



Stability criteria for Bayesian calibration of reservoir sedimentation models

Kilian Mouris¹ · Eduardo Acuna Espinoza^{1,2} · Sebastian Schwindt¹ · Farid Mohammadi¹ · Stefan Haun¹ · Silke Wieprecht¹ · Sergey Oladyshkin¹

Received: 16 November 2022 / Accepted: 18 January 2023 / Published online: 3 February 2023
© The Author(s) 2023

Abstract

Modeling reservoir sedimentation is particularly challenging due to the simultaneous simulation of shallow shores, tributary deltas, and deep waters. The shallow upstream parts of reservoirs, where deltaic avulsion and erosion processes occur, compete with the validity of modeling assumptions used to simulate the deposition of fine sediments in deep waters. We investigate how complex numerical models can be calibrated to accurately predict reservoir sedimentation in the presence of competing model simplifications and identify the importance of calibration parameters for prioritization in measurement campaigns. This study applies Bayesian calibration, a supervised learning technique using surrogate-assisted Bayesian inversion with a Gaussian Process Emulator to calibrate a two-dimensional (2d) hydro-morphodynamic model for simulating sedimentation processes in a reservoir in Albania. Four calibration parameters were fitted to obtain the statistically best possible simulation of bed level changes between 2016 and 2019 through two differently constraining data scenarios. One scenario included measurements from the entire upstream half of the reservoir. Another scenario only included measurements in the geospatially valid range of the numerical model. Model accuracy parameters, Bayesian model evidence, and the variability of the four calibration parameters indicate that Bayesian calibration only converges toward physically meaningful parameter combinations when the calibration nodes are in the valid range of the numerical model. The Bayesian approach also allowed for a comparison of multiple parameters and found that the dry bulk density of the deposited sediments is the most important factor for calibration.

Keywords Bayesian calibration · Bayesian inference · Metamodel · Bayesian active learning · Calibration parameter importance · Reservoir sedimentation

Introduction

Artificial reservoirs are crucial infrastructure for providing drinking water, water for irrigation, flood protection, recreation, and hydroelectric power (Zarfl et al. 2015; Schleiss et al. 2016; Kim et al. 2020). However, reservoirs interrupt the longitudinal continuity of fluvial systems (Hinderer et al. 2013; Sun et al. 2021). For instance, low flow velocities lead to sediment deposition in reservoirs. The deposited sediment

is missing in downstream reaches and reduces the active storage capacity of reservoirs (Kondolf 1997). To minimize sediment deposition and ensure sustainable reservoir operation, it is essential to quantify and accurately predict sedimentation processes. State-of-the-art tools for predicting reservoir sedimentation are two (2d) or three (3d) dimensional numerical models coupling hydrodynamics and sediment transport (Haun et al. 2013; Hanmaiahgari et al. 2018; Olsen and Hillebrand 2018; Khorrami and Banihashemi 2021).

Advances in numerical methods and computing power have led to remarkable improvements in the accuracy and speed of numerical models. Every numerical model requires calibration, which is a subjective and time-consuming process. Calibration is particularly important because the equations used in numerical models are based on simplified assumptions that are partly empirical. To calibrate a model, uncertain calibration parameters are adjusted within

✉ Kilian Mouris
kilian.mouris@iws.uni-stuttgart.de

¹ Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, 70569 Stuttgart, Germany

² Institute of Water and River Basin Management – Hydrology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

a physically reasonable range to achieve a good agreement between modeled and measured data with appropriate tolerance (Simons et al. 2000; Oberkampff et al. 2004; Paul and Negahban-Azar 2018). A common approach to calibrating numerical models is the iterative trial-and-error adaption of calibration parameters. However, this method is time-consuming, labor-intensive, and subjectively biased because it does not account for uncertainty in measured data, modeling errors, nor equifinality (Schmelter and Stevens 2013; Muehleisen and Bergerson 2016; Beckers et al. 2020). While Bayesian inference, a type of stochastic calibration, can address some limitations, it requires many iterations and is therefore not practical for use with computationally intensive models (e.g., hydro-morphodynamic models to simulate reservoir sedimentation). Mohammadi et al. (2018), Beckers et al. (2020), and Scheurer et al. (2021) overcame this challenge using metamodels (also known as surrogate models, response surface, reduced model, etc.) to replicate the full complexity of a deterministic numerical model. These studies employed metamodel updating to reduce the total number of evaluations of the original model required to train the metamodel.

Modeling reservoir sedimentation requires specific simplifying assumptions regarding hydrodynamics and morphodynamics. In comparison to the simulation of rivers, fluctuating water levels and outflow conditions due to reservoir operation, and the simultaneous simulation of very shallow shores and tributary deltas along with deep waters are particularly challenging in reservoir modeling. For instance, wetting and drying of mesh nodes at the shoreline of a reservoir require model simplifications (e.g., the definition of a minimum water depth for a cell). In addition, channel erosion and deltaic avulsion might occur at the head of the reservoir. These erosion and avulsion processes and their exact location are hard to predict and result from stochastic environmental forcing (Hajek and Wolinsky 2012; Chadwick et al. 2019). Furthermore, these processes are still an open research topic (e.g., Langendoen et al. 2016) and difficult to simulate accurately, much less with the same model simplifications as the deposition of fine sediments in deeper waters. As a result of global model assumptions, some regions of a reservoir model may not be accurately represented by the numerical model. This is because the model is generally calibrated to accurately represent either fine sediment deposition in deep waters or delta progression and erosion processes at the head of the reservoir, but not both. *This is why we are investigating in this study how complex numerical models for reservoir sedimentation can be calibrated in light of competing model simplifications.* To this end, we test the hypothesis (i) that Bayesian calibration only converges toward physically meaningful calibration parameter combinations when the model is well-conditioned

(i.e., measured data are in the validity domain of model assumptions). The verification of this hypothesis aims to enrich the scientific baseline for modeling complex hydro-morphodynamic processes in reservoirs, which inherently require modeling regions that may be physically invalid. To test this hypothesis, we adapt a Bayesian calibration technique that uses surrogate-assisted Bayesian inversion with a metamodel in the form of a Gaussian Process Emulator (GPE) according to Oladyshkin et al. (2020). The metamodel and its updating build on Bayesian active learning (BAL), which we further improve through the cumulative consideration of measurement and metamodel errors. To test hypothesis (i), we introduce two spatially distinct measurement data scenarios for calibrating a 2d hydro-morphodynamic reservoir sedimentation model of the large Banja reservoir in Albania.

Bayesian calibration typically starts with the definition of calibration parameters and the corresponding physically meaningful parameter ranges (e.g., Kim and Park 2016; Beckers et al. 2020). Based on initial model tests, we selected the four most sensitive parameters in the form of dry-bulk density of deposited sediments ρ_b , critical shear stress for erosion τ_{cr} , critical shear stress for deposition τ_d , and a diameter multiplier γ that defines the grain size distribution. The large number of four calibration parameters presents a challenge for any calibration process and results in a four-dimensional parameter space with millions of combination options, leading to problems regarding maximum floating-point precision. Hence, we implemented optimization strategies for Bayesian calibration intending to bypass precision errors (arithmetic underflow) caused by the multidimensional space of possible calibration parameter combinations. Furthermore, these parameters carry a high degree of uncertainty that must be thoroughly considered during modeling (Schmelter et al. 2015; Villaret et al. 2016). The grain size distribution, the two critical bed shear stresses for cohesive sediments, and the dry-bulk density can only be determined with great effort by field sampling. Therefore, it is important to identify and prioritize the most important parameters when planning field data collection. This insight enables the development of optimized measurement concepts, to reduce costs and workload. *Hence, we investigate whether our modified Bayesian calibration enables the identification of driving calibration parameters for modeling reservoir sedimentation even in a four-dimensional parameter space.* By examining the importance of four potentially important parameters driving reservoir sedimentation, we test the hypothesis (ii) that at least one of the four calibration parameters plays a dominant role in the fluvial deposition of suspended load in reservoirs. Therefore, we aim to identify the most important calibration parameter that should be addressed in sampling campaigns at reservoirs.

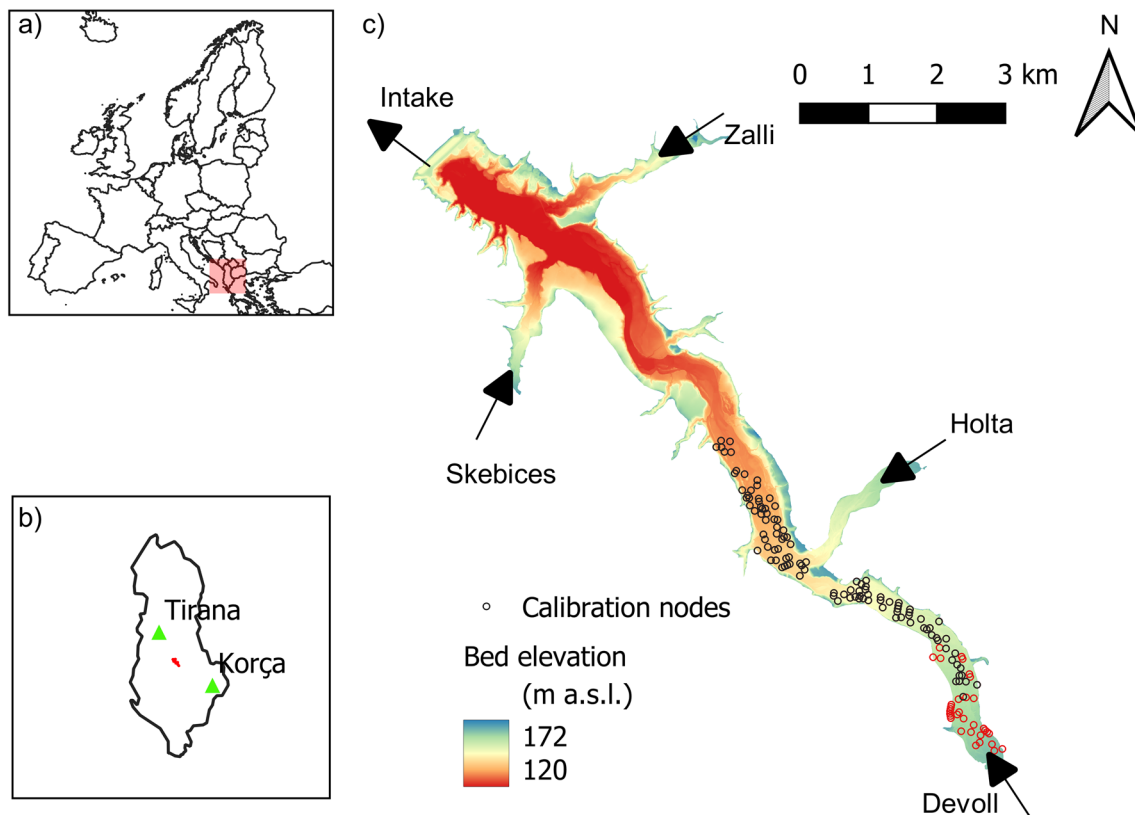


Fig. 1 Location of the study area; **a** European context, **b** national context, and **c** the bathymetry of the Banja reservoir with indication of the calibration nodes, major tributaries and turbine intake. The red calibration nodes are excluded in the VALDOME data scenario

Materials and methods

Study area

The Banja Reservoir

In this study, we numerically simulated hydro-morphodynamic processes in the Banja Reservoir at the Devoll River in central Albania. With a length of 196 km, the Devoll River is the third longest river in Albania and has its source in the Gramos Mountains near the Greek border. The river flows northwestward and is dammed after approximately 160 km, forming the Banja reservoir (see Fig. 1). The reservoir was commissioned in 2016 and has a length of 14 km, a maximum water depth of 60 m close to the dam, and a surface area of approximately 14 km², leading to a storage capacity of approximately 400 million m³. It is mainly fed by the Devoll River (89%, $MQ \approx 33 \text{ m}^3 \text{ s}^{-1}$), Holta River (9%), and two smaller tributaries (Zalli and Skebices River, 1% each). The catchment of the reservoir is characterized by dry and hot summers and wet winters, resulting in low summer, high winter, and high spring flows. Since snowfall is frequent in winter at high elevations, the flow regime is driven by precipitation and

snowmelt. The sediment yield of the Banja catchment is particularly high due to high rainfall erosivity on steep terrains composed of loose soils (Walling and Webb 1996; Borrelli et al. 2020; Mouris et al. 2022).

