aggregates the effects of the physical predictors. It does not mean that sediment transport is deterministic, or that η is physically measurable. It means: It means: η represents how favorable conditions are, $\sigma(\cdot)$ converts favorability into probability.

Positive regime model without ε

A transform that is defined at zero

We need a transform that behaves like a log for large values but is defined at 0 without adding ε. We use asinh, defined exactly below.

Definition 12 (scale parameter)

$$s>0 \tag{15}$$

is a concentration scale (units kg/m^3). It controls how strongly the transform compresses large values. In the example I reported below, we choose s = 1 kg§m^3 for simplest arithmetic. The parameter s is a positive scaling constant that defines the reference magnitude used to normalize inflow suspended sediment concentration before applying the transformation in the positive-regime model. More explicitly, s has the same physical units as sediment concentration; it rescales Cin into a dimensionless quantity prior to transformation; it controls how rapidly the transformation transitions between linear and logarithmic behavior.

The scale parameter s therefore determines the concentration level at which the model transitions from linear sensitivity to logarithmic compression. It represents a characteristic sediment concentration separating low-intensity transport from high-intensity transport regimes. In practical terms, values of Cin below s are treated as low to moderate transport, while values above s are treated as high transport, where compression is desirable to avoid domination by extremes. Thus, s plays a role analogous to a reference concentration scale.

Moreover, the scale parameter s plays three essential roles in the model: (1) Numerical stability as it prevents excessive compression or amplification of values in the transformed space, (2) Interpretability of coefficients by setting a meaningful scale, regression coefficients in the positive-regime mean model correspond to changes around a physically interpretable concentration level, (3) Replacement of ε; unlike the classical log-plus-epsilon model, s is not an arbitrary numerical fix but a structural scale parameter with physical meaning.

There are three legitimate strategies to determine s, depending on modeling philosophy: (1) Data-driven choice (most common), where s may be set equal to a characteristic statistic of the observed positive sediment concentrations, such as the median, geometric mean, or a lower quantile. This choice ensures a good balance between linear and logarithmic regimes, and a minimal distortion of both low and high concentrations; (2) Physically informed choice where s may be chosen to represent a physically meaningful threshold concentration, for example a typical background transport level or an erosion activation scale. This approach is suitable when domain knowledge is strong; (3) Fixed reference scale (simplification) where s may be fixed a priori to a convenient reference value to simplify interpretation and computation, provided that the chosen value lies within the range of observed concentrations. This is the approach adopted in the illustrative example.

In the numerical example, we fixed $s = 1$ kg§m^3. This choice was made deliberately and transparently, for the following reasons: (1) Pedagogical clarity: setting $s = 1$ removes unnecessary scaling factors and allows the reader to follow each numerical step without additional arithmetic complexity, (2) Unit consistency: with $s = 1$, the argument of the transformation remains numerically equal to the concentration value expressed in kg/m^3, simplifying interpretation, (3) Representative magnitude: the example concentrations, ranging from approximately 1 to 4 kg/m^3, are of the same order of magnitude as s, ensuring that both linear and logarithmic behaviors are demonstrated.

Importantly, fixing $s = 1$ kg/m^3 in the example does not imply that this value is optimal or universal; it is chosen solely for demonstration purposes.

For readers, here are publication-ready formulations if needed:

(1) The scale parameter s defines a reference concentration that governs the transition between linear and logarithmic behavior in the positive-regime transformation, thereby controlling the sensitivity of the model across different sediment transport intensities.

(2) Unlike the arbitrary offset parameter commonly introduced in logarithmic transformations, the scale s has a clear physical interpretation as a characteristic sediment concentration and can be selected based on data-driven or physically informed criteria.

(3) In the illustrative numerical example, s was fixed to 1 kg/m^3 to simplify calculations and improve interpretability, without loss of generality.

As a one-sentence synthesis, one may write the following: The scale parameter s serves as a physically interpretable normalization constant that controls the behavior of the positive-regime transformation, replacing ad hoc numerical corrections with a meaningful concentration reference.

Definition 13 (asinh function, exact definition)

For any real u, the following can be written:

$$\operatorname{asinh}(u) = \ln\left(u + \sqrt{u^2 + 1}\right) \tag{16}$$

This is defined for $u = 0$ because:

$$\operatorname{asinh}(0) = \ln\left(0 + \sqrt{0+1}\right) = \ln(1) = 0 \tag{17}$$

Definition 14 (positive-regime transformed variable)

For positive concentrations, we define the following:

$$y^{+}(t) = \operatorname{asinh}\left[\frac{C_{\text{in}}(t)}{s}\right] \tag{18}$$

The variable $y^{+}(t)$ is defined only for the positive sediment transport regime, i.e. for observations such that $C\text{in}(t) > 0$, as indicated in Eq. (8). The superscript "+" explicitly indicates that this variable is defined exclusively on the positive-sediment regime, and is not used when sediment concentration is zero. From the physical meaning point view, $y^{+}(t)$ represents a dimensionless, transformed measure of sediment transport intensity during periods of active sediment mobilization. It is not a direct physical quantity, but rather a mathematical representation of suspended sediment concentration designed to preserve physical interpretability while enabling stable and predictive modeling.

More specifically, it is worth noting the following:

$y^{+}(t)$

(1) quantifies the relative magnitude of suspended sediment concentration with respect to a physically meaningful reference scale s.

(2) It encodes sediment intensity in a form that remains well-defined, smooth, and continuous over the entire positive range of concentrations.

(3) It enables additive modeling of sediment dynamics without violating the underlying physics of sediment transport.

In one-sentence synthesis, the variable $y^{+}(t)$ is a dimensionless, transformed representation of inflow suspended sediment concentration during active transport conditions, designed to preserve physical interpretability, eliminate numerical indeterminacy at zero, and enable stable predictive modeling within the positive sediment regime.

Unlike the classical model expressed by Eq. (1), no ε appears anywhere in Eq. (18).

The positive-regime model uses the previous transformation which has the following two asymptotic regimes:

(1) For small concentrations, i.e.,

$C_{\text{in}} \ll s$

Thus, Eq. (18) allows writing the following:

$$\operatorname{asinh}\left[\frac{C_{\text{in}}(t)}{s}\right] \approx \frac{C_{\text{in}}(t)}{s} \tag{18a}$$

Then:

$$y^{+}(t) \approx \frac{C_{\text{in}}(t)}{s} \tag{18b}$$

This is nearly linear behavior.

(2) For large concentrations, i.e.,

$C_{\text{in}} \gg s$

Thus, from Eq. (18), the following can be written:

$$\operatorname{asinh}\left[\frac{C_{\text{in}}(t)}{s}\right] \approx \ln\left(\frac{2C_{\text{in}}(t)}{s}\right) \tag{18c}$$

Then:

$$y^{+}(t) \approx \ln\left(\frac{2C_{\text{in}}(t)}{s}\right) \tag{18d}$$

This is a logarithmic-like behavior.

Positive-regime mean model

Definition 15 (positive-regime mean model)

For this, the following is defined:

$$\mu(t+\tau) = \mathrm{b}^{T}\mathbf{x}(t) \tag{19}$$

where **b** is a parameter vector.

In essence, μ $(t + \tau)$ characterizes the average magnitude of inflow suspended sediment concentration expected at the forecast horizon, once sediment transport is present, under the prevailing conditions at time *t*. It represents also the expected value of the transformed sediment concentration at the forecast horizon $(t + \tau)$, conditional on the positive-transport regime and on the hydrometeorological conditions observed at time *t*.