Measurement data

The initial bathymetry was interpolated onto a numerical mesh from a photogrammetry-based digital elevation model (DEM) from 2016, before filling the reservoir. In addition, the reservoir bathymetry was measured in 2019 with an acoustic Doppler current profiler (ADCP) boat providing approximately 632×10^3 bed level measurements.

The grain size distribution of the suspended sediment was determined based on suspended sediment measurements at the Devoll River upstream of the reservoir (Ardiclioglu et al. 2011) and reservoir bed samples. The per-sample median diameters of the deposited sediment ranged from 5.7 to 37.4 μm with a mean of 10.5 μm , emphasizing the cohesive nature of the deposits. Upstream of the reservoir, the extracted granulometric curve had cohesive characteristics, with 98% of the volume having grain diameters smaller than 60 μm . Neither cobble, gravel, nor coarse sand was present in the study area. Consequently, bedload was not considered

Table 1 Used input data for the Bayesian calibration of the Banja reservoir

Data set name	Survey method	Data type	Purpose in this study
Digital elevation model (2016)	Aerial survey	Georeferenced raster	Initial bathymetry
Measured bed levels (2019)	Bathymetric survey	Georeferenced shapefiles	Target (calibration) bathymetry
Water levels and outflow	Direct measurement	Text files	Liquid boundary conditions (numerical model)
Inflow discharge	Calculated based on water level and outflow	Text files	Liquid boundary conditions (numerical model)
Grain size distribution	Sediment samples (Ardiclioglu et al. 2011) and field survey in 2021	Text files	Used to define the range for the three grain size fractions
Suspended sediment concentration	Soil erosion and sediment transport model calibrated upstream of the reservoir (Mouris et al. 2022)	Text files	Boundary conditions for numerical model
Sediment samples (reservoir bed)	Measurement campaign	Text files	Verification of model assumptions

in the numerical model. The available measurement data for this study are summarized in Table 1.

Numerical full-complexity model

General setup

In this study, we used Telemac-2D (Hervouet 2007) with its sediment transport and bed evolution module GAIA (Audouin et al. 2020) to simulate reservoir sedimentation processes. Telemac-2D abstracts river landscapes with unstructured grids. The here-used unstructured, triangular numerical mesh consisted of 24,241 elements and 12,600 nodes, resulting in element sizes of approximately 40 m. We defined two roughness coefficients to differentiate between the original river course (before filling) and the newly wetted areas. Due to the low flow velocities, the influence of boundary roughness on reservoir hydrodynamics was small and we applied Manning coefficients of $0.032 \text{ s m}^{-1/3}$ (original cobble-gravel-bed river) and $0.06 \text{ s m}^{-1/3}$ elsewhere (many trees and brushes were not removed before the impoundment of the reservoir).

Telemac-2D approximates the shallow water equations with a combined explicit-implicit solver to calculate the flow field. The hydrodynamic module passes the calculated hydrodynamic variables (water depth, depth-averaged flow velocity) and bed shear stress to the GAIA module. We set the numerical model parameters with the premise of maximizing computational and numeric stability while keeping computing time short. Therefore, we applied a finite element numerical scheme and treated tidal flats (or dry-wet elements) according to software recommendations to use only positive water depths (Hervouet et al. 2011). Furthermore, the method of characteristic solves the advective part of the hydrodynamic equations and improves stability (a result of preliminary model tests). The mixing length turbulence model serves to calculate the turbulent viscosity coefficient,

which is similar to the k- ϵ model when the transverse shear stress is the main turbulence generator, as in the case of a reservoir, but requires 20% less computing time (Dorfmann and Zenz 2016).

To calculate the depth-averaged concentration $C(x, y, t)$ of tracers (i.e., fine particles) in (g L^{-1}), the 2d advection-diffusion-equation is solved.

$$\frac{\partial(hC)}{\partial t} + \frac{\partial(huC)}{\partial x} + \frac{\partial(hvC)}{\partial y} = \frac{\partial}{\partial x} \left(h\epsilon \frac{\partial C}{\partial x} \right) + \frac{\partial}{\partial y} \left(h\epsilon \frac{\partial C}{\partial y} \right) + (E - D) \quad (1)$$

where h is the water depth (m), u (m s^{-1}) and v (m s^{-1}) are the depth-averaged components of flow velocity, ϵ is the turbulent diffusivity of the sediment ($\text{m}^2 \text{ s}^{-1}$), and E and D are the erosion and deposition fluxes ($\text{kg m}^{-2} \text{ s}^{-1}$), respectively. We applied the default treatment of the diffusion term in Eq. (1) to increase numerical stability. In addition, we chose the “Edge-based N-Scheme” to solve the advective term because it provides mass conservative results and treats tidal flats. The erosion and deposition fluxes for cohesive sediment are calculated as follows:

$$D = \omega_s C \left(1 - \frac{\tau_b}{\tau_d} \right) \text{ if } \tau_b > \tau_d \quad (2)$$

$$E = M \left(\frac{\tau_b}{\tau_{ce}} - 1 \right) \text{ if } \tau_b > \tau_{ce} \quad (3)$$

where M is the Krone-Partheniades erosion constant ($\text{kg m}^{-2} \text{ s}^{-1}$), τ_b is the bed shear stress (N m^{-2}), τ_{ce} is the critical shear stress for erosion (N m^{-2}), τ_d is the critical shear stress for deposition (N m^{-2}), and ω_s is the settling velocity (m s^{-1}). The settling velocity is a function of the mean sediment diameter, the ratio of the sediment and water densities, and the kinematic viscosity of the water. The measurement data (see above) had shown that the deposits predominantly

consisted of cohesive sediment, and therefore, only suspended transport was considered in this study.

After determining the erosion and deposition fluxes and calculating the net transport flux per element, GAIA updates the bed level using the Exner equation (Paola and Voller 2005). For a detailed description of the free surface flow and sediment modeling algorithms used, the reader is referred to Hervouet (2007, 2020) and Audouin and Tassi (2020). The steering file for the Telemac-2D simulations is available at Acuna Espinoza et al. (2022).

The focus of the numerical model was on the time-efficient simulation of suspended sediment transport in a large reservoir to ease repetitive calibration runs. Therefore, bedload was not considered and a coarse mesh resolution was used. Due to these simplifying assumptions, but also because of the general limitation of numerical models, it is not possible to accurately predict channel avulsion and erosion through previously deposited cohesive sediments (Hajek and Wolinsky 2012; Liang et al. 2015). Furthermore, channel bank failure depends on the sediment type, moisture content, and seepage processes (Luppi et al. 2009; Rinaldi and Nardi 2013; Olsen and Haun 2020). Therefore, bank failure cannot be simulated with a numerical setup for reservoir sedimentation due to fine particle deposition. Since bank failure processes only occur at the head of the reservoir, the bed level changes in this domain cannot be predicted in a physically correct and stable manner. Still, the above-introduced model setup is valid in deep-water model domains outside of the shallow deposition delta of the Devoll River.

Boundary conditions

For the simulation of reservoir sedimentation over three years, between the two surveys from August 2016 and August 2019, we defined the reservoir inflow Q_{in} ($m^3 s^{-1}$) as a function of measured water levels and measured outflow Q_{out} ($m^3 s^{-1}$) based on a routing equation. More detailed information can be found in SI 1.

Since the suspended sediment concentrations at the tributaries were not known for the simulation period, we implemented a previously developed indirect calculation method (Mouris et al. 2022). The indirect method builds on a calibrated soil erosion and sediment transport model with a monthly resolution ($tons month^{-1}$) to calculate the suspended sediment yield (SSY) of the catchment of the Banja reservoir. We divided the SSY from Mouris et al. (2022) at the Devoll River by the monthly inflow volume to prescribe suspended sediment concentrations (SSC) at the liquid model boundaries. Thus, SSC was constant for every month but varied from month to month. The mean SSC at the Devoll River for the calibration period was $1.36 kg m^{-3}$ with a maximum of $4.0 kg m^{-3}$ in September 2017.

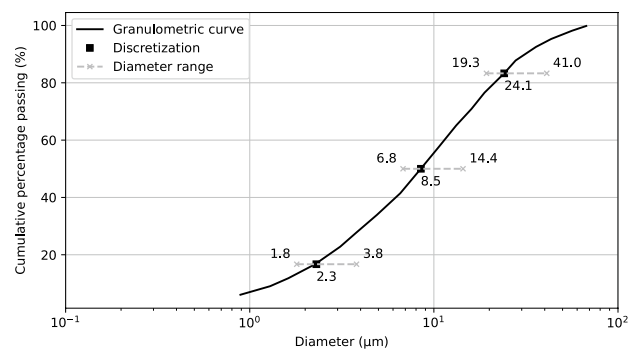


Fig. 2 Granulometric curve with the minimum and maximum grain sizes defined by the γ -multiplier

Calibration parameters

This study optimized four calibration parameters in the form of dry-bulk density of deposited sediments ρ_b , critical shear stress for erosion τ_{cr} , critical shear stress for deposition τ_d , and a diameter multiplier γ for settling velocities. The calibration parameter values were to be adapted to yield a possibly best simulation of the measured bed level changes Δz_{meas} between 2016 and 2019.

The dry-bulk density ρ_b and consolidation processes of mud-sand mixtures strongly depend on the sand content (van Rijn and Barth 2019). Because more than 98% of the deposited sediment in the Banja reservoir is cohesive, we defined ρ_b based on reported literature values for very low ($< 10\%$) sand content. We considered the dry-bulk density a quasi-random variable with equally likely values (i.e., uniformly distributed) between $200 kg m^{-3}$ and $500 kg m^{-3}$ (van Rijn and Barth 2019; van Rijn 2020).

The critical shear stresses for erosion τ_{cr} and deposition τ_d control the exchange rate between suspended and deposited sediment. To define quasi-random, uniformly distributed value ranges for τ_{cr} and τ_d , we referred to field and laboratory tests with sediment mixtures with similar characteristics (grain size distribution, bulk density) as in the Banja reservoir. To this end, we tested value ranges for τ_{cr} between 0.05 and 0.4 Pa (Kornman and Deckere 1998; Widdows et al. 1998; Houwing 1999; Lumborg 2005; Shi et al. 2012; van Rijn 2020), and for τ_d between 0.01 and 0.1 Pa (Krone 1962; Lumborg 2005; Shi et al. 2012).

The deposition pattern in the reservoir also depends on the particle size that drives the settling velocity ω_s (see Eq. (2)). We applied the granulometric curves of suspended sediment upstream of the reservoir, which were subjected to considerable variability (i.e., uncertainty) in the model domain. Figure 2 plots the granulometric curve defined by three diameters representing the lower, middle, and upper third of the total volume. To account for uncertainty, we multiplied every diameter by a factor γ that takes uniformly distributed values between

0.8 and 1.7. Thus, the upper limit of grain sizes was 41 μm, which was larger than 95% of the sediment sample volume. The lower limit of 1.8 μm (2.3 μm • 0.8) was based on preliminary model runs, in which we tested the smallest possible sediment particles that remain in suspension and have an almost negligible influence on the deposition volume. Table 2 shows the resulting value ranges for the four calibration parameters considered in this study.