Formally, it can be written as follows:

$$\mu\left(t+\tau\right)=\mathrm{E}\left[\,y^{+}\left(t+\tau\right)\middle|\,Z\left(t+\tau\right)=0;\,\mathbf{x}(t)\,\right] \tag{20}$$

Positive-regime conditional model

Definition 16 (positive-regime conditional model)

We define what follows:

$$y^{+}\left(t+\tau\right)\Big|\left[\,Z\left(t+\tau\right)=0,\,\mathbf{x}(t)\right]\approx N\left[\,\mu\left(t+\tau\right);\,\sigma^{2}\right] \tag{21}$$

Meaning: conditioned on being in the positive regime ($Z = 0$), and the transformed variable y^{+} is modeled as Normal with mean μ and variance σ^{2}. The vertical bar "|" means "given that". Thus, this above expression reads: "The distribution of $y^{+}(t + \tau)$, given that sediment is present and given the predictors $\mathbf{x}(t)$, is normal." This is crucial as: the distribution applies only in the positive regime, and zero events are excluded.

The symbol:

$$N\left[\,\mu\left(t+\tau\right);\,\sigma^{2}\right]$$

means normal distribution, also called Gaussian distribution. It is not a number and not a parameter. It is a probability distribution, i.e. a mathematical object that describes how a random variable can vary.

Definition 17 (variance parameter)

$$\sigma^{2}>0 \tag{22}$$

is the conditional variance of y^{+} in the positive regime.

Inverse transform back to concentration

From Definition 13, the inverse of asinh is sinh. Specifically, if:

$$y^{+}\left(t+\tau\right)=\operatorname{asinh}\left[\frac{C_{\mathrm{in}}\left(t+\tau\right)}{s}\right] \tag{18}$$

This variable y^{+} is a mathematical representation of concentration used to stabilize variability and make modeling tractable.

Then, the following can be written:

$$\frac{C_{\mathrm{in}}}{s}=\sinh\left(y^{+}\right) \tag{23}$$

Thus, the following can be derived:

$$C_{\text{in}} = s\sinh\left(y^{+}\right) \tag{24}$$

Governing Eq. 2 for zero probability model

$$\hat{C}_{\text{in}}^{+}\left(t+\tau\right) = s\sinh[\mu\left(t+\tau\right)] \tag{25}$$

is the predicted inflow suspended sediment concentration at lead time τ.
We will set τ =1 day (one-day-ahead forecast), because τ is precisely the "forecast lead time".

Final predicted concentration (mixture expectation)

A single point prediction that includes both "zero" and "positive" possibilities is the following:

Governing Equation 3 (final forecast)

$$\hat{C}\text{in}\left(t+\tau\right) = \mathrm{E}\left[\, C_{\text{in}}\left(t+\tau\right)\middle|\,\mathbf{x}(t)\,\right] = \left[1-\pi\left(t+\tau\right)\right]\hat{C}_{\text{in}}^{+}\left(t+\tau\right) \tag{26}$$

That is to say:

$$\hat{C}\text{in}\left(t+\tau\right) = \left[1-\pi\left(t+\tau\right)\right]\hat{C}_{\text{in}}^{+}\left(t+\tau\right) \tag{26a}$$

Eq. (26a) matches the final predictive form of the two-regime ("zero-or-positive") model that the authors derived. This equation is the mixture (probability-weighted) prediction: it combines: a model for occurrence (zero vs. positive), and a model for magnitude given occurrence.

One may provide the few following observations:

$$\hat{C}\text{in}\left(t+\tau\right)$$

is the overall (unconditional) predicted inflow suspended sediment concentration at lead time τ.

$$\pi\left(t+\tau\right) = \mathrm{P_r}[Z\left(t+\tau\right)=1\,\middle|\,\mathbf{x}\left(t\right)]\text{, [Eq. (9)]}$$

is the predicted probability that the concentration is zero at (t + τ), i.e., zero-transport regime.

$$1-\pi\left(t+\tau\right) = \mathrm{P_r}[Z\left(t+\tau\right)=0\,\middle|\,\mathbf{x}\left(t\right)]\text{, [Eq. (10)]}$$

is the predicted probability that sediment transport occurs, i.e., positive regime.

$$\hat{C}^{+}_{\mathrm{in}}\left(t+\tau\right)$$

is the predicted concentration conditional on being in the positive regime, i.e., given the following:

$$C_{\mathrm{in}}\left(t+\tau\right)>0$$

The zero regime contributes $0\times\pi$, and the positive regime contributes

$$\hat{C}^{+}_{\mathrm{in}}\times\pi$$

Note that,

$$\hat{C}^{+}_{\mathrm{in}}$$

describes sediment intensity when transport is present, whereas

$$C_{\mathrm{in}}$$

represents the expected inflow sediment concentration across all hydrological conditions. Distinguishing between the two parameters allows the modeling framework to decouple the processes governing sediment occurrence from those governing sediment magnitude, thereby avoiding the conceptual and numerical ambiguities inherent in single-regime formulations.

Remember that:

$$\hat{C}\mathrm{in}\left(t+\tau\right)$$

represents the model-based estimate of the suspended sediment concentration entering the reservoir at the future time $(t+\tau)$, conditional on the predictor variables observed or inferred at time t. In addition, the previous quantity, e.g.:

$$\hat{C}\mathrm{in}\left(t+\tau\right)$$

which is the model-derived estimate of future inflow suspended sediment concentration, must be distinguished from the measured concentration

$$C_{\mathrm{in}}\left(t+\tau\right)$$

as it constitutes a predictive estimate rather than an observed physical measurement.

Positive-regime formulation and choice of scaling parameters

One can estimate the scale s jointly with the other positive-regime parameters by treating s as an unknown model parameter and choosing the value of s that makes the observed data most probable under the positive-regime probabilistic model. Below is a fully explicit numerical example, no shortcuts in the logic, using the positive-regime model and estimating s, b_0, b_1, and σ^2 together.

We consider only positive transport events, i.e. observations where:

$$C_{\text{in}}\left(t+\tau\right) > 0$$

For those events, we assume: (1) Ci > 0 is the observed inflow sediment concentration (positive regime), (2) Pi is the predictor, herein precipitation at time t, (3) τ = 1 day, so Ci corresponds to "next-day" concentration.

Transform (depends on s)

From Eq. (18), define the following transformed variable (positive regime):

$$y_i\left(s\right) = \operatorname{asinh}\left(\frac{C_i}{s}\right) \tag{18e}$$

with the following exact definition [Es. (16)]:

$$\operatorname{asinh}\left(u\right) = \ln\left(u + \sqrt{u^2 + 1}\right) \tag{16}$$

Herein, $s > 0$ (units: kg/m^3) is unknown.

Positive-regime mean model

Define the following multi-predictor feature vector:

$$\mathbf{x}\left(t\right) = \begin{bmatrix} 1 \\ P\left(t\right) \\ P_{\text{ant}}\left(t\right) \\ Q_{\text{in}}\left(t\right) \end{bmatrix} \tag{27}$$

where Qin(t) is inflow discharge at time t [m^3/s]. The leading "1" is included so the model can represent a baseline level.

Before choosing any "linear" form, we acknowledge what is true, allowing writing the following:

There exists some but generally unknown function *G* such that:

$$\mu(t+\tau) = G\left[\mathbf{x}(t)\right] \tag{28}$$

This is not a choice; it is simply saying: "the conditional mean depends on predictors through some relationship." But *G* is unknown, so we need an approximating form that: (1) is estimable from data, (2) is interpretable, and (3) is stable with limited data.

There are two standard, fully rational derivation routes. We will give both; they reach the same final form.

Route A (Taylor expansion / local linearization)

Step A1

Assume *G* is smooth (differentiable) in a region of interest. This is a modeling assumption; in practice, hydrological response functions are often treated as smooth within the range of observed conditions.

Step A2

Choose a reference operating point $\mathbf{x}_0$ as a typical state, e.g., mean predictors over the training set. We do not need its explicit value to derive the form.