Bayesian calibration

Bayesian inference

To calibrate a numerical model using Bayesian inference, we inferred the posterior distribution $p(\omega|z_{meas})$ of the model calibration parameters (and hence the corresponding model responses) based on the measured bed levels z_{meas} and defined initial ranges for the calibration parameters. The posterior distribution $p(\omega|z_{meas})$ is the result of evaluating Bayes’ theorem in the context of model updating:

$$p(\omega|z_{meas}) = \frac{p(z_{meas}|\omega) \cdot p(\omega)}{p(z_{meas})} \tag{4}$$

where $p(\omega)$ is the prior probability distribution that defines the initial probability of the calibration parameters before considering new or additional evidence (z_{meas}). $p(z_{meas}|\omega)$ is the so-called likelihood function and indicates how well the metamodel reproduces the measured data z_{meas} given a parameter combination ω . $p(\omega|z_{meas})$ is the posterior probability distribution (i.e., the updated probability of the calibration parameters ω given measured data z_{meas}), which is expected to be narrower than $p(\omega)$ (Box and Tiao 1992; Olyshkin and Nowak 2019). $p(z_{meas})$ is a normalization factor, often referred to as Bayesian model evidence (BME), and is important when different posterior distributions are being compared with each other or several competing models are being evaluated (Mohammadi et al. 2018). Assuming that the deviations between the measured bed levels z_{meas} and the modeled bed levels z_{mod} are normally distributed and independent, the likelihood function $p(z_{meas}|\omega)$ is calculated proportionally to the sum of squared errors δ_i^2 between measured and simulated bed levels $z_{meas} - z_{mod}$ weighted by the total error e_i , where i indicates the calibration node.

$$p(z_{meas}|\omega) \propto \exp\left(-0.5 \left[\frac{\delta_1^2}{e_1^2} + \frac{\delta_2^2}{e_2^2} + \dots + \frac{\delta_n^2}{e_n^2} \right] \right) \tag{5}$$

This study provides additional novelty by improving BAL because of how we implement the measurement error e_{meas} and the metamodel error e_{meta} . In particular, we calculated the total errors e_i for each calibration node i as the sums of $e_{meas,i}$ and $e_{meta,i}$ according to the following descriptions.

The measurement errors e_{meas} resulted from the interpolation of the bed level measurements at the calibration nodes of the numerical mesh and uncertainties of field measurements. We used an interpolation radius of 3 m around the calibration nodes to average the bed level. Thus, the number of measurements per calibration node varied from 1 to 35 (8 on average). The variable amount of measurements available for averaging affected the confidence in the averaged values because, for instance, 15 measurements are more representative than two. The mean measurement error e_{meas} was approximately 0.4 m (measurement precision according to operator) where possible sources of errors were a high concentration of suspended sediment near the bottom, uncertainties in the water level of the reservoir, and the movement of the ADCP boat due to waves. Thus, to calculate the measurement errors $e_{meas,i}$ at every node, we introduce Eq. (6) where s_i is the number of observation points within the 3-m radius:

$$e_{meas,i} = \frac{\text{adapted error}}{1 + \ln(s_i)} \tag{6}$$

where an *adapted error* of 1.02 m was computed iteratively to ensure that the average value of e_{meas} for the total number of calibration nodes was 0.4 m. Thus, for example, $e_{meas,i}$ for a calibration node where the bed level was calculated based on 26 survey points is 0.24 m, while two survey points resulted in an $e_{meas,i}$ of 0.6 m.

In addition, we accounted for a metamodel error e_{meta} in the likelihood function because the metamodel is just an approximation of the full-complexity numerical model. We calculated e_{meta} through a leave-one-out cross-validation (LOO-CV), in which the model is repeatedly fitted on n-1 calibration nodes. Then, we calculated the LOO-CV error for each calibration node and training point. The LOO-CV error variance per calibration node was subsequently calculated and implemented as metamodel error $e_{meta,i}$. Finally,

Table 2 Calibration parameters and their value ranges for uniform distributions U(min, max) considered in this study

Calibration parameter		Investigated range		Prior assumption
Critical shear stresses for erosion	τ_{cr} (Pa)	0.05	0.4	$U(0.05, 0.4)$
Critical shear stresses for deposition	τ_d (Pa)	0.01	0.1	$U(0.01, 0.1)$
Dry-bulk density	ρ_b (kgm^{-3})	200.0	500.0	$U(200, 500)$
Diameter multiplier	γ (-)	0.8 • d	1.7 • d	$U(0.8, 1.7)$

the total error e_i included in the likelihood function is composed of the calibration node-specific measurement error $e_{meas,i}$ (6) and the metamodel error $e_{meta,i}$:

$$e_i^2 = e_{meas,i}^2 + e_{meta,i}^2 \tag{7}$$

If the total errors e_i (Eq. (7)) were significant, the influence of the difference between the measured and modeled bed level on the likelihood score decreased (Eq. (5)).

Metamodel construction

Equation (4) can be approximated through Monte Carlo sampling, which requires thousands of numerical model evaluations. However, models that simulate hydrodynamic and morphodynamic processes may require a long computing time, making it computationally impractical to perform thousands of trials. To circumvent unacceptably long computing time, we employed a surrogate-assisted (referring to the metamodel being a surrogate for a full-complexity model, Oladyshkin et al. 2020) Bayesian inversion technique, which replaces the full-complexity numerical model with a metamodel. In particular, a metamodel emulates the output trends of a complex model but requires orders of magnitude less computing time (Beckers et al. 2020; An et al. 2022). Here, we used a Gaussian process emulator (GPE) as metamodel, which is discussed in more detail by Rasmussen and Williams (2006). As the GPE requires the definition of a kernel, we used a radial basis (i.e., squared exponential covariance) function (RBF) kernel in this study. The RBF needs the definition of length scales and their boundaries. The resulting GPE metamodel can then be trained with numerical model responses resulting from various combinations of possible calibration parameter values. Thus, the GPE metamodel was fitted toward a multidimensional response surface where the number of dimensions corresponds to the number of calibration parameters. Note that the metamodel cannot generally replace the numerical model and only serves the purpose of accelerating model calibration.

Bayesian active learning through metamodel training

In this study, we used the GPE metamodel to approximate the prior $p(\omega)$ through 10^6 random Monte Carlo samples. The quality of the surrogate-assisted Bayesian calibration depends on the ability of the metamodel to replicate the full-complexity model. The more training points used to train the metamodel, the better the predictions, since more information is provided to the metamodel with fewer gaps in the parameter space (i.e., fewer gaps need to be closed through stochastic interpolation). However, filling the entire parameter space with training points with a computationally expensive full-complexity model is practically

not feasible because it requires several hours to compute one training point (sums up to more than 500 years of computing time in our case). To bypass long computing time, we applied BAL, which identifies optimal regions in the parameter space for calibrating parameters as a function of metamodel responses. BAL iteratively improves the metamodel in those regions of the parameter space that are most important for Bayesian inference (Oladyshkin and Nowak 2019; Oladyshkin et al. 2020).

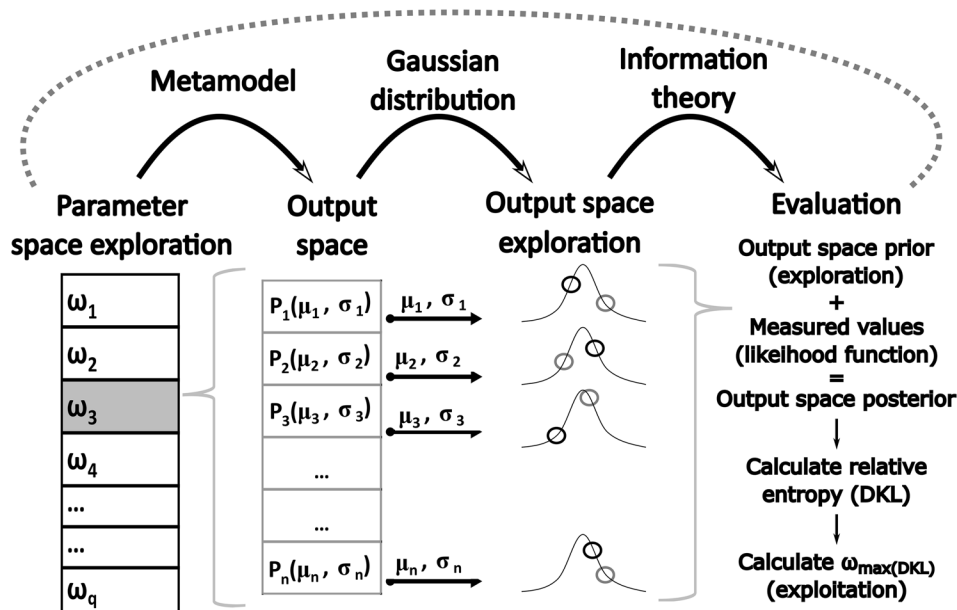
Before starting the BAL process, a prior probability distribution $p(\omega)$ was assigned to every calibration parameter. Initially, a uniform probability distribution between two limit values was assumed. The next step is to compute an initial metamodel, using m parameter realizations and the corresponding full-complexity numerical model runs to train the metamodel. BAL starts with iteratively updating the initial metamodel with new training points so that the metamodel predictions better represent the full-complexity model. For this purpose, we sampled q parameter realizations ω_i that compete to be the next training point (exploration).

The parameter realizations constitute the *parameter space* and each combination is evaluated in the metamodel to generate an *output space*. Here, we had n outputs, associated with the location of our calibration nodes. An advantage of Gaussian processes for generating the metamodel is that each prediction of the output space consists of a mean μ_n and a standard deviation σ_n . Therefore, one can explore the *output space* using a multivariate Gaussian distribution. Figure 3 shows the BAL workflow and exemplary features two random exploration samples (black and gray circles), which in our study, are not just two but 10^5 random exploration samples forming the *output space prior*. To yield the *output space posterior* distribution, we considered two options, notably rejection sampling and Bayesian reweighting. Due to the high dimensionality of the output space (142 calibration points), the rejection rate for the first case was too high, and we chose Bayesian reweighting. For this purpose, we renormalized each value of the prior’s likelihood by their total sum to generate the posterior distribution. Consequently, all realizations (i.e., Monte Carlo samples) of the prior contributed to the posterior statistics (i.e., length scales), proportional to their likelihood. Once the prior $p(\omega)$ and posterior $p(\omega|z_{meas})$ distributions had been generated, we used Eq. (8) to evaluate the so-called relative entropy $D_{KL}(p(\omega|z_{meas}), p(\omega))$ (also referred to as Kullback–Leibler divergence) between both distributions (Kullback and Leibler 1951; Oladyshkin et al. 2020).

$$D_{KL}(p(\omega|z_{meas}), p(\omega)) = -\ln[BME] + E_{p(\omega|z_{meas})}(\ln[p(z_{meas}|\omega)]) \tag{8}$$

where $E_{p(\omega|z_{meas})}$ is the average of the posterior sample’s likelihood (through the likelihood function, cf.

Fig. 3 Flow diagram explaining the Bayesian active learning method applied in this study



Equation (5)). In this context, the relative entropy expresses the information gain from the prior to the posterior distribution.