Step A3

Apply first-order Taylor expansion of *G* around $\mathbf{x}_0$. A first-order Taylor expansion gives the following:

$$G(\mathbf{x}) \approx G(\mathbf{x}_0) + \nabla G(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0) \tag{29}$$

where

$$\nabla G(\mathbf{x}_0)$$

is the gradient vector of partial derivatives evaluated at $\mathbf{x}_0$,
The term

$$\nabla G(\mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0)$$

is the first-order change from $\mathbf{x}_0$ to $\mathbf{x}$.

Step A4

Rearrange into an intercept plus a weighted sum. So, expand as follows:

$$G(\mathbf{x}) \approx G(\mathbf{x}_0) - \nabla G(\mathbf{x}_0)^T \mathbf{x}_0 + \nabla G(\mathbf{x}_0)^T \mathbf{x} \tag{30}$$

Now define the following parameters:

$$\mathrm{b}_0 := G(\mathbf{x}_0) - \nabla G(\mathbf{x}_0)^T \mathbf{x}_0 \tag{31}$$

$$\mathbf{b} := \nabla G(\mathbf{x}_0)\mathbf{x} \tag{32}$$

Then, Eq. (30) reduces to the following:

$$G(\mathbf{x}) \approx \mathrm{b}_0 + \mathbf{b}^T \mathbf{x} \tag{33}$$

Step A5

Write it explicitly for multiple predictors; so, if:

$$\mathbf{x} = \left[\ 1,\ P,\ P_{\mathrm{ant}},\ Q_{\mathrm{in}}\ \right]^T \tag{34}$$

then

$$\mathbf{b} = \left[\ \mathrm{b}_0, \mathrm{b}_1, \mathrm{b}_2, \mathrm{b}_3\ \right]^T \tag{35}$$

and Eq. (28) reduces to the following:

$$\mu(t + \tau) \approx b_0 + b_1 P(t) + b_2 P_{\mathrm{ant}}(t) + b_3 Q_{\mathrm{in}}(t) \tag{36}$$

This is the multi-predictor extension derived as a first-order approximation of an unknown smooth mean function.

Route B (additive effects + parsimony)

This route is more conceptual but equally rigorous.

Step B1

Separate “baseline” and “incremental effects”, so we assert the conditional mean can be decomposed as follows: (1) a baseline level when drivers are minimal, and (2) plus increments due to each driver.

Step B2

Assume each driver contributes approximately linearly on the transformed scale. Because $y+$ is designed to tame nonlinearity, it is reasonable to start with first-order (linear-in-parameters) contributions.

Step B3

The sum contributions is as follows:

$$\mu(t + \tau) = \text{baseline} + \text{effect of}\, P + \text{effect of}\, P_{\mathrm{ant}} + \text{effect of}\, Q_{\mathrm{in}} \tag{37}$$

and "effect" is represented by a coefficient times the predictor value, yielding the same form as Eq. (36), so writing the following:

$$\mu(t+\tau) = b_0 + b_1 P(t) + b_2 P_{\text{ant}}(t) + b_3 Q_{\text{in}}(t) \tag{36}$$

Numerical example

The following numerical example is designed to provide a complete and transparent illustration of the proposed two-regime predictive model for inflow suspended sediment concentration. The example demonstrates, step by step, how the final forecast

$$\hat{C}\text{in}(t+\tau)$$

is obtained by combining (1) the predicted probability of sediment occurrence and (2) the predicted sediment concentration conditional on active transport, in accordance with the governing predictive Eq. (26a), recalled as follows:

$$\hat{C}\text{in}(t+\tau) = \left[1 - \pi(t+\tau)\right]\hat{C}^{+}_{\text{in}}(t+\tau) \tag{26a}$$

Specifically, the example aims to illustrate how historical observations are used to estimate the probability of zero sediment occurrence

$$\pi(t+\tau)$$

how the magnitude of sediment concentration during positive transport events

$$\hat{C}^{+}_{\text{in}}(t+\tau)$$

is modeled and predicted, and how these two components are subsequently combined to produce an unconditional prediction of inflow suspended sediment concentration.

For clarity and reproducibility, the numerical example considers a one-day forecast horizon, i.e., $\tau = 1$, and a simplified set of predictor variables. All model parameters are estimated explicitly from a small synthetic dataset, and each computational step is shown in detail, without approximation or shortcut.

At the end of the numerical example, the reader is expected to clearly understand how the proposed predictive framework translates input data into a final sediment concentration forecast, and how the probability-weighted structure of the model naturally accounts for both zero-transport and active-transport conditions.

Given data

To keep the arithmetic fully checkable, we use the following data:

$\tau = 1$ day,

Feature vector

$$\mathbf{x}(t) = \left[1,\ P(t)\right]^T \text{ [Eq. (7)]}$$

Scale $s = 1$ kg/m^3

We provide precipitation at day t and the inflow observed sediment concentration at day $(t + 1)$. Table 1 reports $n = 6$ training samples.

Table 1: Data used in the numerical example

Training index *i*	Day *ti*	*P*(*ti*) mm/day	Observed *C*in (*ti* + 1) kg/m³
1	1	0	0.0
2	2	5	1.2
3	3	10	2.5
4	4	0	0.0
5	5	20	4.0
6	6	15	3.0

Build the zero indicator *Z*

By definition 7 (zero indicator), [Eq. (8)], the following can be written:

$$Z\left(t_i + 1\right) = \begin{cases} 1, & C_{\text{in}}\left(t_i + 1\right) = 0 \\ 0, & C_{\text{in}}\left(t_i + 1\right) > 0 \end{cases}$$

So, we get the following Table:

Table 2: Building the zero indicator

Training index *i*	Observed *C*i (*ti* + 1) kg/m³	$Z\left(t_i + 1\right)$
1	0.0	1
2	1.2	0
3	2.5	0
4	0.0	1
5	4.0	0
6	3.0	0

Thus, the number of zeros is as follows:

$$m = \sum_{i=1}^{n} Z\left(t_i + 1\right) = 2$$

Estimating the probability π (intercept-only logistic)

To make the derivation fully explicit and solvable by hand, we choose the following simplest logistic model:

$\pi\left(t + \tau\right) = \pi = \text{constant}$, and there is no dependence on *P*.

This statement refers only to the zero-occurrence (logistic) sub-model, i.e. the model governing whether sediment is present or absent. It means that: (1) precipitation *P* is not used to explain the probability of zero sediment, (2) the probability of zero sediment is assumed to be constant across all conditions.

Precipitation is not used in the zero-occurrence model, but it is still required for the positive-regime (magnitude) model. The numerical example involves two distinct sub-models, each with a different role: (1) Zero-occurrence model (logistic part) answering the question "Will sediment transport occur at all? the predictor vector is $\mathbf{x}(t) = [1]$, meaning that precipitation is intentionally excluded herein. This sub-model uses only the presence or absence of sediment, not its magnitude; (2) Positive-regime model (sediment magnitude) answers a different question: "If sediment transport occurs, how large is the sediment concentration"? For this second question, precipitation is physically meaningful and necessary, and the model is used as a predictor of sediment intensity.

Thus, precipitation values are listed in Table 1 because they are required to model sediment magnitude in the positive regime, even though they are deliberately excluded from the zero-occurrence model, where the probability of zero sediment is assumed constant for illustrative purposes.

In the above illustration, the probability of zero sediment occurrence was assumed to be constant, such that $\pi(t + \tau) = \pi$, in order to provide a baseline representation independent of precipitation. This constant-probability formulation can be viewed as a special case of the logistic model in which the linear predictor reduces to an intercept term only.