To this end, every BAL iteration involves the calculation of $D_{KL}(p(\omega|z_{meas}), p(\omega))$ for q samples from the *parameter space*. After evaluating $D_{KL}(p(\omega|z_{meas}), p(\omega))$, we calculated the parameter combination $\omega_{max(DKL)}$ that produced the maximum value of relative entropy to select the stochastically best-performing values for the calibration parameters in this iteration step (exploitation). We used the set of calibration parameters with the highest relative entropy to re-run the numerical full-complexity model and prepared the next BAL iteration step. In particular, the results of the new full-complexity model run serve as new training points for the GPE at the beginning of the next BAL iteration step. The BAL iterations continue until a stop criterion is reached, which is typically the convergence of relative entropy and BME (Oladyshkin et al. 2020). In this study, we additionally considered the evolution of the root-mean-square error (RMSE) after every BAL iteration. The BAL workflow (Fig. 3) and creation of the initial metamodel are explained in detail in the supplemental material SI 2. The complete procedure is implemented in a Python code (Acuna Espinoza et al. 2022).

Selection of calibration nodes for model calibration

The numerical model of the Banja reservoir was calibrated toward measured bed levels at the end of the three-year simulation period from 2016 to 2019. However, we could not use the totality of the available 632×10^3 bed level measurements because we needed to meet two criteria. First, the measurements needed to comply with the computational

mesh and we agglomerated multiple measurements into one at the calibration nodes of the mesh. Second, the number of BAL iterations depends on the number of calibration nodes, and a large number of nodes can result in the so-called *curse of dimensionality* (Bellman 1957), which we will discuss later in light of the results. For instance, if we used 3500 measurement points, the multivariate Gaussian density for calculating the prior output space would have 3500 dimensions of spatially explicit bed level change.

Therefore, we only used nodes located at a maximum distance of 1.5 m from a measured point for calibration, and we agglomerated all measurements in a 3-m radius at the resulting calibration nodes into one bed level value. Further, we did not consider measurements in the downstream section of the reservoir, as we are only interested in the upstream area, where most sediments deposit. These selection filters resulted in 142 calibration nodes at which we evaluated modeled bed levels in the calibration process (see also Fig. 1). For testing the hypothesis (i) that Bayesian calibration only converges toward physically meaningful model parameter combinations when the model is well-conditioned, we introduced two scenarios of measurement data available for the calibration process. First, we considered all 142 calibration nodes that define the MAXME (MAXimum MEasurements) data scenario (black and red calibration nodes in Fig. 1). Second, we removed points in regions where deltaic avulsion and channel erosion occurred according to the observation from 2016 to 2019 to define a VALDOME (VALID DOME MEasurements) data scenario, where all calibration nodes are in the domain of validity of the numerical model. In particular, we removed points adjacent to dry areas (tidal flats) and all measurements where the model uncertainty from the MAXME data scenario was high, as indicated by

LOO-CV error greater than 5.5 m based on an expert assessment. These two removal criteria essentially excluded model regions where avulsion and channel erosion occurred at the head of the reservoir, which the full-complexity model will not be able to simulate correctly. The application of these removal criteria left 109 calibration nodes that we used for the VALDOME scenario (black calibration nodes in Fig. 1).

Experimental procedure

Bayesian calibration stability

The proposed optimization of the Bayesian calibration scheme refers to the extension of the BAL framework, notably the adaptive implementation of errors in the likelihood function through LOO-CV, and its application to four calibration parameters. With these two novel aspects of BAL, we investigated the robustness of Bayesian calibration regarding the quality of the numerical model and in light of equifinality. Therefore, we applied Bayesian calibration to the two above-introduced data scenarios (MAXME and VALDOME).

To prepare the BAL iterations, we ran the full-complexity model with 15 calibration parameter combinations to train the initial metamodel. 13 of the parameter combinations stemmed from random sampling in the parameter space, and the remaining two corresponded to theoretically maximum and minimum sedimentation (i.e., high/low τ_{cr} , low/high τ_d , low/high ρ_b , high/low γ , respectively). A minimum of one training point would be sufficient for the initial metamodel, but more initial training points for BAL can reduce the total time required to achieve convergence. We tracked BME, and RMSE to evaluate if the calibration reached convergence regarding uncertainty and error (see the above section on Bayesian active learning). However, convergence may not be achieved if multiple high-probability regions cause exploitation to jump between very different calibration parameter combinations in the BAL iterations. In these cases, BAL theoretically bounces back and forth eternally between nearly equally likely combinations of calibration parameters. This phenomenon, known as equifinality, poses a great challenge for model calibration (e.g., Franks et al. 1997).

To address equifinality, we analyzed the BAL convergence in the two measurement data scenarios (see above) with 55 iterations according to literature recommendations (Mohammadi et al. 2018; Beckers et al. 2020; Scheurer et al. 2021). We verified hypothesis (i) if the VALDOME scenario led to more unique and physically meaningful maximum likelihood regions than the MAXME data scenario, and less significant, later, or no convergence in the MAXME data scenario. To this end, we investigated the evolution of BME, RMSE, and the variability of the four calibration parameters in the last five BAL iteration steps. To assess

the ability of the metamodel to reproduce the results of the full-complexity model, we compared the results predicted by the metamodel with those predicted by the numerical model. We present the global model accuracy after the Bayesian calibration by comparing the calculated and measured bed level changes after running the two data scenarios.

Importance of calibration parameters

The Bayesian calibration looks for the best-fit combination of the four calibration parameters τ_{cr} , τ_d , ρ_b , and γ to investigate optimization methods for multidimensional calibration parameter spaces and find the most relevant parameters driving reservoir sedimentation in this numerical model. The four calibration parameters are known to be relevant for hydro-morphodynamic processes in reservoirs (Haun et al. 2013; Dutta and Sen 2016; Hillebrand et al. 2016). However, to our best knowledge, the four parameters have never been directly compared with each other due to the limited capacities of subjective trial-and-error calibration. The adapted Bayesian framework and the VALDOME scenario enable us to perform such a comparison of the four calibration parameters. Thus, we aim to test hypothesis (ii) that at least one of the calibration parameters τ_{cr} , τ_d , ρ_b , or γ plays a governing role in the fluvial deposition of suspended sediments in reservoirs. To this end, we made use of a multi-parameter plot of the posterior distributions (Eq. (5)) of the four calibration parameters for both data scenarios. We will accept hypothesis (ii) if at least one of the four calibration parameters has a considerably narrower posterior distribution than the other parameters. This parameter will be more important than the other calibration parameters because it has the smallest uncertainty (i.e., narrowest posterior) of the maximum likelihoods.

Results

Convergence speed

In the VALDOME scenario, the BME began converging toward a value of approximately 10^{-31} after the 46th BAL iteration, and we ran in total 55 iterations to monitor the convergence trend (see Fig. 4). The BME for the MAXME scenario fluctuated around a value of 10^{-37} during the 55 BAL iterations, and no clear convergence trend was observed. In addition, Fig. 4 also shows the evolution of the RMSE between the metamodel and full-complexity model results for every tested parameter combination used as a training point in BAL. The plots reveal that the RMSE is higher for the MAXME scenario, and more importantly, there is no decreasing trend for this scenario. In contrast, the evolution of the RMSE for the VALDOME scenario decreased.

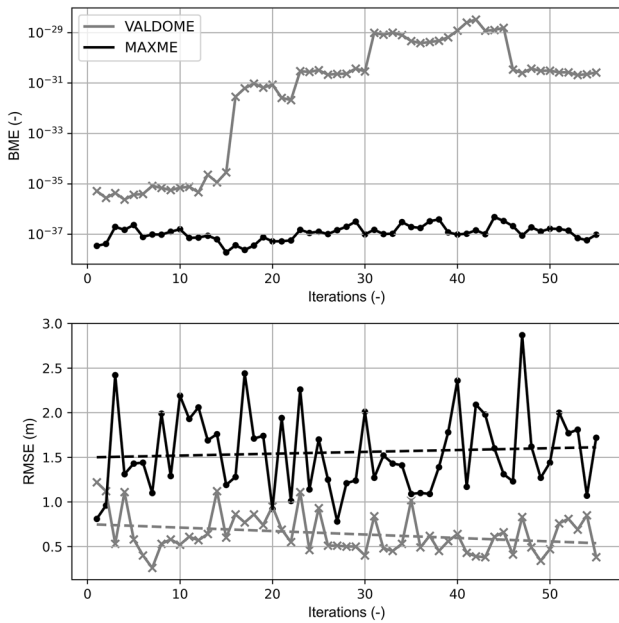


Fig. 4 BME (top) and RMSE evolution including linear trend lines (bottom) for the 55 BAL iterations and both data scenarios

Figure 5 shows the variability of the four calibration parameters in the last five BAL iterations for the VALDOME scenario in black and the MAXME scenario in gray. Comparing the two data scenarios shows that the MAXME scenario had significantly higher variability and no physical convergence for τ_{cr} and τ_d . In contrast, there was hardly any variability of ρ_b in both scenarios, whereas the variability in γ was slightly higher in the VALDOME scenario. The total computing

Fig. 5 Variability of the four calibration parameters for the five last BAL iterations 50–55 and both data scenarios

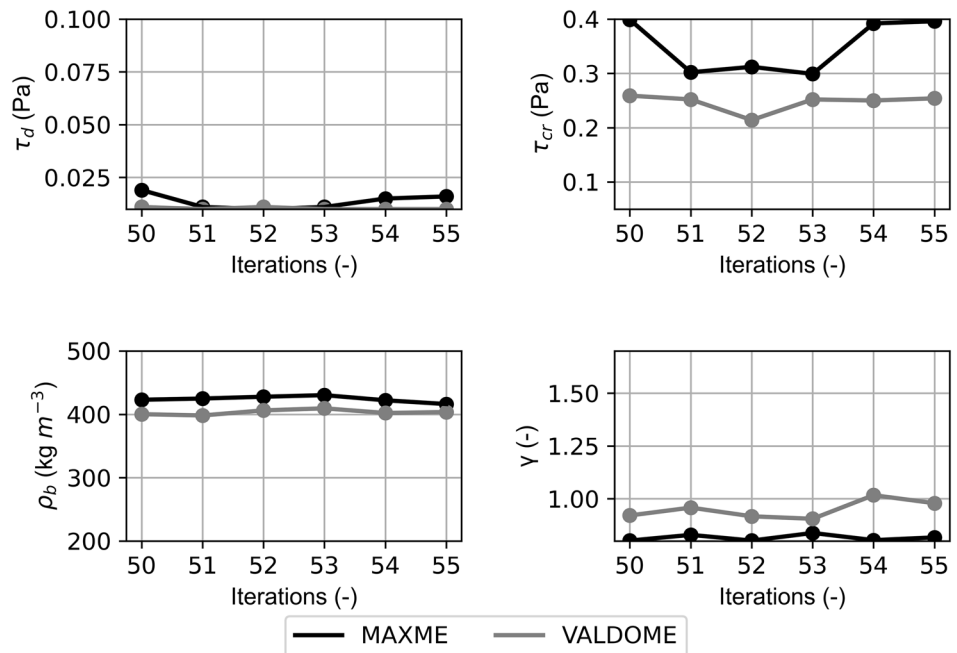


Table 3 Calibrated parameters for the morphodynamic model of the Banja reservoir using BAL

Calibration Parameter	Prior assumption	Maximum likelihoods	
		MAXME	VALDOME
τ_{cr} (Pa)	U(0.05, 0.4)	0.39	0.25
τ_d (Pa)	U(0.01, 0.1)	0.02	0.01
ρ_b ($kg\ m^{-3}$)	U(200, 500)	416.2	403.6
γ (-)	U(0.8, 1.7)	0.82	0.98

time for the BAL iterations per scenario was approximately one month (on 12 Cores using AMD Ryzen 9 5950 \times 16- (32) @ 3.4 GHz processor), which was only possible with the coarse mesh resolution.

Posterior distributions and importance of calibration parameters

Maximum likelihoods

Table 3 shows the maximum likelihood of posterior distributions for the calibration parameters, which is the realization of the Monte Carlo sample with the highest likelihood and comparable to a deterministic best-fit solution. The maximum likelihood of τ_{cr} was 0.39 Pa, close to the upper limit considering all calibration nodes (MAXME). In the physically relevant-only (VALDOME) scenario, τ_{cr} was 0.25 Pa. The critical shear stress for deposition τ_d was close to the lower limit at 0.02 Pa and 0.01 Pa in the MAXME and VALDOME scenarios, respectively. The maximum likelihood of

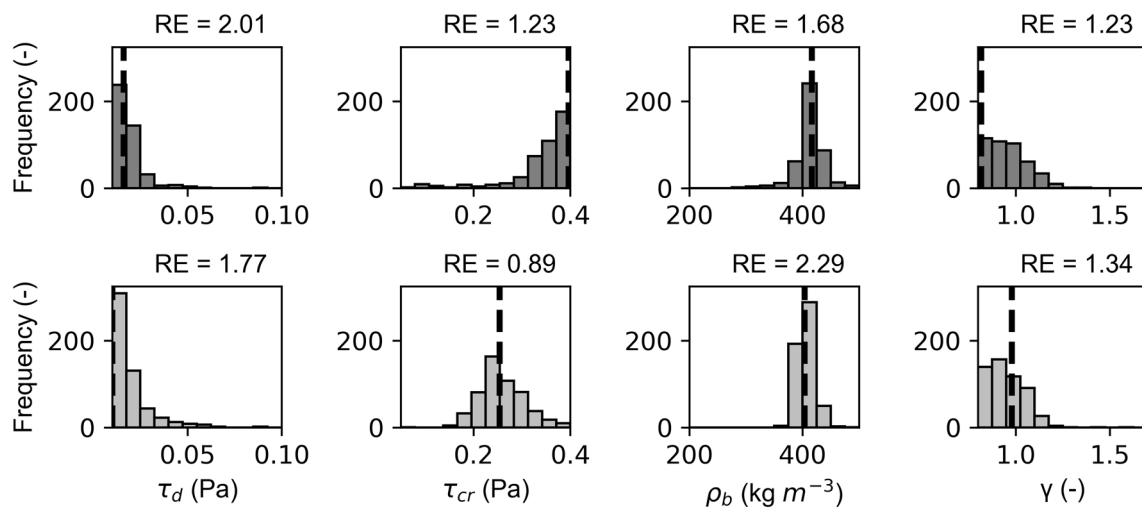


Fig. 6 Posterior distributions in the parameter space, and associated relative entropy (RE) at the end of the MAXME scenario in gray (top) and the VALDOME scenario in light gray (bottom). The dashed vertical lines indicate the maximum likelihood values of the calibration parameters

ρ_b was close to 410 kg m^{-3} in both scenarios. The diameter multiplier γ was 0.82 in the MAXME scenario and 0.98 in the VALDOME scenario. A detailed analysis of the posterior parameter space is provided below.

Posterior parameter distributions

The posterior distributions of the calibration parameters indicate the uncertainty in the maximum likelihoods listed in Table 3. Figure 6 shows the individual posterior histograms for each calibration parameter after the MAXME scenario at the top (433 posterior samples) and after the VALDOME scenario at the bottom (540 posterior samples).

As a standardized measure to identify driving calibration parameters and evaluate their uncertainty, we calculated the Kullback–Leibler divergence (Kullback and Leibler 1951), also known as relative entropy (RE), to measure the information gain between the initial (uniform) prior and the final posterior probability distribution for every calibration parameter. High RE characterizes a narrow distribution, which represents high information gain and low uncertainty in the maximum likelihoods.

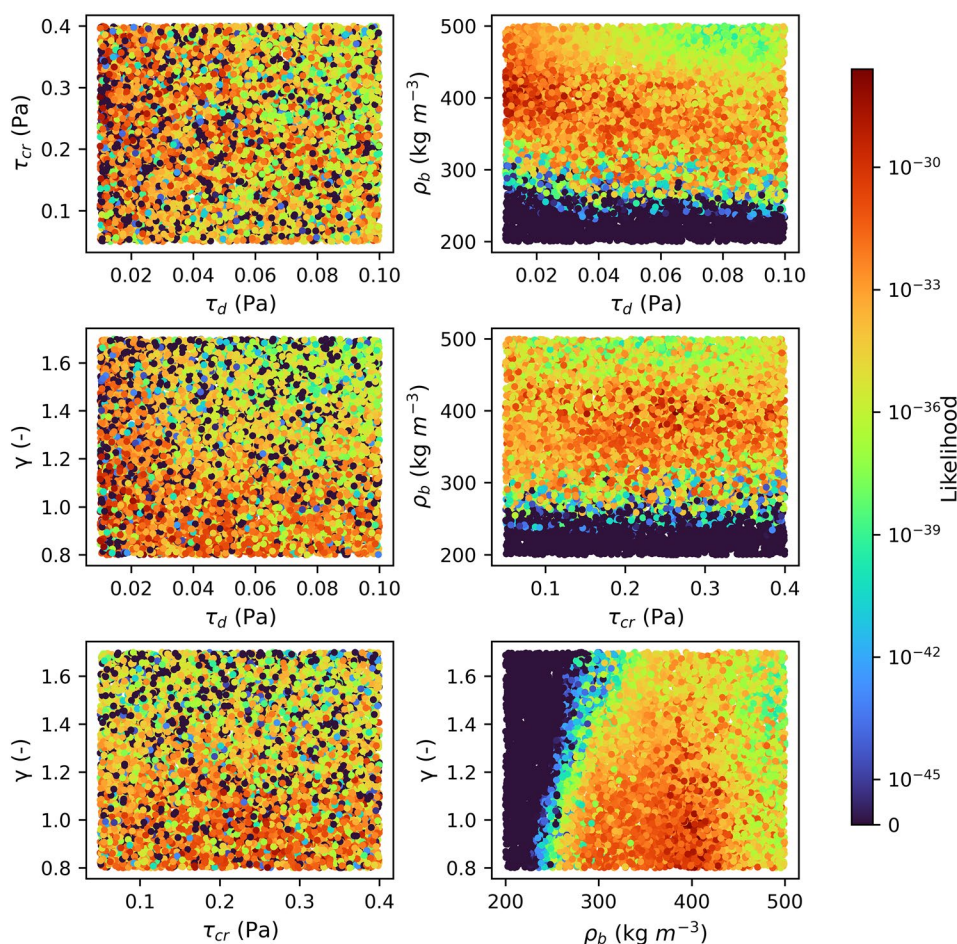
Figure 6 shows that ρ_b and τ_d have the narrowest posterior distribution in both calibration scenarios, which indicates that these parameters are the most restrictive and important in the calibration process. This finding is also supported by the high RE of 1.68 and 2.29 for ρ_b and 2.01 and 1.77 for τ_d . That is, only values close to the maximum likelihood value (dashed line) led to accurate results. However, the maximum likelihood for τ_d was close to the lower limit in both data scenarios. The histogram for τ_{cr} differs significantly between the two different scenarios. For the MAXME scenario, the distribution peaks close to the upper limit, and the RE was

1.23 whereas the histogram for the VALDOME scenario peaks at 0.25 Pa and gets wider, characterized by a lower RE of 0.89. The histogram of the diameter multiplier γ peaks in both scenarios close to the lower limit of the initial range and the RE slightly increased from 1.23 to 1.34 for the VALDOME scenario.

The meaningfulness and qualitative significance of the yielded maximum likelihoods can also be interpreted by examining data patterns and regions of high and distinguishable maximum likelihoods in the parameter space. For this purpose, Fig. 7 illustrates the likelihood of all possible calibration parameter combinations of τ_d , τ_{cr} , γ , and ρ_b at the end of the VALDOME scenario. Similar plots of the results for the MAXME scenario can be found in SI Fig. 2. Since a four-dimensional parameter space cannot be plotted graphically, we created six two-dimensional plots of the possible combinations. The three plots on the right of Fig. 7 clearly show that the likelihoods for $\rho_b < 300 \text{ kg m}^{-3}$ are very small and quite small for $\rho_b > 450 \text{ kg m}^{-3}$ (in line with Fig. 6). Thus, values of ρ_b significantly lower or higher than the maximum likelihood did not lead to accurate results, and the data pattern of ρ_b confirms its high relative importance compared to the other three calibration parameters. In contrast, the boundary between high and low probabilities in the data pattern for τ_d was less distinct, with the highest likelihoods occurring for $\tau_d < 0.05 \text{ Pa}$. τ_{cr} had the least pronounced data pattern, and high probabilities occurred almost throughout the entire range. In addition, the data pattern for γ showed high likelihoods over a wide range with $\gamma < 1.2$.

Furthermore, there was no significant correlation between the calibration parameters (see SI Fig. 3), which indicates that the calibration parameters were well chosen and independent. If there were high correlations between

Fig. 7 Likelihood values along the six possible parameter space combinations at the end of the VALDOME scenario



the calibration parameters, they would contain redundant information and the variation in one parameter could be compensated by a change of another parameter. In such a case, calibration would be restricted to determining the ratios between the parameters.

Simulated bed level changes

Figure 8 shows the cumulative bed level changes and water depth in the Banja reservoir after the end of the three-year simulation period with the calibration parameters for the VALDOME scenario listed in Table 3. In the region near the Devoll River tributary, the water became shallow and several channels formed. The highest deposits of more than 4 m occurred in the upstream part of the reservoir. At low water levels, some of the deposited sediment in the upstream part of the reservoir was eroded, resulting in the formation of smaller channels. These channels did not occur in permanently impounded regions (shown in dark blue). The sediment deposit height decreased in flow direction because of the decreasing flow velocity and the continuous settling of sediment particles. In the reservoir, the deposition heights

in flow direction were less than 4 m after 2.4 km, less than 1.5 m after 4.6 km, and less than 0.5 m after 8.0 km.

Model accuracy

Agreement between GPE metamodel and numerical model

Since the final calibration parameters stem from the metamodel, we performed two analyses to evaluate the calibration quality. First, we compared the metamodel with the 2d hydro-morphodynamic model to quantify how well the metamodel mimics the full-complexity model results. Second, we compared the calibrated numerical model with the measurement data to evaluate the final model quality.