The general model of

$$\eta\left(t + \tau\right)$$

is given by Eq. (11). However, in the present case, one may write the following:

$$\eta\left(t + \tau\right) = a_0$$

since it follows directly from defining the predictor vector as [Eq. (7)] with no effect of the precipitation *P*:

$$\mathbf{x}(t) = [1,\ 0]^T = [1]^T$$

which implies that no hydrometeorological variables, such as precipitation, are assumed to influence the probability of zero sediment occurrence. Indeed, this modeling choice implies that the occurrence of zero sediment events is characterized by a constant baseline probability, rather than being explicitly driven by rainfall intensity or other hydrological predictors. Thus, the derived following result:

$$\eta\left(t + \tau\right) = a_0$$

arises naturally from the choice of an intercept-only predictor vector, reflecting the assumption that precipitation has no explicit effect on the probability of zero sediment occurrence in the illustrative example.
Thus, according to Eq. (14), the following can be written:

$$\pi = \sigma\left(a_0\right) = \frac{1}{1 + \exp\left(- a_0\right)}$$

Likelihood for Bernoulli observations

Each

$$Z_i := Z\left(t_i + 1\right)$$

is Bernoulli with parameter π. The probability mass function is as follows:

$$\mathrm{P_r}\left(Z_i = z_i\right) = \pi^{z_i}\left(1 - \pi\right)^{1 - z_i}$$

Assuming independence across the index i, the likelihood is as follows:

$$L\left(\pi\right) = \prod_{i=1}^{n} \pi^{z_i}\left(1 - \pi\right)^{1 - z_i}$$

Now, group powers of π and $(1 - \pi)$ such as writing what follows:

$\sum z_i = m$, and

$$\sum\left(1 - z_i\right) = n - m$$

Thus, on may write the following:

$$L\left(\pi\right) = \pi^m\left(1 - \pi\right)^{n - m}$$

Let’s take the natural log. Thus, the following can be written:

$$\ell(\pi) = \ln L(\pi) = m\ln\pi + (n - m)\ln(1 - \pi)$$

Now, differentiate with respect to π. The following is derived:

$$\frac{d\ell}{d\pi} = \frac{m}{\pi} - \frac{n - m}{1 - \pi}$$

Let's set derivative to zero for maximum likelihood; thus, the following is written:

$$\frac{m}{\pi} - \frac{n - m}{1 - \pi} = 0$$

Thus, the following can be written:

$$\frac{m(1 - \pi) - (n - m)\pi}{\pi(1 - \pi)} = 0$$

The denominator is positive for $0 < \pi < 1$, so the numerator must be zero such as writing the following:

$$m(1 - \pi) - (n - m)\pi = 0$$

Expanding the previous equality yields the following:

$$m - m\pi - n\pi + m\pi = 0$$

As the $-\ m\pi$ and $+\ m\pi$ cancel, the following can be written:

$$\hat{\pi} = \frac{m}{n} = \frac{2}{6} = \frac{1}{3} = 0.333333...$$

Therefore, the following can be written:

$$1 - \hat{\pi} = \frac{n - m}{n} = \frac{2}{3} = 0.666666...$$

This means explicitly what follows:

(1) $\hat{\pi}$ is the estimated probability that next-day sediment is zero,

(2) $(1 - \hat{\pi})$ is the estimated probability that next-day sediment is positive.

This result is statistically exact under the Bernoulli modeling assumption adopted for sediment occurrence.

In other words, one may state the following:

$$\hat{\pi} = 1/3$$

is the maximum-likelihood estimate of the probability of zero sediment; it represents the empirical probability of zero sediment occurrence in the sample used for the numerical example. Out of the $n = 6$ observed cases, exactly $m = 2$ correspond to zero suspended sediment concentration, while 4 cases correspond to positive sediment transport.

On the other side, the following can be written: From a physical standpoint, the result below

$$1 - \hat{\pi} = 2/3$$

indicates that, in the considered dataset, sediment transport is active most of the time, but not permanently. This is fully consistent with the intermittent and event-driven nature of sediment dynamics, where transport occurs during sufficiently energetic hydrological conditions and vanishes during quiescent periods. Conversely,

$$\hat{\pi} = 1/3$$

quantifies the non-negligible frequency of zero-transport conditions, which classical log-scale models are unable to represent explicitly without ad hoc numerical corrections.

Moreover, the key point lies in the fact that the previous numerical values have a strong implication for the predicted model. Indeed, they have a direct and transparent role in the final predictive model expressed by Eq. (26a), recalled as follows:

$$\hat{C}\mathrm{in}\left(t+\tau\right) = \left[1 - \hat{\pi}\left(t+\tau\right)\right]\hat{C}_{\mathrm{in}}^{+}\left(t+\tau\right) \tag{26a}$$

This means the following: (1) the model does not always predict sediment, even when the positive-regime model yields a positive value; and (2) the predicted sediment concentration is explicitly down-weighted by the probability of occurrence.

In this considered numerical example, the predicted magnitude from the positive regime is reduced by a factor 2/3, reflecting the fact that sediment transport does not occur systematically.

These probability values illustrate a fundamental advantage of the proposed two-regime framework: (1) The occurrence probability π is estimated from data, not imposed; (2) Zero-transport events are treated as a structural regime, not as numerical anomalies; and (3) The final prediction is a probability-weighted expectation, not a blind extrapolation.

Classical models implicitly assume $1 - \pi = 1$, meaning that sediment is always present, which is contradicted by the previous empirical result, i.e.:

$$\overset{\wedge}{\pi} = 1/3$$

It is important to emphasize that: (1) the dataset is intentionally small and illustrative, (2) the values 1/3 and 2/3 are not meant to be universal, but demonstrative, (3) they show that the model behaves correctly even with limited data.

In larger datasets, π^ would naturally converge toward the long-term frequency of zero sediment events, making the framework even more robust.

Furthermore, the estimated probabilities

$$\overset{\wedge}{\pi} = 1/3 \text{ and } 1 - \overset{\wedge}{\pi} = 2/3$$

demonstrate that zero and active sediment transport regimes coexist in the observed data, thereby justifying the probabilistic two-regime formulation and highlighting the limitations of classical single-regime sediment models.

Positive regime data and transform *y*+

Positive cases are those with the following:

$$C_{\text{in}}\left(t_i + 1\right) > 0$$

where indices are (Table 2): i = 2, 3, 5, 6, corresponding to the following:

$$Z\left(t_i + 1\right) = 0$$

Let's set the scale s = 1 kg/m^3. For each positive case, and according to Eqs. (16) and (18e), compute the following:

$$y_i^+ = \operatorname{asinh}\left(C_i\right) = \ln\left(C_i + \sqrt{C_i^2 + 1}\right)$$

Computing each y+ explicitly in accordance with the previous formula yield the Table 3 below:

Table 3: Values of *y*+

i	*Ci*	*y*+
2	1.2	1.01597313
3	2.5	1.64723115
5	4	2.09471255
6	3	1.81844646

In addition, the positive-regime dataset is as reported in Table 4 below:

Table 4: Positive-regime dataset

Case	$P(t)$ mm/day	Cin $(t + 1)$ kg/m^3	$y+$ = asinh (C)
A	5	1.2	1.01597313
B	10	2.5	1.64723115
C	20	4	2.09471255
D	15	3	1.81844646

Fit the positive-regime mean model by least squares

For a "positive case", let's adopt the index j that means an observation such as follows:

$$C_{\text{in}}\left(t_j + \tau\right) > 0$$

For such a case, according to Eq. (18), we define the transformed response

$$y_j^+ = \text{asinh}\left[\frac{C_{\text{in}}\left(t_j + \tau\right)}{s}\right]$$

According to Eq. (36), we define the conditional mean of the transformed variable under the positive regime by the following:

$$\mu_j = \mu\left(t_j + \tau\right) = b_0 + b_1 P\left(t_j\right)$$

For compactness we adopt the following:

$$P_j := P\left(t_j\right)$$

Hence, the following can be written:

$$\mu_j = b_0 + b_1 P_j$$

In addition, we assume that, given predictors and given positive regime, what follows:

$$y_j^+ \,\Big|\, \left[\, C_{\text{in}}\left(t_j + \tau\right) > 0,\, P_j \right] \approx N\left(\, \mu_j ; \sigma^2 \right)$$

When the example says: For each positive case *j*, prediction is as follows:

$$\hat{y}_j = b_0 + b_1 P_j$$

it is referring to the point prediction of the following transformed variable: y_j^+.

A point prediction must be a single number. In probabilistic modeling, the most standard point prediction is the conditional expected value because it is optimal under squared-error loss. So, we define the point predictor as follows:

$$\hat{y}_j := \mathrm{E}\left[\, y_j^+ \,\middle|\, C_{\text{in}}\left(t_j + \tau\right) > 0,\, P_j \,\right]$$

We start from the previous conditional model which recalled as follows:

$$y_j^+ \,\Big|\, \left[\, C_{\text{in}}\left(t_j + \tau\right) > 0,\, P_j \right] \approx N\left(\, \mu_j\, ;\, \sigma^2 \right)$$

A basic property of a normal distribution is as follows:

If:

$$Y \sim N\left(\, \mu ; \sigma^2 \right)$$, then

$$\mathrm{E}\left[Y\right] = \mu$$

Let's apply this to the conditional distribution of

$$y_j^+$$

Since conditionally it is normal with mean

$$\mu_j$$

we have the following:

$$\mathrm{E}\left[\, y_j^+ \,\middle|\, C_{\text{in}}\left(t_j + \tau\right) > 0,\, P_j \,\right] = \mu_j$$

However, previously, we defined the following:

$$\mu_j = b_0 + b_1 P_j$$

Therefore, one may write what follows:

$$\hat{y}_j = \mathrm{E}\left[\, y_j^+ \,\middle|\, C_{\text{in}}\left(t_j + \tau\right) > 0,\, P_j \,\right] = \mu_j = b_0 + b_1 P_j$$

Thus, for each positive case *j*, prediction is as follows:

$$\hat{y}_j^{+} = b_0 + b_1 P_j$$

Build the design matrix X and target vector y

There are n+ = 4 positive samples.

Design matrix *X* has one row per sample and one column per parameter: first column is 1 (for b_0), and second column is *P* (for b_1). Thus, the following can be written:

$$X = \begin{bmatrix} 1 & 5 \\ 1 & 10 \\ 1 & 20 \\ 1 & 15 \end{bmatrix}$$

Target vector *y* is the list of observed *y*+ values reported in Table 4. Thus, the following can be written:

$$y = \begin{bmatrix} 1.01597313 \\ 1.64723115 \\ 2.09471255 \\ 1.81844646 \end{bmatrix}$$

Least squares objective

The residual vector is as follows:

$$r = y - Xb$$

where

$$b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

Let's *J*(**b**) denotes the objective function, that is, a scalar-valued function that measures how well a given parameter vector **b** explains the observed data. In other words: **b** is the decision variable, i.e., the parameters we want to estimate, and *J*(**b**) quantifies the misfit, or, equivalently, the lack of agreement, between model predictions and observations. The goal of estimation is to find the value of **b** that optimizes (usually minimizes) this function.

Thus, the least-squares objective function is defined as follows:

$$J(\mathbf{b}) = r^T r = (y - Xb)^T (y - Xb)$$

which represents the sum of squared residuals over all positive-regime observations.

Minimizing *J*(**b**) yields the parameter vector linked to **b** that provides the best fit to the data in the sense of minimizing the total quadratic prediction error on the transformed scale. This estimation procedure is employed only at the parameter-calibration stage of the positive-regime model and does not intervene in the derivation of the governing predictive equations. Under the assumption that the conditional errors in the positive regime are independent and normally distributed with constant variance, minimization of the least-squares objective is equivalent to maximum likelihood estimation. Consequently, the least-squares criterion provides both a statistically optimal and computationally efficient means of estimating model coefficients once the probabilistic two-regime structure has been established.

Minimizing *J*(**b**) gives the following normal equation:

$$X^T X b = X^T y$$

Now, let's compute these entry by entry, such as the following can be written:

(1) Compute $X^T X$

$$X^T X = \begin{bmatrix} \sum 1^2 & \sum(1 \times P) \\ \sum(P \times 1) & \sum P^2 \end{bmatrix}$$

Compute sums as follows:

$$\sum 1^2 = 4$$

$$\sum P = 5 + 10 + 20 + 15 = 50$$

$$\sum P^2 = 5^2 + 10^2 + 20^2 + 15^2 = 25 + 100 + 400 + 225 = 750$$

Thus:

$$X^T X = \begin{pmatrix} 4 & 50 \\ 50 & 750 \end{pmatrix}$$

(2) Compute $X^T y$

$$X^T y = \begin{bmatrix} \sum y \\ \sum(Py) \end{bmatrix}$$

with

$$\sum y = 1.01597313 + 1.64723115 + 2.09471255 + 1.81844646 = 6.5763632$$

$\Sigma(Py)$

- $5 \times 1.01597313 = 5.0798655$
- $10 \times 1.64723115 = 16.472311$
- $20 \times 2.09471255 = 41.894250$
- $15 \times 1.81844646 = 27.2766975$

Thus, the sum is as follows:

$$\Sigma(Py) = 5.0798655 + 16.472311 + 41.894250 + 27.2766975 = 90.723124$$

Then, the following can be written:

$$X^T y = \begin{bmatrix} \Sigma y \\ \Sigma(Py) \end{bmatrix} = \begin{bmatrix} 6.5763632 \\ 90.723124 \end{bmatrix}$$

Solve exactly the following:

$$\left(X^T X\right)b = X^T y$$

Thus:

$$\begin{bmatrix} 4 & 50 \\ 50 & 750 \end{bmatrix}\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 6.5763632 \\ 90.723124 \end{bmatrix}$$

This is the following system:

$$4b_0 + 50b_1 = 6.5763632$$

$$50b_0 + 750b_1 = 90.723124$$

It is easy to derive that the final result is as follows:

$$b_0 = 0.7922324 \; , \; b_1 = 0.068148672$$

So, the fitted positive-regime mean model is as follows:

$$\mu(t+1) = 0.7922324 + 0.068148672 P(t)$$

Final prediction for a new day

Assume today is day *t* and precipitation is as follows:

$$P(t) = 8\,\text{mm/day}$$

Step 1: zero probability prediction (intercept-only)

Earlier, we estimated the following:

$$\hat{\pi} = \frac{1}{3}, \text{ and } 1 - \hat{\pi} = \frac{2}{3}$$

Step 2: positive-regime mean

$$\mu(t+1) = 0.7922324 + 0.068148672 \times 8 = 1.3374220$$

Step 3: positive-regime concentration

We used $s = 1$, and $\tau = 1$ day, so, from Eq. (25), the following can be written:

$$\hat{C}_{\text{in}}^{+}(t+1) = \sinh\left[\mu(t+1)\right] = \sinh(1.3374220)$$

By definition, one may write the following:

$$\sinh(u) = \frac{e^{u} - e^{-u}}{2}$$

For $u = 1.3374220$, the final result is as follows:

$$\sinh(1.3374220) = 1.7736$$

Thus:

$$\hat{C}_{\text{in}}^{+}(t+1) = 1.7736\,\text{kg/m}^3$$

Step 4: final mixture prediction

$$\hat{C}_{\text{in}}(t+1) = \left(1 - \hat{\pi}\right)\hat{C}_{\text{in}}^{+}(t+1) = \frac{2}{3} \times 1.7736$$

The final result is as follows:

$$\hat{C}_{\text{in}}(t+1) \approx 1.1824\,\text{kg/m}^3$$

Thus, one may conclude the following:

(1) The model predicts the following target, directly, with $\tau = 1$ day:

$$\hat{C}\mathrm{in}\left(t+\tau\right)$$

(2) It replaces the classical log + epsilon idea by explicitly modeling two physically distinct regimes:

(2.1) **zero regime** with probability π,

(2.2) **positive regime** with probability $(1 - \pi)$.