To evaluate the quality of the metamodel results, the numerical model was run with the optimal calibration parameter combinations shown in Table 3. The inclusion of all calibration nodes (MAXME) led to a Pearson's correlation r of 0.92, an RMSE of 0.87 m, and a mean absolute error (MAE) of 0.45 m (Fig. 9). Hence, the metamodel reproduced the numerical model results with good accuracy. However, the metamodel significantly overestimated bed level changes for some calibration nodes in the upstream part near the

Fig. 8 Cumulative bed level changes (left) and water depth (right) after the end of the simulation period of the VALDOME scenario

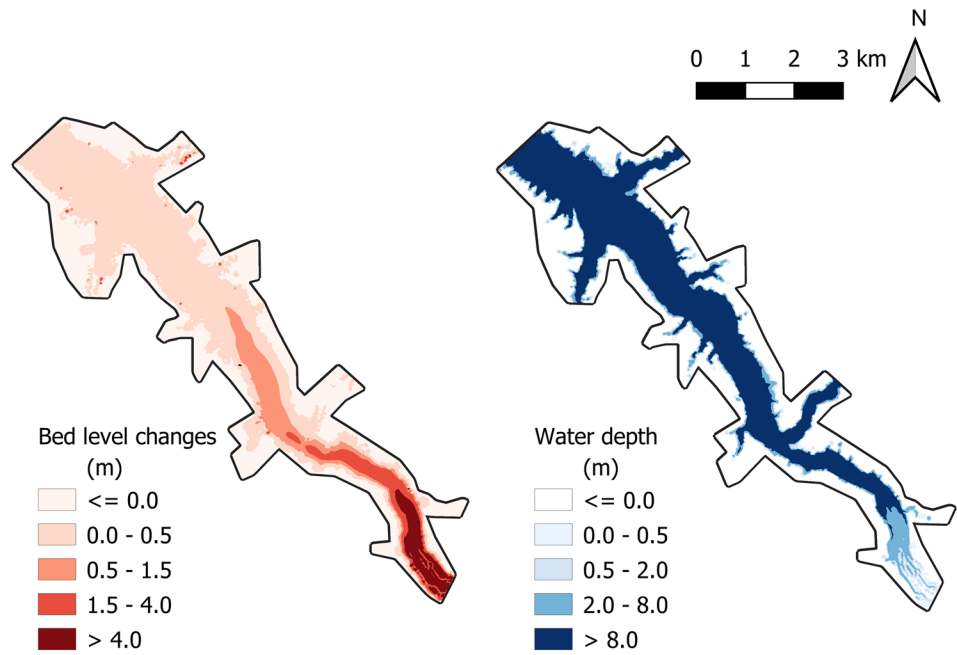
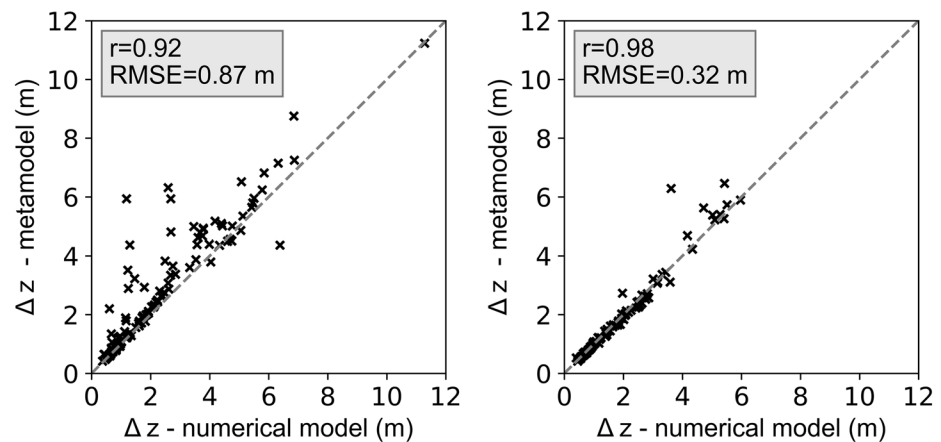


Fig. 9 Scatter plot of the computed bed level changes Δz from the metamodel and the numerical model after the MAXME scenario at the left and VALDOME scenario at the right. The dashed line represents the hypothetical perfect metamodel accuracy



Devoll River tributary. Excluding these upstream nodes (VALDOME) resulted in a significantly better r of 0.98, an RMSE of 0.32, and an MAE of only 0.13 m. Therefore, the trained metamodel accurately emulated the full-complexity numerical model results at the end of the simulation period.

To assess the regions of high uncertainty in the metamodel, we calculated the LOO-CV errors for the 142 calibration nodes and the 70 parameter combinations used as training points. We averaged the absolute values of the differences for each point to estimate the expected error between the metamodel and the full-complexity model. This analysis is important because a high model metamodel error causes decreased influence of the difference between the measured and modeled bed level on the likelihood (Eq. 5). Therefore, calibration nodes with a large LOO-CV error indicate high uncertainty in the metamodel and carry

less weight in the final likelihood calculation. The LOO-CV mean errors for the MAXME scenario ranged from 0.04 to 4.04 m (0.8 m on average). The highest errors occurred near the upstream boundary, which is also shown in SI Fig. 4. In contrast, the LOO-CV errors for the VALDOME were significantly smaller and ranged from 0.03 to 1.76 m (0.35 m on average).

Modeled and measured agreement

To evaluate the quality of the calibrated hydro-morphodynamic numerical model, we calculated Pearson’s r and the RMSE for both scenarios. The MAXME scenario led to an r of 0.70, RMSE of 1.62 m, and MAE of 1.17 m. The VALDOME scenario yielded a similar r of 0.66, a smaller RMSE of 1.04, and a smaller MAE of 0.91 m.

Fig. 10 Scatter plot of modeled and simulated bed level changes Δz for the MAXME scenario at the left and VALDOME scenario at the right. The dashed line represents the hypothetical perfect model accuracy

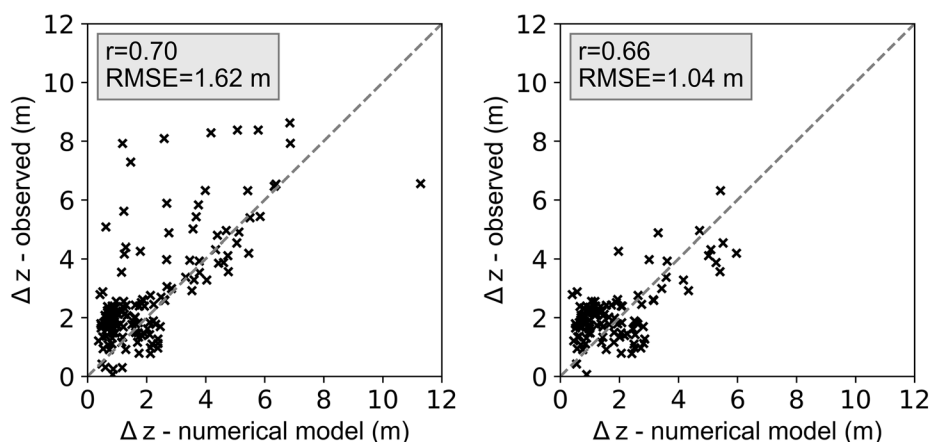
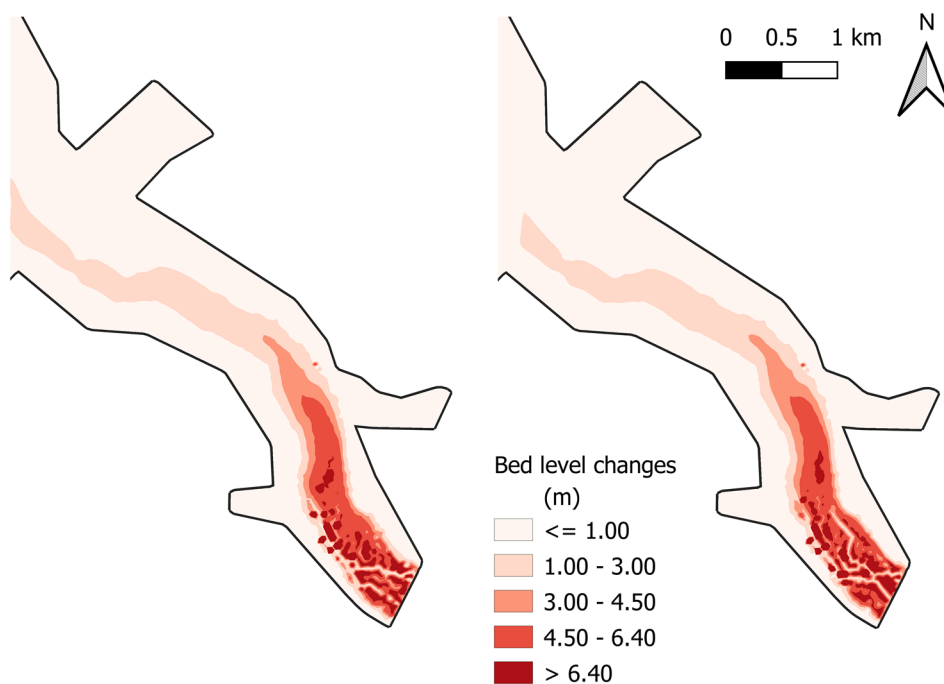


Fig. 11 Simulated bed level changes (2016–2019) in the upstream part of the reservoir for two different simulations with similar calibration parameters at the end of the simulation period



Thus, both scenarios yielded satisfactory agreement between the measured and modeled results according to global statistics indicating a good representation of the sedimentation patterns in the Banja reservoir (see Fig. 10). Still, the MAE and RSME were not negligible but significantly lower in the VALDOME scenario. In the MAXME scenario, there were 11 calibration nodes with bed level change errors greater than 2 m compared to only three such nodes in the VALDOME scenario (see SI Fig. 5). In addition, Fig. 10 indicated that the numerical model tends to underestimate small measured bed level changes by approximately 1 m in both scenarios.

Discussion

Deposition patterns and model deviations

Figure 11 shows the results of the bed level evolution in the upstream part of the reservoir after three years and for two simulations with similar parameter combinations. The figure shows several channels with high topographic gradients at different locations. Thick sediment deposits occurred next to these channels, particularly at mesh nodes that were only temporarily wet in the simulation period. These nodes can be inside a channel (small sediment deposits) in one model run and outside the channel (thick sediment deposits) in the next model run. Although the physical model environment

is similar, the patterns in Fig. 11a and b are very different, which indicates numerical instabilities that the metamodel attempted to emulate by drastically changing the calibration parameter values (see Fig. 5).

The coarse mesh used in this study (to reduce computing time) affected the accuracy of the results and numerical stability. In addition, the model only considered suspended sediment transport and was not able to reproduce shallow water regions experiencing channel erosion, deltaic avulsion, or bank failures. The simplification assumptions made the model physically not fully well defined for simulating morphological processes at the head of the reservoir. In the MAXME scenario, the BAL iterations attempted to overcome mismatches between measured and modeled erosion channels by prominent changes in the calibration parameter values, indicating equifinality. As a result, the BAL iterations were unstable and did not converge. This is reflected in the higher fluctuations of the calibration parameters (Fig. 5) and maximum likelihoods near the limit of the investigated range for three of the four calibration parameters in the MAXME scenario. In addition, the BME (Fig. 4) did not converge because every BAL iteration tried to explore numerical instability in the delta region. Yet, the adapted Bayesian calibration worked well in the model domain where suspended sediment deposition could be reproduced and the model was stable (VALDOME). The BAL converged toward a solution that is confirmed by a very low RMSE of 0.32 m and a high Pearson's r of 0.98 between the full-complexity and the metamodel. However, the RMSE of 1.04 m of the numerical model regarding measured data was significantly smaller compared to the MAXME scenario but not negligible, which is due to the limitations of the modeling approach. For instance, complex three-dimensional hydrodynamics and stratified flow cannot be represented by a 2d model, which is expected to affect the deposition pattern in the reservoir. Also, we assumed a constant ρ_b for the entire reservoir, while consolidation occurs over time and the density increases (Mehta et al. 1989; Winterwerp and Kesteren 2004; Lo et al. 2014; Hoffmann et al. 2017). Accordingly, the average bulk density in a reservoir is often heterogeneous and varies over time, which is not reflected in our model assumptions. In addition, some boundary conditions were not measured directly, and therefore, subject to additional uncertainty. For example, the inflow into the reservoir was calculated from measured outflow, reservoir water levels, and hydrological model outputs. Also, the sediment yields of the tributaries stem from a model with monthly resolution only (Mouris et al. 2022). Considering the uncertainty related to these model simplifications, as well as the mean measurement error of approximately 0.4 m, the final model quality is acceptable in the VALDOME scenario. In light of the instability of the MAXME scenario, we verify the hypothesis (i) that Bayesian calibration only converges

toward physically meaningful model parameter combinations when the model is well-conditioned (i.e., measured data are in the validity domain of model assumptions).

Relevant calibration parameters

Since the Bayesian calibration only converges toward physically meaningful model parameter combinations when the calibration nodes are in the range of validity of the numerical model, we only used the VALDOME scenario to identify the calibration parameter importance for reservoir sedimentation modeling.