(3) It avoids ln (0) without any ε by using the asinh transform, which is defined at zero.

(4) The final prediction is the following probability-weighted, derived from Eqs. (25) and (26a):

$$\hat{C}\mathrm{in}\left(t+\tau\right)=\left[1-\hat{\pi}\left(t+\tau\right)\right]s\sinh[\mu\left(t+\tau\right)]$$

It can be also simply written in the following form:

$$\hat{C}\mathrm{in}\left(t+\tau\right)=\left(1-\hat{\pi}\right)\hat{C}_{\mathrm{in}}\left(t+\tau\right)$$

This form makes the following interpretation explicit:

4.1. If

$$\hat{\pi}=1$$

sediment never occurs → predicted concentration is zero,

4.2. If

$$\hat{\pi}=0$$

sediment always occurs → prediction reduces to the positive-regime model,

4.3. If

$$0<\hat{\pi}<1$$

sediment occurs intermittently → prediction is a probability-weighted expectation.

(5) $\hat{\pi}$ = Probability of zero sediment concentration, which can be mathematically written as follows:

$$\hat{\pi}=\mathbb{P}\left[C_{\mathrm{in}}\left(t_i+\tau\right)=0\right]$$

It quantifies the frequency of hydrologically inactive or sediment-free periods.

$\left(1 - \hat{\pi}\right)$ = Probability of positive sediment concentration (sediment occurrence), which can be translated in the following mathematical form:

$$\left(1 - \hat{\pi}\right) = \mathbb{P}\left[C_{\text{in}}\left(t_i + \tau\right) > 0\right]$$

It quantifies the frequency of sediment-producing conditions.

This separation reflects the intermittent nature of sediment transport, which classical log-scale or rating-curve models cannot represent explicitly.

The above definition follows directly from the two-regime structure of the predictive model: (1) Zero-sediment regime, representing periods with no suspended sediment transport (quiescent or baseflow conditions), (2) Positive-sediment regime, representing periods of active sediment transport driven by hydrological forcing.

(6) The Bernoulli variable governing regime selection is defined as follows:

$$Z\left(t + \tau\right) = \begin{cases} 1, & C_{\text{in}}\left(t + \tau\right) = 0 \\ 0, & C_{\text{in}}\left(t + \tau\right) > 0 \end{cases}$$

with

$$\mathbb{P}\left[Z = 0\right] = \pi \text{, and } \mathbb{P}\left[Z = 1\right] = 1 - \pi$$

Physical consistency and predictive superiority of the proposed two-regime sediment modeling framework

Accurate prediction of inflow suspended sediment concentration remains a challenging problem due to the inherently intermittent and event-driven nature of sediment transport processes. Traditional modeling approaches, whether based on log-transformed regressions, empirical rating curves, or standard time-series techniques, often rely on mathematical constructs that do not fully reflect the underlying physical mechanisms. In particular, the widespread use of logarithmic transformations with arbitrary offset parameters introduces conceptual inconsistencies and numerical artifacts, especially in the presence of zero or near-zero sediment concentrations. This highlights a fundamental mismatch between model formulation and physical reality, where periods of no transport and active transport are treated within a single continuous framework.

In response to this limitation, the present work introduces a fundamentally different modeling perspective grounded in the explicit recognition of two distinct sediment transport regimes. By separating the occurrence of sediment transport from its magnitude, the proposed framework aligns the mathematical structure of the model with the physical behavior of the system. This leads to a probabilistic formulation that is both analytically rigorous and physically interpretable, avoiding ad hoc numerical corrections and providing a direct prediction of the quantity of interest. The following discussion aims to clarify the conceptual foundations of this approach, emphasize its methodological significance, and position it within the broader context of predictive modeling in hydro-sedimentology.

Nature of the model: A fundamental departure from classical approaches

The model developed in this study does not constitute a mere statistical variant among existing approaches such as ARIMA models, Markov chains, or Kalman filters; rather, it is founded on a structural reformulation of the underlying physical problem. Indeed, the classical literature largely relies on transformations of the form log $(C + \varepsilon)$, where ε is an arbitrary parameter introduced solely to circumvent the mathematical indeterminacy at zero, without any physical meaning. As rigorously demonstrated in this work, ε has no hydraulic or sedimentological interpretation, introduces systematic bias, particularly at low concentrations, and obscures a fundamental physical reality: the existence of two distinct regimes, namely the absence and the presence of sediment transport. Consequently, the proposed model does not aim to refine or adjust an existing regression framework, but rather to resolve a foundational conceptual inconsistency inherent in conventional approaches.

Main contribution: explicit separation of physical mechanisms

The central contribution of this work lies in the recognition that suspended sediment concentration time series do not constitute a single continuous process, but rather a two-regime system: a zero-transport regime, corresponding to the physical absence of sediment, and an active-transport regime, associated with genuine erosion and transport processes. This distinction is grounded in physical reality and not merely in statistical convenience. Consequently, conventional approaches such as ARIMA models or Markov chains implicitly describe a single continuous dynamic, whereas the proposed framework introduces a mixed probabilistic structure combining a Bernoulli process for occurrence and a conditional model for magnitude. This represents a genuine shift in paradigm, moving from empirical time-series formulations toward a probabilistic model explicitly governed by the physics of sediment transport.

Comparison with ARIMA, Kalman, and Markov models

ARIMA and Kalman-based models inherently assume a continuous signal representation and do not explicitly account for zero values as a distinct physical regime. In practice, they often rely on transformations such as logarithmic or Box-Cox forms, which indirectly reintroduce the difficulties associated with the ε term used to handle zero concentrations. While Markov chain approaches are capable of representing discrete states, they do not directly yield a predictive law for sediment concentration with a clear physical interpretation of magnitude. In contrast, the proposed framework explicitly models sediment occurrence through a probability of transport, separates it from the conditional modeling of concentration magnitude, and provides a closed-form predictive expectation that remains both mathematically consistent and physically interpretable.

Model validation

Regarding the importance of validation using real-world time series, it is useful to distinguish between two complementary levels of analysis.

(a) Fundamental level (Scope of the present work)

The primary objective of this study is to establish a coherent governing model, eliminate mathematical artifacts such as the ε parameter, and introduce a predictive structure grounded in physical principles. At this stage, the emphasis is placed on ensuring conceptual consistency and mathematical rigor, both of which are rigorously demonstrated within the framework of the proposed formulation.

(b) Applied level (Future work)

Extensive validation, as highlighted, naturally constitutes a necessary subsequent stage. This includes application to long-term datasets across multiple sites, a diversity of climatic conditions, and systematic quantitative comparison using performance metrics such as RMSE, NSE, log-score, and related criteria. This phase is explicitly identified in the study as a key direction for future research and model assessment.

Key point

The most essential point is the following: classical models attempt to correct a mathematical issue, namely the indeterminacy of log(0), whereas the proposed framework recognizes that this is fundamentally a physical problem corresponding to the absence of sediment transport. This distinction provides the basis for introducing a Bernoulli process to model occurrence and a conditional model to represent transport intensity.

Constructive outlook

The need to evaluate the model using diverse real-world datasets and to conduct systematic comparisons with competing approaches is fully acknowledged. Such developments represent a natural continuation of the present work and will provide a rigorous basis for quantifying the improvements and added value introduced by the proposed formulation.

Summary statement

In summary, the proposed model should not be viewed as a mere statistical alternative, but rather as a reconstruction of the modeling framework itself, aimed at aligning the mathematical formulation with the physical reality of sediment transport processes.