Figure 6 shows that the density ρ_b was the most restrictive (i.e., constraining) calibration parameter due to its narrow-shaped posterior distribution, which was not imposed by the initial value ranges (Table 2). This is also evident in the clear data pattern of ρ_b across the parameter space (Fig. 7), where high likelihoods occurred only in a very narrow range. Yet, many studies exclude ρ_b from the calibration process and use fixed literature values or empirical equations to obtain a representative value (Foster and Charlesworth 1994; Verstraeten and Poesen 2001; Banasik et al. 2021). Our findings suggest that ρ_b should be either calibrated or directly measured, rather than simply derived from the literature. This finding is important because, for instance, models for calculating the sediment yield are often calibrated against the volume change in lakes or reservoirs. Since the volume of the deposited sediments is directly related to the dry-bulk density of the sediments, an incorrect value for ρ_b results in an incorrect calculation of sediment masses. For example, if the sediment inflow is underestimated, the error can be compensated for by a lower dry-bulk density for the deposited sediment. In this study, the Bayesian calibration led to a reasonable value of 403.6 kg m⁻³ in the VALDOME scenario (van Rijn and Barth 2019; van Rijn 2020).

According to the posterior distribution (see Fig. 6), τ_d was the second most restrictive (i.e., constraining) parameter with a small maximum likelihood of 0.01 Pa. Yet, the maximum likelihood was located at the lower limit of the investigated range, which suggests that the Bayesian calibration would have tried an even smaller τ_d if possible. Hence, the posterior distribution should be interpreted carefully, as a broader range may result in a wider distribution. A possible explanation for why the Bayesian calibration preferred small values of τ_d is the maximization of suspended load trajectories. Since fine particles are kept in suspension by turbulence even at low flow velocities, the BAL attempted to compensate for the insufficient model assumptions regarding 3d turbulence (mixing length model) by decreasing τ_d . Furthermore, the actual shear stresses in a large reservoir are very small. Thus, only very small τ_d values affect the deposition process in the numerical model, especially since

we disregarded measured data in the shallow delta region in the VALDOME scenario.

The calibrated diameter multiplier γ was 0.98, which falls into the lower half of the initial range. γ was not very restrictive and yielded high likelihoods for a comparatively wide range with $\gamma < 1.2$. These small γ values indicate that the observed particle size diameters and the corresponding settling velocities were rather too large and smaller particle sizes and settling velocities lead to better results. Generally, the grain size distribution of a suspended sediment load sample represents the present hydraulic conditions. Therefore, we recommend using grain sizes in a reasonable range for calibration or varying the grain size distributions as a function of discharge. The stochastic approach led to representative grain size ranges: $d_{17} = 1.89$ to 2.25 μm , $d_{50} = 7.0$ to 8.33 μm , and $d_{83} = 19.76$ to 23.62 μm . However, flocculation processes can alter the settling velocity of cohesive particles (Dyer and Manning 1999; Winterwerp and Kesteren 2004), and therefore, the actual grain sizes of individual particles can be even smaller.

The optimum τ_{cr} was 0.25 (VALDOME) and the corresponding likelihood pattern was not very pronounced and had little impact on the calibration process, because only the very upstream calibration nodes were affected by erosion and resuspension. In contrast to a free-flowing river, sedimentation dominates in the reservoir due to the large water depths and low flow velocities. Thus, there were significant differences between the two sets of calibration nodes, underlining that particularly the upstream delta section of the reservoir was controlled by τ_{cr} . Hence, the RE further decreased with the exclusion of the calibration nodes in the upstream part (VALDOME), which emphasizes the diminishing importance and higher uncertainty of τ_{cr} .

Ultimately, we verify hypothesis (ii) since the Bayesian calibration identified ρ_b as the driving calibration parameter in the fluvial deposition of suspended sediments in reservoirs. In contrast, γ or τ_{cr} had significantly less influence on the final sedimentation pattern. Parameters with narrow posterior distributions and high relative entropy compared to a uniform distribution can be interpreted as driving and restrictive, while parameters with wide posterior distributions can be interpreted as less important and uncertain. The narrow posterior distribution for τ_d suggests a high information gain through BAL, with the maximum likelihood at the lower limit. Consequently, small τ_d lead to more accurate results, although the importance of τ_d cannot be objectively assessed.

The curse of dimensionality

The so-called curse of dimensionality (see also the methods section on Bayesian calibration) forced us to limit the number of calibration nodes. Even though we limited the number

of calibration nodes, the dimensions of the response surface were still too high and both scenarios were subjected to the curse of dimensionality. This phenomenon occurred because of the exponential term of the likelihood function (Eq. (5)), which represents the (negative) weighted sum of the squared difference between the measured and modeled bed level change. The more calibration nodes we used, the larger the negative value of the sum becomes. In consequence, the exponential term became a number so close to zero that the precision of a computer is insufficient to express it. This problem, known as arithmetic underflow (e.g., Coonen 1980), caused the likelihood function to become zero, which does not allow for the calculation of convergence scores and selection of a next training point. To solve this problem, we artificially increased the total error in Eq. (7) by multiplying it by 5. The artificial error amplification was equally applied to all individual errors and represented the smallest integer amplification factor that avoided arithmetic underflow. Since the amplification factor was constant, the rank of the output realization remained unchanged.

The curse of dimensionality also affected the number of Monte Carlo (MC) samples that could be drawn to approximate the posterior distribution in Eq. (5). With increasing dimensionality, the required computing power for a representative sample increased exponentially. In consequence, the region with the highest density became more restrictive and the vast majority of the probability density function was concentrated in low-likelihood areas. To balance representativeness, the curse of dimensionality, and computing time, we limited the sample size to 10^5 MC realizations.

The curse of dimensionality also affects the generation of the posterior distribution through rejection sampling (Smith and Gelfand 1992) or the here-used Bayesian weighting strategy, as most of the samples were concentrated in low-likelihood areas. Thus, the weight of nearby all samples was close to zero or arithmetic underflow occurred. The above-introduced error multiplier helped to avoid these arithmetic underflow issues by increasing the width of the high-likelihood region and enabling a representative posterior.

Conclusion

The region where the model simplifications were not entirely valid caused stability issues in the upstream part of the reservoir, where small channels with low water depths led to high topographic gradients and large model uncertainty. Hence, the inclusion of all calibration nodes resulted in a degradation of model accuracy, fluctuating Bayesian model evidence, and higher variability of the four calibration parameters in the last five BAL iterations. In addition, the maximum likelihood values of the calibration parameters were located near the limit of the

investigated range. Consequently, Bayesian calibration only converged toward physically meaningful parameter combinations when the model was well-conditioned (i.e., when the measurement data are in physically representative regions of the model domain). The final model quality was still affected by the limitations of the 2d numerical model, leading to a considerable mean absolute error of approx. 1 m regarding the modeled deposition height.

Bayesian calibration identified the dry-bulk density as the driving and most important parameter to simulate the fluvial deposition of suspended sediments in reservoirs. Thus, the dry-bulk density should be prioritized in data collection, already before setting up a reservoir sedimentation model. In contrast, the particle diameter multiplier and the critical shear stress for erosion had less influence on the deposition pattern as can be seen from the wider posterior distribution. The importance of the critical shear stress for deposition could not be objectively assessed because the maximum likelihood is located at the lower limit of the initial range. Yet, small values led to better results because the BAL tried to maximize suspended load trajectories to compensate for insufficient model assumptions about 3d turbulence that keeps fine particles in suspension.

Ultimately, this study shows that a robust Bayesian calibration can also be achieved when global model simplification hypotheses cannot be applied to the entire model domain, requiring that the measurement data for calibration must be from model domains where the simplifying assumptions are valid. Furthermore, our modified BAL approach accounted for both measurement and metamodel errors, enabling a multi-parametric comparison and identification of driving calibration parameters even in four-dimensional parameter space.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40808-023-01712-7>.

Acknowledgements This study was carried out in the framework of the DIRT-X project, which is part of AXIS, an ERA-NET initiated by JPI Climate, and funded by FFG Austria, BMBF Germany, FORMAS Sweden, NWO NL, and RCN Norway with co-funding from the European Union (Grant No. 776608). The fourth author is funded by the German Research Foundation (DFG, 327154368 – SFB 1313). The fifth author is indebted to the Baden-Württemberg Stiftung for the financial support from the Eliteprogramme for Postdocs. We also thank Nils Rütther and Slaven Conevski for providing us with input data and Maria Fernanda Morales Oreamuno and Anna Cerf for their help and fruitful discussions.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability All the codes and steering files of the numerical model are publicly available in an online repository. Further datasets used in this study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- An Y, Yan X, Lu W et al (2022) An improved bayesian approach linked to a surrogate model for identifying groundwater pollution sources. *Hydrogeol J* 30:601–616. <https://doi.org/10.1007/s10040-021-02411-2>
- Ardiclioglu M, Kocileri G, Kuriqi A (2011) Assessment of Sediment Transport in the Devolli River. In: 1st International Balkans Conference on Challenges of Civil Engineering
- Audouin Y, Benson T, Delinares M et al (2020) Introducing GAIA, the brand new sediment transport module of the TELEMAT. <https://doi.org/10.5281/ZENODO.3611600>. -MASCARET system
- Audouin Y, Tassi P (2020) GAIA User Manual
- Acuna Espinoza E, Mouris K, Schwindt S, Mohammadi F (2022) Surrogate Assisted Bayesian Calibration. Version 0.1.0. https://github.com/eduardoAcunaEspinoza/surrogated_assisted_bayesian_calibration/tree/v0.1.0
- Banasik K, Hejduk L, Krajewski A, Wasilewicz M (2021) The intensity of siltation of a small reservoir in Poland and its relationship to environmental changes. *CATENA* 204:105436. <https://doi.org/10.1016/j.catena.2021.105436>
- Beckers F, Heredia A, Noack M et al (2020) Bayesian calibration and validation of a large-scale and Time-Demanding Sediment Transport Model. *Water Resour Res* 56. <https://doi.org/10.1029/2019WR026966>
- Bellman R (1957) *Dynamic programming*. Princeton Univ. Pr, Princeton, NJ
- Borrelli P, Robinson DA, Panagos P et al (2020) Land use and climate change impacts on global soil erosion by water (2015–2070). *Proceedings of the National Academy of Sciences* 117:21994–22001. <https://doi.org/10.1073/pnas.2001403117>
- Box GEP, Tiao GC (1992) *Bayesian Inference in Statistical Analysis*, 1st edition. Wiley-Interscience, New York
- Chadwick AJ, Lamb MP, Moodie AJ et al (2019) Origin of a preferential avulsion node on Lowland River Deltas. *Geophys Res Lett* 46:4267–4277. <https://doi.org/10.1029/2019GL082491>
- Coonen (1980) Special feature an implementation guide to a proposed Standard for floating-point arithmetic. *Computer* 13:68–79. <https://doi.org/10.1109/MC.1980.1653344>
- Dorfmann C, Zenz G (2016) The depth-averaged Mixing Length turbulence model for Telemac-2D. *Proceedings of the XXIIIrd TELEMAT-MASCARET User Conference 2016*, 11 to 13 October 2016, Paris, France 163–168
- Dutta S, Sen D (2016) Sediment distribution and its impacts on Hirakud Reservoir (India) storage capacity. *Lakes Reserv: Res Manag* 21:245–263. <https://doi.org/10.1111/lre.12144>
- Dyer KR, Manning AJ (1999) Observation of the size, settling velocity and effective density of flocs, and their fractal dimensions. *J Sea Res* 41:87–95. [https://doi.org/10.1016/S1385-1101\(98\)00036-7](https://doi.org/10.1016/S1385-1101(98)00036-7)

- Foster IDL, Charlesworth SM (1994) Variability in the physical, chemical and magnetic properties of reservoir sediments; implications for sediment source tracing. IAHS Publ no 224:153–160
- Franks SW, Beven KJ, Quinn PF, Wright IR (1997) On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. *Agric For Meteorol* 86:63–75. [https://doi.org/10.1016/S0168-1923\(96\)02421-5](https://doi.org/10.1016/S0168-1923(96)02421-5)
- Hajek EA, Wolinsky MA (2012) Simplified process modeling of river avulsion and alluvial architecture: connecting models and field data. *Sediment Geol* 257–260:1–30. <https://doi.org/10.1016/j.sedgeo.2011.09.005>
- Hanmaiahgari PR, Gompa NR, Pal D, Pu JH (2018) Numerical modeling of the Sakuma dam reservoir sedimentation. *Nat Hazards* 91:1075–1096. <https://doi.org/10.1007/s11069-018-3168-4>
- Haun S, Kjærås H, Løvfåll S, Olsen NRB (2013) Three-dimensional measurements and numerical modelling of suspended sediments in a hydropower reservoir. *J Hydrol* 479:180–188. <https://doi.org/10.1016/j.jhydrol.2012.11.060>
- Hervouet J-M (2007) *Hydrodynamics of Free Surface Flows: Modelling with the Finite Element Method*, 1. edition. Wiley, Chichester; Hoboken, NJ
- Hervouet J-M (2020) TELEMAC-2D User Manual
- Hervouet J-M, Razafindrakoto E, Villaret C (2011) Dealing with dry zones in free surface flows. A New Class of Advection Schemes
- Hillebrand G, Klassen I, Olsen NRB (2016) 3D CFD modelling of velocities and sediment transport in the Iffezheim hydropower reservoir. *Hydrol Res* 48:147–159. <https://doi.org/10.2166/nh.2016.197>
- Hinderer M, Kastowski M, Kamelger A et al (2013) River loads and modern denudation of the Alps — a review. *Earth-Sci Rev* 118:11–44. <https://doi.org/10.1016/j.earscirev.2013.01.001>
- Hoffmann T, Hillebrand G, Noack M (2017) Uncertainty analysis of settling, consolidation and resuspension of cohesive sediments in the Upper Rhine. *Int J River Basin Manag* 15:401–411. <https://doi.org/10.1080/15715124.2017.1375509>
- Houwing E-J, Estuarine (1999) *Coastal and Shelf Science* 49:545–555. <https://doi.org/10.1006/ecss.1999.0518>
- Khorrani Z, Banihashemi MA (2021) Development of a non-coupled algorithm for simulating long-term sedimentation in the Zonouz dam reservoir, Iran. *J Soils Sediments* 21:545–560. <https://doi.org/10.1007/s11368-020-02714-z>
- Kim G-E, Kim J-H, Yoo S-H (2020) Assessing the environmental benefits of multi-purpose water uses of hydropower reservoirs on the Han River in South Korea. *Energy Environ* 31:1167–1180. <https://doi.org/10.1177/0958305X19882407>
- Kim Y-J, Park C-S (2016) Stepwise deterministic and stochastic calibration of an energy simulation model for an existing building. *Energy Build* 118:455–468. <https://doi.org/10.1016/j.enbuild.2016.10.009>
- Kondolf GM (1997) Hungry water: Effects of Dams and Gravel Mining on River channels. *Environ Manag* 21:533–551. <https://doi.org/10.1007/s002679900048>
- Kornman BA, Deckere EMGTD (1998) Temporal variation in sediment erodibility and suspended sediment dynamics in the Dollard estuary. *Geol Soc Spec Publ* 139:231–241. <https://doi.org/10.1144/GSL.SP.1998.139.01.19>
- Krone RB (1962) Flume studies of transport of sediment in estuarial shoaling processes. Hydraulic Engineering Laboratory and Sanitary Engineering Research Laboratory, Berkeley, CA, USA
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22:79–86. <https://doi.org/10.1214/aoms/1177729694>
- Langendoen EJ, Mendoza A, Abad JD et al (2016) Improved numerical modeling of morphodynamics of rivers with steep banks. *Adv Water Resour* 93:4–14. <https://doi.org/10.1016/j.advwatres.2015.04.002>
- Liang M, Voller VR, Paola C (2015) A reduced-complexity model for river delta formation; part 1: modeling deltas with channel dynamics. *Earth Surf Dyn* 3:67–86. <https://doi.org/10.5194/esurf-3-67-2015>
- Lo EL, Bentley SJ, Xu K (2014) Experimental study of cohesive sediment consolidation and resuspension identifies approaches for coastal restoration: Lake Lery, Louisiana. *Geo-Mar Lett* 34:499–509. <https://doi.org/10.1007/s00367-014-0381-3>
- Lumborg U (2005) Modelling the deposition, erosion, and flux of cohesive sediment through Øresund. *J Mar Syst* 56:179–193. <https://doi.org/10.1016/j.jmarsys.2004.11.003>
- Luppi L, Rinaldi M, Teruggi LB et al (2009) Monitoring and numerical modelling of riverbank erosion processes: a case study along the Cecina River (central Italy). *Earth Surf Process Landf* 34:530–546. <https://doi.org/10.1002/esp.1754>
- Mehta AJ, Hayter EJ, Parker WR et al (1989) Cohesive sediment transport. I: process description. *J Hydraul Eng* 115:1076–1093. [10.1061/\(ASCE\)0733-9429\(1989\)115:8\(1076\)](https://doi.org/10.1061/(ASCE)0733-9429(1989)115:8(1076))
- Mohammadi F, Kopmann R, Guthke A et al (2018) Bayesian selection of hydro-morphodynamic models under computational time constraints. *Adv Water Resour* 117:53–64. <https://doi.org/10.1016/j.advwatres.2018.05.007>
- Mouris K, Schwindt S, Haun S et al (2022) Introducing seasonal snow memory into the RUSLE. *J Soils Sediments*. <https://doi.org/10.1007/s11368-022-03192-1>
- Muehleisen RT, Bergerson J (2016) Bayesian Calibration - What, Why And How. In: *International High Performance Buildings Conference*. Purdue University, West Lafayette, IN, USA
- Oberkampf WL, Trucano TG, Hirsch C (2004) Verification, validation, and predictive capability in computational engineering and physics. *Appl Mech Rev* 57:345–384. <https://doi.org/10.1115/1.1767847>
- Oladyshkin S, Mohammadi F, Kroeker I, Nowak W (2020) Bayesian active learning for the gaussian process Emulator using information theory. *Entropy* 22:890. <https://doi.org/10.3390/e22080890>
- Oladyshkin S, Nowak W (2019) The connection between bayesian inference and information theory for Model Selection, Information Gain and Experimental Design. *Entropy* 21:1081. <https://doi.org/10.3390/e21111081>
- Olsen NRB, Haun S (2020) A numerical geotechnical model for computing soil slides at banks of water reservoirs. *Int J Geo-Eng* 11:22. <https://doi.org/10.1186/s40703-020-00129-w>
- Olsen NRB, Hillebrand G (2018) Long-time 3D CFD modeling of sedimentation with dredging in a hydropower reservoir. *J Soils Sediments* 18:3031–3040. <https://doi.org/10.1007/s11368-018-1989-0>
- Paola C, Voller VR (2005) A generalized Exner equation for sediment mass balance. *J Geophys Res Earth Surf* 110. <https://doi.org/10.1029/2004JF000274>
- Paul M, Negahban-Azar M (2018) Sensitivity and uncertainty analysis for streamflow prediction using multiple optimization algorithms and objective functions: San Joaquin Watershed, California. *Model Earth Syst Environ* 4:1509–1525. <https://doi.org/10.1007/s40808-018-0483-4>
- Rasmussen CE, Williams CKI (2006) *Gaussian processes for machine learning*. MIT Press, Cambridge, Mass
- van Rijn LC (2020) Erodibility of mud–sand Bed Mixtures. *J Hydraul Eng* 146:04019050. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0001677](https://doi.org/10.1061/(ASCE)HY.1943-7900.0001677)
- van Rijn LC, Barth R (2019) Settling and consolidation of soft mud–sand layers. *J Waterw Port Coast Ocean Eng* 145:04018028. [https://doi.org/10.1061/\(ASCE\)WW.1943-5460.0000483](https://doi.org/10.1061/(ASCE)WW.1943-5460.0000483)
- Rinaldi M, Nardi L (2013) Modeling interactions between Riverbank Hydrology and Mass failures. *J Hydraul Eng - ASCE* 139:1231–1240. [https://doi.org/10.1061/\(ASCE\)JHE.1943-5584.0000716](https://doi.org/10.1061/(ASCE)JHE.1943-5584.0000716)
- Scheurer S, Schäfer Rodrigues Silva A, Mohammadi F et al (2021) Surrogate-based bayesian comparison of computationally

- expensive models: application to microbially induced calcite precipitation. *Comput Geosci* 25:1899–1917. <https://doi.org/10.1007/s10596-021-10076-9>
- Schleiss AJ, Franca MJ, Juez C, De Cesare G (2016) Reservoir sedimentation. *J Hydraul Res* 54:595–614. <https://doi.org/10.1080/00221686.2016.1225320>
- Schmelter ML, Stevens DK (2013) Traditional and bayesian statistical models in Fluvial Sediment Transport. *J Hydraul Eng* 139:336–340. [https://doi.org/10.1061/\(ASCE\)HY.1943-7900.0000672](https://doi.org/10.1061/(ASCE)HY.1943-7900.0000672)
- Schmelter M, Wilcock P, Hooten M, Stevens D (2015) Multi-Fraction bayesian sediment transport model. *J Mar Sci Eng* 3:1066–1092. <https://doi.org/10.3390/jmse3031066>
- Shi BW, Yang SL, Wang YP et al (2012) Relating accretion and erosion at an exposed tidal wetland to the bottom shear stress of combined current–wave action. *Geomorphology* 138:380–389. <https://doi.org/10.1016/j.geomorph.2011.10.004>
- Simons RK, Canali GE, Anderson-Newton GT, Cotton GK (2000) Sediment transport modeling: Calibration, Verification, and evaluation. *Soil Sediment Contam* 9:261–289. <https://doi.org/10.1080/10588330091134239>
- Smith AFM, Gelfand AE (1992) Bayesian statistics without tears: a sampling-resampling perspective. *Am Stat* 46:84–88. <https://doi.org/10.2307/2684170>
- Sun J, Zhang F, Zhang X et al (2021) Severely declining suspended sediment concentration in the heavily dammed Changjiang Fluvial System. *Water Resour Res* 57. <https://doi.org/10.1029/2021WR030370>. e2021WR030370
- Verstraeten G, Poesen J (2001) Variability of dry sediment bulk density between and within retention ponds and its impact on the calculation of sediment yields. *Earth Surf Process Landf* 26:375–394. <https://doi.org/10.1002/esp.186>
- Villaret C, Kopmann R, Wyncoll D et al (2016) First-order uncertainty analysis using algorithmic differentiation of morphodynamic models. *Comput Geosci* 90:144–151. <https://doi.org/10.1016/j.cageo.2015.10.012>
- Walling DE, Webb BW (1996) Erosion and sediment yield: a global overview. In: *Proceedings of the Exeter Symposium*. IAHS, Exeter, UK, pp 3–19
- Widdows J, Brinsley MD, Bowley N, Barrett C (1998) A benthic annular flume for In Situ Measurement of suspension Feeding/ Biodeposition rates and Erosion potential of intertidal cohesive sediments. *Estuar Coast Shelf Sci* 46:27–38. <https://doi.org/10.1006/ecss.1997.0259>
- Winterwerp JC, van Kesteren WGM (2004) *Introduction to the physics of cohesive sediment in the marine environment*. Elsevier, Amsterdam; Boston
- Zarfl C, Lumsdon AE, Berlekamp J et al (2015) A global boom in hydropower dam construction. *Aquat Sci* 77:161–170. <https://doi.org/10.1007/s00027-014-0377-0>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.