Main Findings

The first major finding of this study is that classical sediment concentration models are not predictive by construction. Approaches based on log-transformed sediment concentration or empirical rating curves implicitly assume a single continuous regime and require arbitrary numerical corrections, such as the introduction of an offset parameter to handle zero values. These practices obscure the physical meaning of the modeled variable and fail to represent the intermittent nature of sediment transport observed in measured time series.

A second key finding is that suspended sediment concentration time series are inherently two-regime processes. Observations clearly demonstrate the coexistence of periods with zero sediment transport and periods with active transport. Treating these fundamentally different physical states within a single regression framework leads to conceptual inconsistency. The explicit introduction of a binary zero indicator allows sediment occurrence to be modeled as a distinct process rather than as a numerical anomaly.

The study further demonstrates that sediment occurrence and sediment magnitude are governed by different mechanisms and must be modeled separately. The occurrence of sediment transport is naturally described by a Bernoulli process, whose probability can be linked to hydrometeorological predictors, while the magnitude of sediment concentration, conditional on occurrence, can be described by a continuous model. This separation provides a physically meaningful and statistically coherent framework that is absent from classical approaches.

Another important finding concerns the use of the inverse hyperbolic sine (asinh) transformation for modeling positive sediment concentrations. Unlike the logarithmic transformation, the asinh transformation is exactly defined at zero and does not require any artificial offset parameter. It behaves linearly at low concentrations and logarithmically at high concentrations, thereby preserving sensitivity near transport thresholds while preventing extreme values from dominating the model. This transformation yields a stable and physically interpretable representation of sediment intensity.

The proposed two-regime framework leads to a closed-form predictive equation for inflow suspended sediment concentration, expressed as a probability-weighted expectation that combines sediment occurrence probability and conditional sediment magnitude. This formulation directly targets the quantity of interest at a specified lead time and avoids back-transformation bias. The predictive structure is transparent and grounded in probability theory rather than empirical curve fitting.

The numerical example confirms the internal consistency and interpretability of the model. Estimated probabilities of zero and positive sediment occurrence reflect the observed frequency of sediment-free and sediment-active conditions. The resulting predictions are naturally moderated by occurrence probability, preventing systematic overprediction during periods when sediment transport is unlikely. This behavior contrasts sharply with classical models, which implicitly assume continuous sediment presence.

Finally, the study finds that the proposed framework offers clear advantages for reservoir sediment management and forecasting applications. By explicitly predicting both the likelihood and magnitude of sediment inflow, the model provides actionable information for operational decision-making, sediment routing strategies, and long-term reservoir sustainability assessments. The framework is flexible, extensible to multiple predictors, and compatible with both data-driven and physically based modeling approaches.

CONCLUSION

This study has addressed a long-standing and fundamental limitation in the modeling and prediction of inflow suspended sediment concentration to dam reservoirs. Classical approaches, whether based on log-transformed regression models or empirical rating curves of the form $C = \mathrm{a}Q^{\mathrm{b}}$, have been shown to be inadequate for predictive purposes, as they rely on indirect variable transformations, introduce arbitrary numerical corrections, implicitly assume a single continuous regime, and fail to represent the intermittent and event-driven nature of sediment transport. In particular, the widespread use of logarithmic transformations with an artificial offset parameter ε masks the physical distinction between zero-transport and active-transport conditions and leads to bias and ambiguity at low concentrations.

In response to these shortcomings, this paper has developed a new probabilistic two-regime governing model for predicting inflow suspended sediment concentration. The proposed framework is grounded in a clear physical interpretation of sediment dynamics, recognizing that measured sediment time series are characterized by two fundamentally distinct regimes: a zero or negligible transport regime and a positive, active transport regime. Rather than forcing these regimes into a single regression equation, the model explicitly separates sediment occurrence from sediment magnitude.

The zero-transport regime is modeled through a Bernoulli process, whose probability is represented by a logistic formulation that maps hydrometeorological predictors into a well-defined probability of zero sediment occurrence. Conditional on sediment presence, the magnitude of suspended sediment concentration is modeled using a positive-regime formulation based on the inverse hyperbolic sine transformation. This transformation preserves logarithmic behavior at high concentrations while remaining exactly defined at zero, thereby eliminating the need for any arbitrary numerical offset and providing a physically interpretable scale parameter.

A key outcome of the proposed approach is the derivation of a closed-form predictive expression for inflow suspended sediment concentration, expressed as a probability-weighted expectation that combines sediment occurrence and conditional magnitude into a single forecast at a specified lead time. All governing equations were derived explicitly, and parameter estimation procedures were formulated in a transparent and reproducible manner using likelihood-based and least-squares principles. A fully detailed numerical example demonstrated the internal consistency of the model, the interpretation of estimated probabilities, and the practical steps required for implementation.

Compared to classical log-scale models and rating-curve formulations, the proposed two-regime framework offers several decisive advantages. It directly targets the physical quantity of interest, future inflow suspended sediment concentration, rather than an intermediate transformed variable; it removes arbitrary tuning parameters and preserves dimensional consistency; it provides a probabilistic interpretation of sediment occurrence; and it naturally accounts for the intermittent nature of sediment transport. By decoupling the processes governing sediment occurrence from those governing sediment intensity, the model avoids conceptual and numerical artifacts inherent in single-regime approaches.

Beyond its theoretical contribution, the proposed model has direct practical implications for reservoir sediment management. By explicitly predicting both the likelihood and magnitude of sediment inflow at a given forecast horizon, the framework provides a sound basis for anticipating sediment pulses, optimizing sediment routing operations, and supporting operational decision-making in dam reservoirs, particularly in arid and semi-arid regions where sedimentation remains a critical concern.

Future work may extend the proposed framework by incorporating additional hydrological predictors, exploring spatial variability, coupling the model with physically based runoff and erosion modules, and validating its performance on long-term observed datasets across diverse climatic and geomorphological contexts. Nevertheless, the present study establishes a robust, physically consistent, and predictive alternative to classical sediment concentration models and provides a new foundation for probabilistic sediment forecasting in reservoir systems.

Declaration of competing interest

The author declares that he has no known competing financial interests or personal relationships that could influence the work reported in this article.

REFERENCES

ABD RAHMAN A.N., OTHMAN F., WAN JAAFAR W.J., AHMED ELSHAFIE A.H.K. (2023). An assessment of floods' characteristics and patterns in Pahang, Malaysia, Larhyss Journal, No 55, pp. 89-105.

ACHOUR B., MEHTA D., AZAMATHULLA H. (2024). A New Trapezoidal Flume for Open Channel Flow Measurement Design, Theory, and Experiment, Larhyss Journal, No 59, pp. 157-179.

ANNANDALE G.W. (2013). Quenching the Thirst: Sustainable Water Supply and Climate Change, Book, CreateSpace Independent Publishing Platform (a self-publishing imprint of Amazon/Createspace), North Charleston, USA, 250p.

ASSELMAN N.E.M. (2000). Fitting and interpretation of sediment rating curves, Journal of Hydrology, Vol. 234, pp. 228-248.

ATHMANI H., BOUKEHLIFI KOUIDER D., BENSEFIA S., DJAFRI S.A. (2025). Flood risk assessment in arid regions based on hydraulic modeling with HEC-RAS. case study of wadi Tamda in Doucen, Algeria, Larhyss Journal, No 61, pp. 81-109.

BAUDHANWALA D., KANTHARIA V., PATEL D., MEHTA D., WAIKHOM S. (2023). Applicability of SWMM for urban flood forecasting a case study of the western zone of Surat city, Larhyss Journal, No 54, pp. 71-83.

BEAR J. (1972). Dynamics of Fluids in Porous Media, Book, 1st Edition, American Elsevier Publishing Company, Inc., New York, NY, USA, 764p.

BENFETTA H., OUADJA A., ACHOUR B., REMINI B. (2016). Capacity loss in dams located in arid and semi-arid zones. case of Gargar, Bouhanifia, Ouizert and Foum El Gherza dams, Larhyss Journal, No 25, pp. 183-201. (In French)

BEN SAID M., HAFNAOUI M.A., HACHEMI A., MADI M., BENMALEK A. (2024). Evaluating the effectiveness of the existing flood risk protection measures along wadi Deffa in El-Bayadh city, Algeria, Larhyss Journal, No 59, pp. 7-28.

BOUGAMOUZA A., REMINI B., SAKHRAOUI F. (2020). Analytical study of sediment evolution in the lake of the Foum El Gherza dam (Biskra, Algeria), Larhyss Journal, No 43, pp. 169-179.

CHADEE A.A., RATHORE K., CHOUDHARY L.A., VERMA S., MEHTA D. (2024). The korba coal mining zone in India assessment of risk health and pollutant sources, Larhyss Journal, No 60, pp. 113-131.

CHADEE A., NARRA M., MEHTA D., ANDREW J., AZAMATHULLA H. (2023). Impact of climate change on water resource engineering in Trinidad and Tobago, Larhyss Journal, No 55, pp. 215-229.

DO T.V.H., PHAM H.G., KIEU Q.L., TRAN T.N.H. (2025). The typical mechanisms and factors leading to flash floods in small watersheds in the mountainous region of Vietnam, a case study in the CHU VA stream watershed, Larhyss Journal, No 61, pp. 141-168.

EZZ H. (2025). Unexpected flooding in Mersa Matruh, Egypt - Investigating causes, hydrological analysis, and flood risk assessment, Larhyss Journal, No 61, pp. 371-399.

GRAF W.H. (1984). Hydraulics of Sediment Transport, Water Resources Publications, LLC, Highlands Ranch, Colorado, USA, 513p.

HELSEL D.R., HIRSH R.M. (2002). Statistical Methods in Water Resources, U.S. Geological Survey (USGS) Techniques of Water-Resources Investigations, Book 4, Chapter A3, U.S. Geological Survey (USGS), Reston, Virginia, VA, USA.

ICOLD (International Commission on Large Dams). (1989). Sedimentation Control of Reservoirs, Bulletin No 67, Paris, France.

KARPATNE A., WATKINS W., READ J., KUMAR V. (2017). Physics-guided neural networks (PGNN): An application in lake temperature modeling, Publisher Identifier: arXiv preprint arXiv:1710.11431, submitted October 31, 2017, and subsequently updated in versions.

LEWIS J. (1996). Turbidity-controlled suspended sediment sampling for runoff-event load estimation, Water Resources Research, Vol. 32, Issue 7, pp. 2299-2310.

MEHTA D., YADAV S.M. (2021). An analysis of rainfall variability and drought over Barmer District of Rajasthan, Northwest India, Water Supply, Vol. 21, Issue 5, pp. 2505-2517.

MEHTA D., YADAV S. (2024). Rainfall runoff modelling using HEC-HMS model: case study of Purna river basin, Larhyss Journal, No 59, pp. 101-118.

MEHTA D.J., PRAJAPATI K.J. (2017). Simulation of existing water distribution network at Punagam Area of Surat City using WaterGEMS software, In ASCE India Conference 2017, December 12-14, Reston, VA: American Society of Civil Engineers (ASCE), pp. 312-321.

MEHTA D., ACHOUR B., PASTAGIA J., AZAMATHULLA H., VERMA S. (2023). Review of reservoir operation, Larhyss Journal, No 56, pp. 193-214.

MEZENNER N., BENKACI T., BERMAD A., DECHEMI N. (2022). Dam reservoir operation optimization using genetic algorithm and principal component analysis simulation model - case of dam Ghrib, Larhyss Journal, No 51, pp. 145-160.

MONTGOMERY D.C., PECK E.A., VINING G.G. (2012). Introduction to Linear Regression Analysis, Book, 5th Edition, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, New Jersey, NJ, New York, 672p.

MORRIS G.L., FAN J. (1998). Reservoir Sedimentation Handbook: Design and Management of Dams, Reservoirs, and Watersheds for Sustainable Use, McGraw-Hill Professional (McGraw-Hill Book Co.), New York, USA, 848p.

PALMIERI A., SHAH F., ANNANDALE G.W. (2003). Reservoir conservation - The RESCON approach, Journal of Hydraulic Research, Vol. 41, Issue 4, pp. 373-384.

PANCHAL S.L., SURYANARAYANA T.M.V. (2025). Optimized operation of a multipurpose reservoir by evolutionary algorithm for Panam reservoir project in eastern Gujarat, India, Larhyss Journal, No 61, pp. 31-52.

PARKER G. (2004). 1D Sediment Transport Morphodynamics with Applications to Rivers and Turbidity Currents, E-book / Lecture notes / Morphodynamics monograph, University of Illinois at Urbana-Champaign, Department of Civil & Environmental Engineering, Ven Te Chow Hydrosystems Laboratory, USA.

RAISSI M., PERDIKARIS P., KARNIADAKIS G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems, Journal of Computational Physics, Vol. 378, pp. 686-707.

REICHSTEIN M., CAMPS-VALLS G., STEVENS B., JUNG M., DENZLER J. (2019). Deep learning and process understanding for data-driven Earth system science, Nature, Vol. 566, pp. 195-204.

REMINI W., REMINI B. (2003). Sedimentation in north African dams, Larhyss Journal, No 2, pp. 45-54. (In French)

REMINI B. (2010). A new approach to fight against dams' siltation: the emerged obstacle technique, Larhyss Journal, No 9, pp. 43-53. (In French)

REMINI B., BENSAFIA D. (2016). Siltation of dams in arid regions - Algerian examples, Larhyss Journal, No 27, pp. 63-90. (In French)

REMINI B., TOUMI A. (2017). The Beni Haroun reservoir (Algeria) is it threatened by siltation? Larhyss Journal, No 29, pp. 249-263.

REMINI B. (2017). A new management approach of dams' siltation, Larhyss Journal, No 31, pp. 51-81. (In French)

REMINI B., BOUABIBSA R., MOUDJED K. (2019). Beni Haroun and Koudiat Acerdoune (Algeria): two large dams threatened by the phenomenon of siltation, Larhyss Journal, No 38, pp. 131-151. (In French)

REMINI B. (2022). Sustainable desilting of dams, Larhyss Journal, No 51, pp. 211-236.

SUTHERLAND R.A., GRAY J.E., CLIFTON G.E. (2000). The use of turbidity measurements to estimate suspended sediment concentrations in rivers, Hydrological Processes, Vol. 14, pp. 1499-1511.

TOUMI A., REMINI B. (2020). Zardezas (Algeria): a dam that is silting up? Larhyss Journal, No 43, pp. 181-196.

USGS (United States Geological Survey). (2006). Techniques and Methods for Estimating Suspended-Sediment Concentration and Load, U.S. Geological Survey, Book 3, Chapter C2.

VERMA S., SAHU R.T., PRASAD A.D., VERMA M.K. (2023). Reservoir operation optimization using ant colony optimization a case study of Mahanadi reservoir project complex Chhattisgarh - India, Larhyss Journal, No 53, pp. 73-93.

WALLING D.E. (1977). Assessing the accuracy of suspended sediment rating curves for a small basin, Water Resources Research, Vol. 13, Issue 3, pp. 531-538.

WU W.S. (2007). Computational River Dynamics, Book, 1st edition (standard hardcover academic edition), Chemical Rubber Company (CRC) Press, 845p.

ZEGAIT R., PIZZO H.S. (2023). Flood control reservoir using VBA simulation case of Idles basin in southern Algeria, Larhyss Journal, No 53, pp. 41-60.